

# Regional *vs* local drivers of phylogenetic and species diversity in stream fish communities

SIMON BLANCHET<sup>\*,†</sup>, MATTHEW R. HELMUS<sup>‡</sup>, SÉBASTIEN BROSSE<sup>†,§</sup> AND GAËL GRENOUILLET<sup>†,§</sup>

<sup>\*</sup>Station d'Ecologie Expérimentale du CNRS à Moulis, USR 2936, Moulis, France

<sup>†</sup>Centre National de la Recherche Scientifique (CNRS), Université Paul Sabatier, École Nationale de Formation Agronomique (ENFA), UMR-5174 EDB (Laboratoire Évolution & Diversité Biologique), 118 route de Narbonne, F-31062 Toulouse cedex 4, France

<sup>‡</sup>Amsterdam Global Change Institute, Department of Animal Ecology, Vrije Universiteit, Amsterdam, Netherlands

<sup>§</sup>Laboratoire Évolution et Diversité Biologique EDB, UMR 5174 UPS, ENFA, Université de Toulouse, Toulouse, France

## SUMMARY

1. Regional and local processes are thought to influence community phylogenetic and species diversity, but have proved challenging to distinguish. We disentangled the effects of local (limiting similarity and habitat filtering) and regional processes (biogeography, history and geographical constraints) on the species and phylogenetic diversity of native riverine fish communities across France.
2. Fish species richness (*SR*), phylogenetic diversity (estimated as the phylogenetic species variability metric, *PSV*) and trait evenness (estimated from 17 ecological and life-history traits) were estimated for fish at 932 sampling sites across 42 river basins in France. A phylogenetic signal in traits and species responses to environment variation were estimated using linear models. Null models were developed to test whether river basins constrain the *PSV* of the communities. Hierarchical linear models were fitted to *PSV* and *SR* to partition the variance explained by local and regional processes.
3. Fish species richness and phylogenetic diversity were negatively correlated across sites ( $r = -0.589$ ,  $P < 0.001$ ). There was a phylogenetic signal in species traits, but only a weak signal in species responses to environmental variation across sites. The null models suggested that river basins constrain community *PSV*. Local-scale predictors (temperature, landscape position and trait evenness) explained the variation in both *PSV* and *SR*, while regional-scale predictors (latitude, longitude and basin surface area) explained the variation only in *SR*. Significant interaction terms between local- and regional-scale predictors were also detected for both metrics.
4. Local-scale processes, habitat filtering and limiting similarity, probably determine the species composition of fish communities across France, while broad-scale, regional processes constrain community composition both directly and indirectly by affecting the strength of these local-scale processes. By analysing phylogenetic diversity and species richness simultaneously, a full picture of the drivers of local-scale diversity can be determined.

*Keywords:* environmental gradient, fish, hierarchical models, latitudinal gradient, metacommunity

## Introduction

Community assembly is driven by processes that operate at both local and regional scales. At regional scales, biogeographical and historical processes such as colonisation, extinction and speciation are important in determining which species can potentially disperse to and establish in a particular community. At the local scale,

community diversity can be determined by biotic and abiotic factors such as the presence of competitor species or environmental variability (Ricklefs, 2004; Graham, 2006; Leibold & McPeck, 2006; Graham *et al.*, 2009). For example, communities under high environmental stress (i.e. communities living in extreme/harsh environments) are expected to have low species richness compared to those not under such stress (Connell, 1978; Chase, 2007;

Correspondence: Simon Blanchet, Station d'Ecologie Expérimentale du CNRS à Moulis, 2 route du CNRS, 09 200 Moulis, France. E-mail: simon.blanchet@ecoex-moulis.cnrs.fr

Helmus *et al.*, 2010). On the other hand, unstressed communities could also have low species richness if the region where the community resides is isolated (Ricklefs, 1987). Regional- and local-scale processes can thus interact to determine community species richness (Harrison & Cornell, 2008).

These processes can result in non-random phylogenetic patterns in the composition of communities (Webb *et al.*, 2002; Cavender-Bares *et al.*, 2009; Vamosi *et al.*, 2009). An often expressed assumption is that if species traits, niches and sensitivities to abiotic and biotic factors show a phylogenetic signal (i.e. phylogenetically similar species are ecologically and functionally similar), then communities assembled through limiting similarity (i.e. when competition limits the closely related taxa) should show relatively high phylogenetic diversity (phylogenetic overdispersion) and contain relatively unrelated species. On the contrary, communities assembled through habitat filtering (i.e. when local biotic and abiotic conditions 'select' for individual species from a regional pool possessing traits fitting these local conditions) should show low phylogenetic diversity (phylogenetic clustering) and contain relatively closely related species (Webb *et al.*, 2002; Cavender-Bares *et al.*, 2009). However, it is now clear that multiple interacting processes determine the phylogenetic diversity of communities and that phylogenetic community structure (i.e. whether or not communities are phylogenetically clustered or overdispersed compared to some null hypothesis) cannot be used to infer an effect of limiting similarity or habitat filtering (Cavender-Bares *et al.*, 2004; Helmus *et al.*, 2007a; Valiente-Banuet & Verdu, 2007; Mayfield & Levine, 2010). For example, Mayfield & Levine (2010) argued that limiting similarity could lead either to phylogenetic clustering or overdispersion. This suggests that more data than just the presence/absence of species across communities and their phylogenetic relationships are required to infer structural processes. At the very least, data on species traits, habitat affinities and regional factors are also needed, and we argue here that these data need to be analysed in a hierarchical framework in order to infer how limiting similarity, habitat filtering and regional-scale processes interact.

The spatial and hierarchical nature at which these processes operate should be explicitly considered when attempting to understand their relative roles in community diversity (Harrison & Cornell, 2008). Hierarchical or multilevel regression models contain parameters estimated by probability models with their own estimated parameters (Gelman & Hill, 2006). These models are more informative than standard regression models

because they provide intuitive variance partitioning among covariates that operate at different spatial scales. Here, we use hierarchical regressions to ask how local-scale processes (limiting similarity and habitat filtering) interact with regional-scale processes and attributes (geographical location and spatial area) to determine community species richness and phylogenetic diversity. Species richness is a classical measure of species diversity, whereas phylogenetic diversity is a measure of diversity incorporating phylogenetic differences between species and hence their evolutionary histories. These two indices represent complementary measures of diversity (Crozier, 1997; Devictor *et al.*, 2010).

In the absence of experimental data, indirect methods must be used to test whether limiting similarity or habitat filtering affects community biodiversity. Since it cannot be assumed that an overdispersed phylogenetic community pattern is indicative of competition (Mayfield & Levine, 2010), trait evenness metrics, where the spacing of species in a community along an axis of trait values is estimated, are often used alternative methods for inferring the signature of limiting similarity in spatial data sets (Cornwell & Ackerly, 2009). Specifically, communities affected by limiting similarity should contain species that are evenly spaced along trait axes (Bowers & Brown, 1982). However, competitive dominants may have clustered, not even, trait distributions depending on the environment in which the community is found and the traits under study (Mayfield & Levine, 2010). Thus, if limiting similarity causes variance in species richness or phylogenetic diversity, there should be a correlation between community trait evenness, species richness and/or phylogenetic diversity, but the correlations may be weak if limiting similarity also causes trait clustering. An effective, non-experimental way to test for the effect of habitat filtering on biodiversity is to sample communities along environmental clines such as nutrient or elevation gradients (Horner-Devine & Bohannon, 2006; Bryant *et al.*, 2008). For example, Helmus *et al.* (2007a) and Newton *et al.* (2007) sampled fish and microbes, respectively, across lake acidity gradients and both found phylogenetic diversity and species richness to decrease with increasing water pH.

The regional context in which communities reside also influences their diversity (Ricklefs, 1987). At the regional scale, patterns in phylogenetic diversity and species richness are expected to reflect the biogeographical history of communities (e.g. colonisation history) and the spatial structure of regions (e.g. the area of the regions, the regional climate). For instance, Algar, Kerr & Currie (2009) highlighted a latitudinal phylogenetic diversity

and species richness gradient in American tree frogs (Hylidae), with South America being the centre of Hylidae evolution and subsequent colonisation and radiation of the Hyline clade into North America (Wiens *et al.*, 2006). However, species richness and phylogenetic diversity do not always covary at regional scales (Crozier, 1997; Devictor *et al.*, 2010). This may be because particular attributes of regions, such as total area, affect richness more than phylogenetic diversity (Rosenzweig & Ziv, 1999; Helmus & Ives, 2012); however, how particular regional attributes affect phylogenetic diversity versus species richness has not been well studied.

The aim of our work was to identify local- and regional-scale drivers of fish phylogenetic diversity and species richness in French rivers. We first estimated phylogenetic signal in ecologically important traits to test whether closely related species have similar traits (Blomberg, Garland & Ives, 2003). We then compared the observed phylogenetic diversity of local communities to null models, to test whether river basins (i.e. the regional pools in our study) influence the phylogenetic composition of local communities residing in the basins (Hugueny, 1989). Finally, we tested for covariation between phylogenetic diversity and species richness at the local scale and explored whether similar processes affected phylogenetic diversity and species richness. We used a combination of hierarchical models and species-specific regressions for this final analysis (Gelman & Hill, 2006; Helmus *et al.*, 2007a). Our work provides a comprehensive picture of the regional- and local-scale processes that determine the freshwater fish diversity across the entirety of France. We worked under several predictions based on previous studies on the environmental and regional determinants of the French fish community diversity.

## Methods

### *French fish community diversity*

We obtained a database of freshwater fish presence/absence across 1110 sampling sites in France. The database was provided by the Office National de l'Eau et des Milieux Aquatiques [ONEMA, see Buisson, Blanc & Grenouillet (2008) for further details]. A standard fishing protocol was conducted yearly during low-flow periods by the ONEMA to collect information on fish assemblages present in pre-defined river sections. To account for potential yearly variation in species composition at the sampling site level (i.e. some rare species may have not been sampled in certain years), our list of species (presence/

absence) at the site level was derived by combining the information for a period ranging from 2004 to 2009 and the species list at the sampling level was derived from six fishing occasions at each site. The status of each species within a catchment (native or translocated from another French catchment) was determined from published sources (Keith and Allardi, 2007). From the 1110 initial sites, we discarded those represented by a single species, and we focused on native composition only. As most, if not all, introductions are recent (i.e. most have been introduced <100 years ago, although some of them such as carp have been introduced ~2000 years ago), it makes little sense to analyse the whole fish (native + non-native species) composition within a framework including biogeographical processes. It is noteworthy that analyses including non-native species as a predictor were run and provided similar results to those not including this predictor (not shown). Our final database contained 28 native fish species distributed in 931 sampling sites in 42 river basins evenly scattered across France. Each river basin was represented by at least two sampling sites; the longer the river basin, the higher the number of sampling sites. Sampling sites were separated by at least ten kilometres to ensure a relative independence between local communities.

We used three mitochondrial genes (cytochrome b, cytochrome oxidase I and ribosomal 16S subunit) to reconstruct phylogenetic relationships between the 28 studied fish species. Sequence data were obtained from GenBank, and alignments were performed using the software ClustalW (Thompson, Higgins & Gibson, 1994). The best model of evolution for each gene was calculated using the software Paup\*4.0b10 (Swofford, 1993) and selected using the software Modeltest (through Akaike Information Criteria –AIC values, Posada & Buckley, 2004). We used maximum likelihood (as implemented in PhyML) to build a phylogenetic tree that accounted for the model of evolution of each gene (Guindon & Gascuel, 2003). *Lampetra planerii* was used as outgroup in the reconstruction of the phylogenetic relationships between the studied fish species.

We correlated fish species richness (*SR*) with phylogenetic species variability (*PSV*) across the 931 communities in our data set. The metric *PSV* is defined as:

$$PSV = \frac{ntrC - \sum C}{n(n-1)}$$

where *n* is the number of species, *C* is a covariance matrix that summarises the phylogenetic relationships of the *n* species, *trC* is the sum of diagonal elements of *C*, and  $\sum C$  denotes the sum of all elements of *C* (Helmus *et al.*, 2007b). The values of *PSV* fall between 0 and 1,

with values closer to one indicating communities comprising unrelated species. The metric is calculated relative to a particular phylogenetic tree, in our case the 28 French fish phylogeny we constructed. The expected values of *PSV* for a set of randomly assembled communities are statistically independent of species richness; thus, it was a sound metric to use to look for covariation among species richness and phylogenetic diversity (Helmus *et al.*, 2007b).

#### *Do closely related fishes have similar traits?*

Community phylogenetic diversity will only be influenced by local-scale ecological processes if it is related to ecologically important traits. We obtained 17 categorical and quantitative ecological traits related to feeding, habitat, reproduction and morphology of the 28 fishes [see Buisson & Grenouillet (2009) and Table S1 for details of these traits]. We performed a principal component analysis (PCA) to reduce dimensionality among the traits and to describe better the overall ecology of the species (Ricklefs & Miles, 1994). To account for the presence of both continuous and discrete variables, the PCA was performed on a matrix of pairwise Gower distances (Gower, 1971) among the 28 species and based on the 17 traits (Laliberté & Legendre, 2010; Villéger, Novack-Gottshall & Mouillot, 2011). The PCA scores of the 28 fish species were taken from the first three axes (accounting for more than 75% of the total variance) to obtain three orthogonal estimates of composite functional trait values for each species ('trait 1', 'trait 2' and 'trait 3', see Table S1 for details).

We addressed whether the three composite traits exhibited phylogenetic signal (i.e. are closely related species more likely to have similar trait values?). We fitted to each trait a phylogenetic generalised least-squares model with a parameter,  $d$ , that gives the strength of phylogenetic signal (Lavin *et al.*, 2008; Helmus *et al.*, 2010). The parameter  $d = 0$  when there is no phylogenetic signal in the trait values (i.e. a star phylogeny best fits the trait data). When  $0 < d < 1$ , there is signal, but it less than expected based on the Brownian motion model of trait evolution, with the expectation matched when  $d = 1$ . Any  $d > 1$  indicates more phylogenetic signal than what is expected by Brownian motion evolution and can be considered indicative of phylogenetic trait conservatism (Losos, 2008).

#### *Do river basins constrain the biodiversity of fish communities?*

We compared two null models to test whether river basins constrain the phylogenetic diversity of local

communities. We determined whether the observed average *PSV* value across all communities ( $PSV_{obs}$ ) differed ( $\alpha = 0.05$ ) from the distributions of two null hypotheses varying in the definition of the regional pool. Our first null hypothesis was that communities could be colonised by all species in France regardless of whether a particular species was found in a particular river basin. Given a matrix with the 28 species as columns and 931 communities as rows, the permutation that tested this first null hypothesis shuffled cells within columns maintaining the observed total prevalence of species across all communities (i.e. the frequency permutation in Kembel *et al.*, 2010). Our second null hypothesis was that communities were assembled only from the species found in the particular basin in which communities were located (i.e. there was a separate regional pool for each of the 42 basins). To do this, we separately permuted each of the 42 basin community-by-species matrices as for the first null model and calculated the mean null *PSV* across all the permuted matrices. If river basins constrain the species present within communities, then the observed *PSV* should fall outside the null distribution of the first model, but not the second. Each permutation was performed 10000 times to create the null distributions.

#### *Hierarchical models to test for the determinants of community biodiversity*

Varying intercept, varying slope, two-level, hierarchical linear models were fitted to the *SR* and *PSV* data. Intercepts and slopes were allowed to vary across river basins, and this accounted for autocorrelation between sampling sites belonging to a same river basin (Gelman & Hill, 2006). A Poisson error term for *SR* and a Gaussian error term for *PSV* best-fitted the data. Five local-scale (i.e. level 1) predictors were fitted to the community diversity data. The assumption we made by fitting the *PSV* data to local-scale predictors was that *PSV* gives an estimate of the ecological similarity of species. This is the assumption that we tested with the phylogenetic signal trait methods described previously. Communities were also grouped by river basin, and the variance among the basins in their estimated intercepts and local-scale predictor regression coefficients was fitted to four regional-scale predictors (i.e. level 2). The assumption we made behind grouping sites to basin was that each community was a sampling from a basin-specific species pool. This was the assumption we tested with our null model tests described previously. By fitting the model with regional-scale predictors, we

specifically addressed the question of what characteristics about river basins affect the phylogenetic diversity and species richness of French freshwater fish communities. We also explicitly tested for interactions between the regional- and local-scale predictors in the model.

We fitted two local-scale abiotic covariates, mean annual air temperature (°C) and landscape position along the river gradient. As streams are reasonably well-mixed waterbodies that easily exchange heat with the atmosphere, air and river water temperatures are strongly positively correlated, especially for large-scale studies (Mohseni & Stephan, 1999; Caissie, 2006). Therefore, air temperatures were used as a surrogate for water temperatures. Air temperatures were extracted from the CRU CL 2.0 (Climatic Research Unit Climatology version 2.0) data set (New *et al.*, 2002) at a resolution of  $10' \times 10'$  and then averaged for the period 1961–1990 to describe the current climate. Landscape position was derived from a PCA on two variables that described the spatial position of sites in the river network: distance from the headwater source and surface area of the drainage basin above the sampling site. These two variables were derived from a Global Information System (GIS) using the software ArcGIS® (Esri, Redlands, CA, USA). The first axis of the PCA, accounting for 93.2% of the total variability, was kept as a synthetic variable describing the position along the upstream–downstream gradient. Low values along this axis indicate upstream sites (Buisson *et al.*, 2008).

Three additional local-scale predictors were estimates of the evenness of trait spacing within communities and were used to infer the influence of limiting similarity. We calculated the standard deviation of nearest neighbour distances (sdNND) that was standardised to account for differences in species richness (Stubbs & Bastow Wilson, 2004; Cornwell & Ackerly, 2009) for each of the three composite traits. Briefly, for each composite trait and each fish community, the species in each community were sorted with respect to their trait values and the trait distance of each gap between successive species was calculated. Then, the standard deviation of the distribution of these gaps was calculated. If the gaps are similar among the species of a community, then this value is low – species are evenly spaced within the community along the trait axis. To remove any correlation between these standard deviations and the species richness of communities, we derived a null distribution of trait-gap standard deviations for each level of species richness seen in our communities by randomly drawing  $n$  trait values 1000 times for each  $n$ . The observed values were then standardised by subtracting the mean of the

distribution corresponding to the  $n$  of the community and dividing by the distribution standard deviation. These standardised values were then multiplied by  $-1$  to obtain a metric that increases with increasing trait evenness. We do not use trait variances or ranges here since they are more prone to be affected by both habitat filtering and limiting similarity (Cornwell & Ackerly, 2009). We termed these three metrics 'evenness 1', 'evenness 2' and 'evenness 3', respectively, for the three composite traits.

We obtained four regional-scale predictors from GIS and grey literature: basin mean latitude, basin mean longitude, basin surface area and the altitudinal range of the basin. These four predictors can all be thought of as regional-scale characteristics that influence local community diversity either through historical contingencies (e.g. some basins may not have been colonised by some lineages) or through broad-scale geographical clines in climate (e.g. in France lower latitudes are associated with more Mediterranean climates).

For each diversity metric (*PSV* and *SR*), we fitted the full model (i.e. five local-scale predictors, four regional-scale predictors and the 20 two-term interactions) and then used a stepwise selection procedure based on AIC to select the best predictors and hence the best model for supporting the data. Since AIC is not fully defined for multilevel hierarchical models, we only used this as a guide to work towards a parsimonious model with all coefficients significant at the 0.05 level (Gelman & Hill, 2006). All the predictors were transformed to z-scores to standardise the slope coefficients ( $\beta$ ) in order to compare the relative strength of the predictors.

#### *Individual species regressions*

In order to explain more fully the *SR* and *PSV* patterns, we used hierarchical regressions to fit the presence/absence of each species to the variables selected in the best-fit models of *PSV* and *SR*. We omitted trait evenness variables since the meaning of such variables at the species level is unclear. Phylogenetic signal in the coefficient estimates was then tested as described previously (Helmus *et al.*, 2007a).

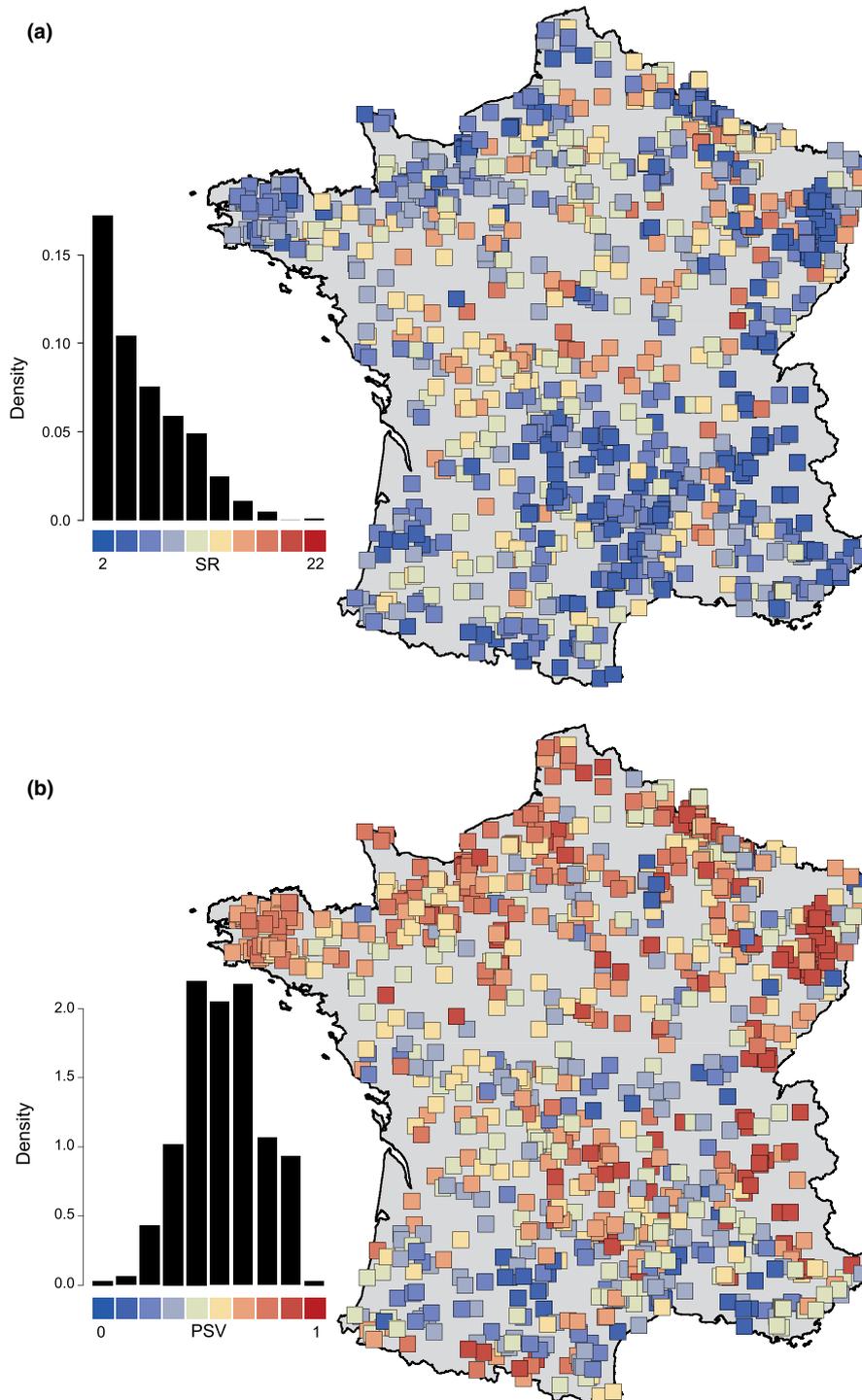
All these and the previously described analyses were performed in R with the *lme4* and *picante* packages (Bates, Maechler & Bolker, 2010; Kembel *et al.*, 2010).

## Results

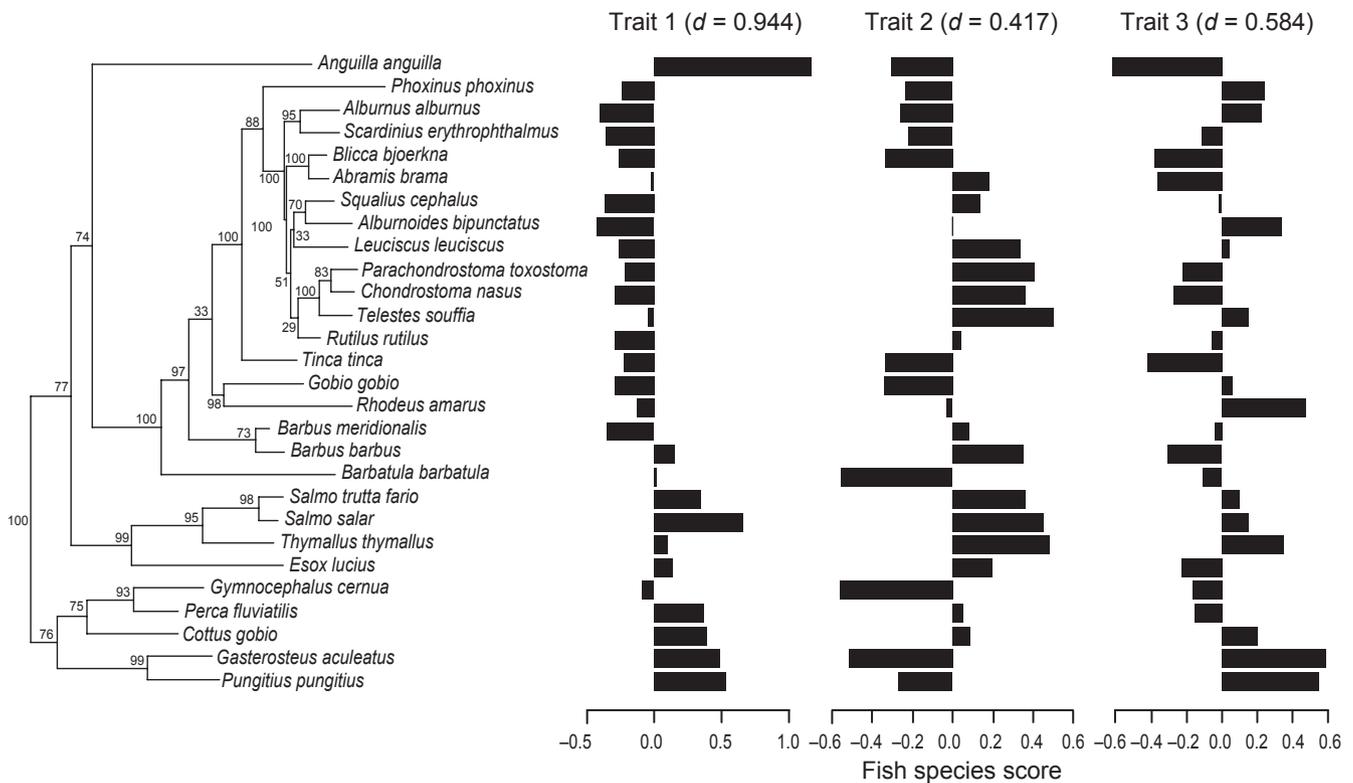
At the local scale, we found a negative correlation between *SR* and *PSV* across French freshwater fish communities

( $r_{\text{Pearson}} = -0.589$ ,  $t = -22.249$ , d.f. = 929,  $P < 0.001$ , Fig. 1). The three composite functional traits (and particularly trait 1) exhibited a phylogenetic signal (Table S2, Fig. 2). The fish communities on average were significantly

overdispersed (i.e. contained relatively unrelated species) compared to null communities assembled irrespective of basin identity ( $PSV_{\text{obs}} = 0.737$ ;  $PSV_{\text{null}} = 0.703$ , 95% C.I. = 0.696–0.710), but were not different from communi-



**Fig. 1** Spatial distribution of (a) fish species richness (SR) and (b) phylogenetic species variability (PSV) across France. Histograms (Kernel density plots) represented the statistical distribution of each metric.



**Fig. 2** Closely related French fish species have similar traits. A principal component analysis was performed on 17 functional traits, and the species scores across each of the first three PCA axes were taken as three composite traits.  $d$  gives the level of phylogenetic signal (Table S2). Numbers on the tree indicate Bayesian posterior probabilities (%).

ties assembled from randomisation within basins ( $PSV_{null} = 0.702$ , 95% C.I. = 0.626–0.860).

The best-fit hierarchical models to explain  $PSV$  and  $SR$  contained similar local-scale predictors (Table 1). Annual mean temperature and landscape position negatively affected  $PSV$ , while these same two predictors positively affected  $SR$ . Trait evenness measured for composite trait 3 negatively affected  $PSV$ , while trait 2 evenness negatively affected  $SR$ . Fish species richness was much more affected by regional-scale processes than was phylogenetic diversity (Table 1). Basin mean latitude, basin mean longitude and basin area were all selected as main effects in the  $SR$  model, while no regional predictors were main effects in the  $PSV$  model. Communities tended to be more species rich in northern rather than southern river basins, in western rather than eastern river basins and in larger rather than smaller river basins (Fig. 1). Three interaction terms between regional- and local-scale predictors were also significant in the  $SR$  model. The three terms involved trait evenness metrics, indicating that the slopes of the relationships between trait evenness and species richness were contingent upon the river basins to which communities belong. In

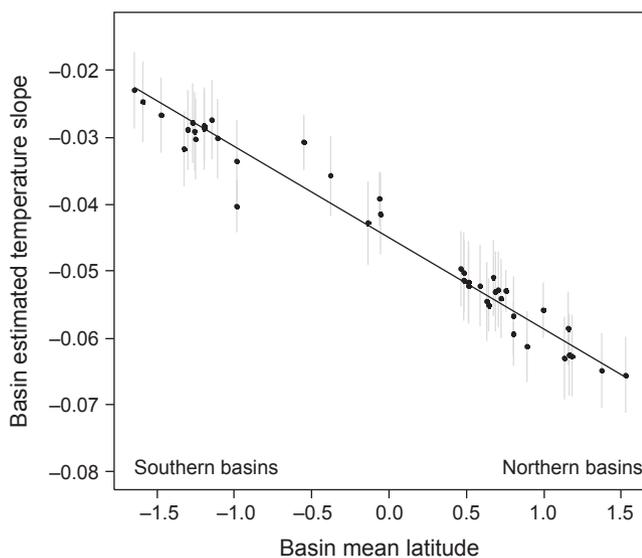
the  $PSV$  model, there was an interaction between latitude and temperature, indicating that the negative relationship between mean temperature and  $PSV$  was more pronounced in northern compared to southern river basins (Fig. 3).

We fitted hierarchical models with basin mean latitude, basin mean longitude, basin area, air temperature and landscape position as covariates to the presence/absence of each of the 28 species in order to understand better how individual species responses to these variables affected community diversity. A significant phylogenetic signal was detected only for the effect of temperature – Cyprinidae species increased, while Salmonidae species decreased with temperature. Since Cyprinidae was the most diverse and derived major clade in our phylogeny, the increase in this clade coupled to the decrease in the Salmonidae clade (*Salmo* and *Thymallus*), a distant clade relative to the cyprinids, probably caused the negative correlation between  $PSV$  and temperature. The prevalence of most species tended to be positively associated with landscape position of the site along the river gradient, as well as with latitude and basin area (Fig. 4).

**Table 1** Multilevel hierarchical model coefficients of the variables that best explain French river fish community phylogenetic diversity (PSV) and species richness (SR)

Predictors	Process involved	Estimate (95% CI)	
		PSV	SR
Basin latitude	Biogeography		0.202 (0.186; 0.252)
Basin longitude	Biogeography		-0.088 (-0.154; -0.022)
Basin area	Biogeography		0.067 (0.026; 0.108)
Temperature	Habitat filtering	-0.480 (-0.564; -0.395)	0.184 (0.138; 0.223)
River gradient	Habitat filtering	-0.295 (-0.366; -0.224)	0.219 (0.182; 0.245)
Trait evenness 3	Limiting similarity	-0.203 (-0.257; -0.149)	
Trait evenness 2	Limiting similarity		-0.050 (-0.077; -0.022)
Basin longitude * Trait evenness 1	Biogeography-limiting similarity interaction		0.044 (0.002; 0.087)
Basin area * Trait evenness 3	Biogeography-limiting similarity interaction		-0.035 (-0.051; -0.019)
Basin latitude * Trait evenness 3	Biogeography-limiting similarity interaction		0.049 (0.013; 0.084)
Basin latitude * Temperature	Biogeography-habitat filtering interaction	-0.126 (-0.217; -0.035)	

Model coefficients are indicated with their 95% C.I. (in brackets). All variables were scaled (z-scores) so the coefficients are directly comparable.



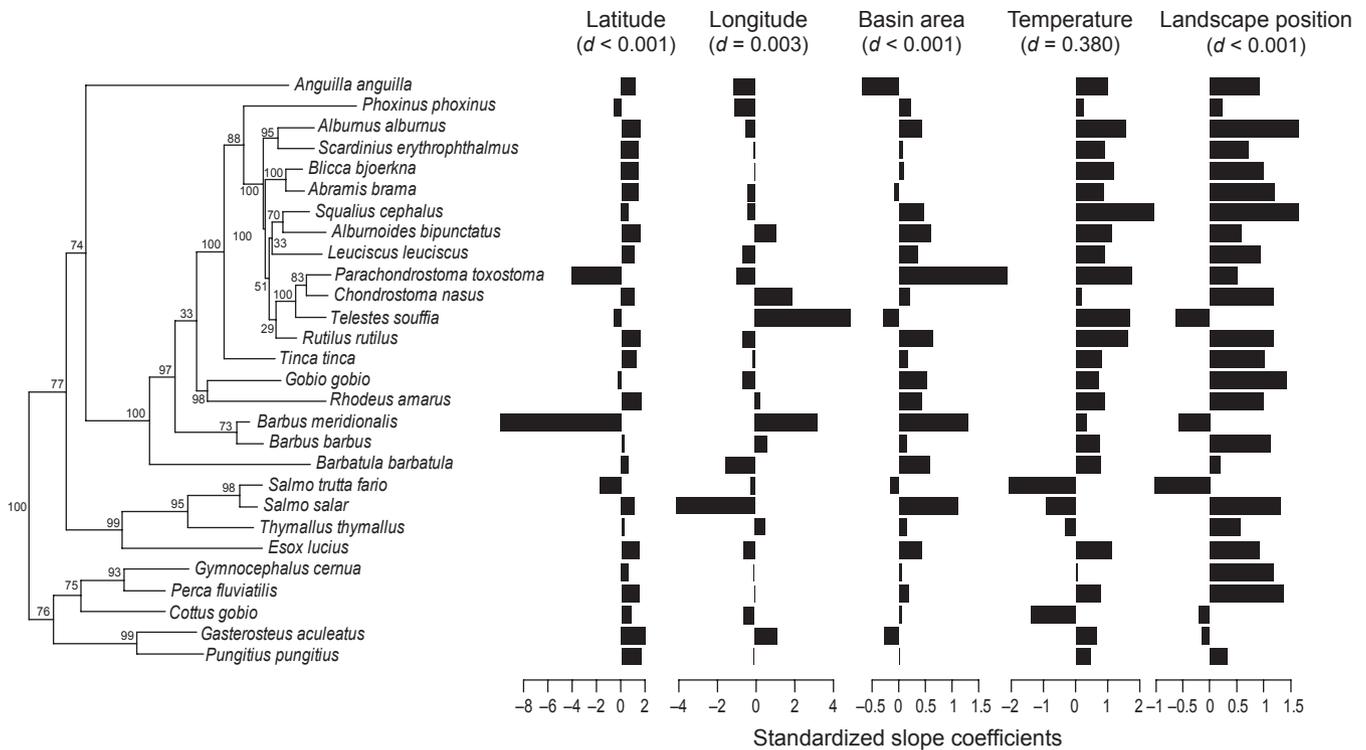
**Fig. 3** The effect of temperature on fish community phylogenetic diversity (PSV) depends on basin latitudinal position. Each dot represents the slope ( $\pm$  SE) of the relationship between PSV and temperature for each river basin. Slopes were estimated using a hierarchical linear model, and the solid line represents the fixed effect across all basins (see methods for details).

## Discussion

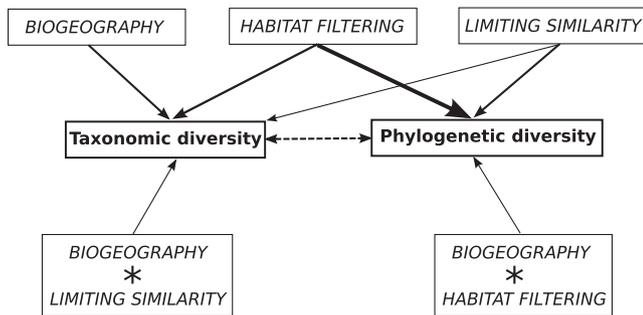
Local- and regional-scale processes are thought to interact to determine the biodiversity of ecological communities. The riverine fish communities of France are no exception. We found that the phylogenetic diversity and species richness of French fish communities are

correlated with a variety of interacting regional- and local-scale factors (synthesised in Fig. 5), suggesting that multiple-scale processes are driving both diversity indices. More precisely, our results suggest that habitat filtering along the same environmental gradients (temperature and landscape position) most likely affects both measures of biodiversity. Limiting similarity (as indirectly measured through trait evenness metrics) mainly affected phylogenetic diversity, and regional-scale processes affected both species richness and phylogenetic diversity, but more so for species richness than phylogenetic diversity (Fig. 5). Below, we explain how limiting similarity, habitat filtering and regional-scale processes may interact to shape biodiversity patterns of French freshwater fishes. Inferring processes from correlative approaches as the one developed here always has to be considered with care. We therefore focus on the main processes supported by our analyses that have shaped French freshwater fish communities.

The biogeographical history of French freshwater fauna is still poorly understood (Persat & Keith, 1997; Costedoat *et al.*, 2006). The only well-accepted pattern is that the Danube drainage acted as one of the most important refuges during the last glaciation period (Reyjol *et al.*, 2007). From this refuge, the most probable colonisation route for French river basins was from the Danubian refuge to close river basins such as the Rhine or the Rhone and then to neighbouring river basins (Reyjol *et al.*, 2007; Costedoat & Gilles, 2009). Hence, in France, we expected north-south and east-west gradients of phylogenetic diversity and species richness if most colonisation events



**Fig. 4** The effect of regional-scale (latitude, longitude and basin area) and local-scale (temperature and landscape position) predictors on the distributions of French riverine fish species. Bars are standardised coefficients from a hierarchical regression of species presence/absence regressed on each predictor. Only temperature exhibits a significant phylogenetic signal as estimated by a phylogenetic regression (see also Table S2).



**Fig. 5** Conceptual representation of the main drivers of taxonomic and phylogenetic diversity. Arrows are derived from the multilevel hierarchical models, and their thickness accounts for the relative importance of each effect. The dotted double arrow indicates the negative correlation between the two diversity metrics.

occurred from the Danubian refuge. Regarding basin area, positive relationships with species and phylogenetic diversity are expected as bigger basins generally contain more species from unrelated clades (Hugueny, Movellan & Belliard, 2011; Helmus & Ives, 2012).

We found that river basins do constrain the community biodiversity of French freshwater fish communities. Indeed, large basins on average had communities with higher species richness than small basins, and fish communities contained more unrelated species than

expected if all fish species present in French catchments were able to colonise freely all river basins. River basins thus define a species pool from which local communities are drawn, and dispersal among these basins is probably constrained by ancient and present-day waterway connections (Reyjol *et al.*, 2007; Hugueny *et al.*, 2011). The prevalence of most species increased with latitude causing community species richness to increase with latitude. This north–south pattern is expected given that colonisation from refuge during the last glaciation occurred through river basins situated in northern France. However, there was no effect of latitude (nor other regional-scale factors) on phylogenetic diversity. This indicates that the fishes that dispersed out of glacial refuge(s) from north to south were not a phylogenetic subset of the present-day native species of France. Instead, closely related species have very different latitudinal patterns (for example, compare the two *Barbus* for latitude in Fig. 4), indicating either differences in where particular species took refuge during the last glaciation and/or differences in dispersal ability that are unrelated to phylogenetic relatedness.

River basins are characterised by an environmental gradient from the river origins to its mouth (Vannote

et al., 1980). For example, river width, depth and water temperature increase, while river slope and water velocity decrease (Montgomery, 1999). Such abiotic changes are associated with changes in the biological diversity of the fauna. Particularly, fish species richness tends to increase from upstream to downstream (Rahel & Hubert, 1991), a general observation that holds true for the French fish fauna (Buisson et al., 2008). Environmental conditions in upstream sites are considered as harsh (e.g. higher thermal amplitudes, Oberdorff et al., 1998), and thus, we hypothesised that communities in the upstream reaches of rivers should contain closely related species and that phylogenetic diversity should increase along the river gradient. Hence, along the river gradient, species richness should covary positively with phylogenetic measures of diversity, with upstream sites containing a pool of species with low species richness and a low phylogenetic diversity.

We found that ecological traits were related to phylogeny, which indicates that French fishes that were more closely related to each other tended to have similar composite trait values. This suggests that the phylogenetic diversity of French riverine fish communities is a good surrogate of the ecological similarity of species within communities. Fish phylogenetic diversity may thus be affected by local-scale processes such as habitat filtering (Peres-Neto, 2004; Helmus et al., 2007a). We found evidence that habitat filtering occurred at upland, cold temperature reaches where salmonids (*Thymallus thymallus*, *Salmo salar* and *S. trutta fario*) and other species, such as *Cottus gobio* distantly related to the salmonids, tend to reside. Habitat filtering for cold-tolerant species decreased from upstream to downstream and communities gained more species than they lost. This mechanism probably accounts for the positive relationships between species richness and water temperature and species richness and landscape position. The species gained were cyprinids, and the species lost were salmonids and other species distantly related to cyprinids. Thus, phylogenetic diversity decreased along the river continuum from upstream to downstream. Habitat filtering probably caused the strong negative correlation between community species richness and phylogenetic diversity (Fig. 5).

However, our results suggest that habitat filtering was modified by the basin characteristics. Indeed, the strength of the negative relationship between temperature and phylogenetic diversity was weaker in southern compared to northern river basins. This may be explained because the lower/minimum observed temperatures in rivers within basins in the south may be higher than those in the north. This loss of cold temper-

ature reaches would decrease the effect of habitat filtering of cold-tolerant species in southern basins. This seems to be the case as we found a weak, but significant, negative relationship between latitude and minimal air temperature across France ( $r = -0.12$ ,  $P < 0.001$ ). Assuming a strong correlation between air and water temperature (Mohseni & Stephan, 1999; Caissie, 2006), such environmental discrepancy between northern and southern river basins may explain the interaction term we highlight here.

This interaction term between basin latitude and temperature in the PSV model also suggests the mechanism by which French river basins constrain the phylogenetic diversity of fish communities. In our null model analyses, we found that communities were significantly overdispersed compared to a null model that randomised species across basins, but were not significantly structured when compared to a null model that randomised species within basins. Basins that had higher spatial variation in temperature – those that contain sites with colder temperatures – are found in more northern latitudes. As PSV was significantly associated with temperature, these northern basins should have contributed most to the observed overdispersed phylogenetic structure of the unconstrained first null model. This can be seen in the strong correlation ( $r^2 = 0.485$ ,  $P < 0.001$ , d.f. = 40) between basin latitude and the mean null 1 standardised effect size [i.e.  $SES = (PSV_{obs} - \text{mean}(PSV_{null1}))/SD(PSV_{null1})$ ]. When compared to the null expectation from model 1, sites in northern river basins were thus more overdispersed than those in southern river basins. Therefore, the geographical and climate structure of basins might constrain the phylogenetic diversity of local communities by influencing the impact of habitat filtering – the spatial location of a basin may determine how much temperature filters the phylogenetic composition of French fish communities.

There is indirect evidence that limiting similarity influences French fish communities. We found that phylogenetically diverse and species-rich communities tend to have less even trait distributions along composite traits 2 and 3 than less-diverse communities. Basin-level characteristics might also influence the effect of limiting similarity on species richness; surface area, latitude and longitude interact with how trait evenness metrics 1 and 3 affect species richness. However, the effect sizes of these evenness metrics on species richness are quite low; most are an order of magnitude lower than those coefficients associated with habitat filtering (Murray & Conner, 2009). In contrast, trait evenness metric 3 does strongly affect phylogenetic diversity. Trait 3 most corre-

lates with fecundity, lifespan and body length (Table S1). According to Mayfield & Levine (2010), these traits can be seen as traits related to the competitive ability differences between species (i.e. traits driving competitive exclusion), rather than to traits related to niche differences (i.e. traits that cause species to limit themselves more than their competitors). Sites having high PSV values are generally characterised by low water temperature, and interestingly, species inhabiting those upstream sites (salmonids, *Cottus gobio*) all have positive values regarding traits metric 3, which explains the low evenness values measured at upstream sites containing high PSV.

In phylogenetically less-diverse communities (i.e. downstream sites), communities are dominated by cyprinids for which trait 3 values vary. Mayfield & Levine (2010) argued that, when traits contributing to differences in competitive ability mainly control community assembly and if traits are significantly related to phylogeny, we can expect competitive interactions to reduce trait evenness (and hence phylogenetic diversity) in communities. Following this reasoning, and assuming that traits contributing to competitive ability differences predominate the assembly of these communities (i.e. those synthesised by trait 3), we can hence speculate that competitive interaction is stronger downstream which led to communities with less even traits and lower phylogenetic diversity. Overall, this would indicate that habitat filtering would be the leading process in upstream communities, whereas limiting similarity would mainly shape species and phylogenetic diversity downstream (or alternatively the effect of habitat filtering may decrease downstream).

In conclusion, by analysing species richness and phylogenetic diversity simultaneously with analyses of individual species responses to spatial environmental and regional gradients, we were able to obtain a detailed picture of how multiple processes might operate to affect the local-scale community composition and biodiversity. Furthermore, by analysing our data in a hierarchical framework, we could infer how local- and regional-scale processes act and interact. Our results, based on freshwater fish, are constrained by the low taxonomic diversity of the French fish fauna together with the strong dominance of cyprinids. The richness and the taxonomy of the species pool are nevertheless explained by historical constraints (i.e. post-glacial recolonisation, Reyjol *et al.*, 2007), and our results probably hold true for Western European fish assemblages that share a similar species pool and the same biogeographical history. Applying a similar approach under different biogeo-

graphical contexts and species pools would be of particular interest to gain a more comprehensive view of the forces that shape both taxonomic and phylogenetic structure of fish assemblages.

Moreover, from a methodological point of view, our analyses were by necessity disjointed: we used separate analyses to assess phylogenetic signal in traits, species responses to the environment, the effect of river basins on phylogenetic diversity, the best predictors of community phylogenetic diversity and species richness and the influence of these predictors on individual species distributions across France. As more integrated methods are developed (e.g. Ives & Helmus, 2011), more sophisticated and predictive analyses of the questions we address here will become possible, and this should clarify the exact processes driving biodiversity patterns in freshwater communities. Regardless, we have demonstrated that regional- and local-scale processes probably interact in complex ways to shape taxonomic and phylogenetic community biodiversity patterns.

### Acknowledgments

MRH was funded by NSF-DBI-0906011 (Bioinformatics postdoc fellowship) and an Amsterdam Global Change Institute postdoctoral fellowship. We are indebted to the Office National de l'Eau et des Milieux Aquatiques (ONEMA) for providing fish data. This work has been carried out in two research units (SEEM and EDB) that are part of the 'Laboratoire d'Excellence' (LABEX) entitled TULIP (ANR-10-LABX-41).

### References

- Algar A.C., Kerr J.T. & Currie D.J. (2009) Evolutionary constraints on regional faunas: whom, but not how many. *Ecology Letters*, **12**, 57–65.
- Bates D., Maechler M. & Bolker B. (2010) *lme4: Linear mixed-effects models using Eigen and S4*, R packages.
- Blomberg S.P., Garland T. & Ives A.R. (2003) Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution*, **57**, 717–745.
- Bowers M.A. & Brown J.H. (1982) Body size and coexistence in desert rodents: chance or community structure? *Ecology*, **63**, 391–400.
- Bryant J.A., Lamanna C., Morlon H., Kerkhoff A.J., Enquist B.J. & Green J.L. (2008) Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 11505–11511.
- Buisson L., Blanc L. & Grenouillet G. (2008) Modelling stream fish species distribution in a river network: the

- relative effects of temperature versus physical factors. *Ecology of Freshwater Fish*, **17**, 244–257.
- Buisson L. & Grenouillet G. (2009) Contrasted impacts of climate change on stream fish assemblages along an environmental gradient. *Diversity and Distributions*, **15**, 613–626.
- Caissie D. (2006) The thermal regime of rivers: a review. *Freshwater Biology*, **51**, 1389–1406.
- Cavender-Bares J., Ackerly D.D., Baum D.A. & Bazzaz F.A. (2004) Phylogenetic overdispersion in Floridian oak communities. *American Naturalist*, **163**, 823–843.
- Cavender-Bares J., Kozak K.H., Fine P.V.A. & Kembel S.W. (2009) The merging of community ecology and phylogenetic biology. *Ecology Letters*, **12**, 693–715.
- Chase J.M. (2007) Drought mediates the importance of stochastic community assembly 10.1073/pnas.0704350104. *Proceedings of the National Academy of Sciences*, **104**, 17430–17434.
- Connell J.H. (1978) Diversity in tropical rain forests and coral reefs - high diversity of trees and corals is maintained only in a non-equilibrium state. *Science*, **199**, 1302–1310.
- Cornwell W.K. & Ackerly D.D. (2009) Community assembly and shifts in plant trait distributions across an environmental gradient in coastal California. *Ecological Monographs*, **79**, 109–126.
- Costedoat C., Chappaz R., Barascud B., Guillard O. & Gilles A. (2006) Heterogeneous colonization pattern of European Cyprinids, as highlighted by the dace complex (*Teleostei : Cyprinidae*). *Molecular Phylogenetics and Evolution*, **41**, 127–148.
- Costedoat C. & Gilles A. (2009) Quaternary pattern of freshwater fishes in Europe: comparative phylogeography and conservation perspective. *The Open Conservation Biology Journal*, **3**, 36–48.
- Crozier R.H. (1997) Preserving the information content of species: genetic diversity, phylogeny, and conservation worth. *Annual Review of Ecology and Systematics*, **28**, 243–268.
- Devictor V., Mouillot D., Meynard C., Jiguet F., Thuiller W. & Mouquet N. (2010) Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: the need for integrative conservation strategies in a changing world. *Ecology Letters*, **13**, 1030–1040.
- Gelman A. & Hill J. (2006) *Data Analysis using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York.
- Gower J.C. (1971) A general coefficient of similarity and some of its properties. *Biometrics*, **27**, 857–874.
- Graham C.H. (2006) Habitat history improves prediction of biodiversity in rainforest fauna. *Proceedings of the National Academy of Sciences*, **103**, 632–636.
- Graham C.H., Parra J.L., Rahbek C. & McGuire J.A. (2009) Phylogenetic structure in tropical hummingbird communities. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 19673–19678.
- Guindon S. & Gascuel O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696–704.
- Harrison S. & Cornell H. (2008) Toward a better understanding of the regional causes of local community richness. *Ecology Letters*, **11**, 969–979.
- Helmus M.R., Bland T.J., Williams C.K. & Ives A.R. (2007b) Phylogenetic measures of biodiversity. *American Naturalist*, **169**, E68–E83.
- Helmus M.R. & Ives A.R. (2012) Phylogenetic diversity area curves. *Ecology*, **93**, S31–S43.
- Helmus M.R., Keller W., Paterson M.J., Yan N.D., Cannon C.H. & Rusak J.A. (2010) Communities contain closely related species during ecosystem disturbance. *Ecology Letters*, **13**, 162–174.
- Helmus M.R., Savage K., Diebel M.W., Maxted J.T. & Ives A.R. (2007a) Separating the determinants of phylogenetic community structure. *Ecology Letters*, **10**, 917–925.
- Horner-Devine M.C. & Bohannan B.J.M. (2006) Phylogenetic clustering and overdispersion in bacterial communities. *Ecology*, **87**, S100–S108.
- Hugueny B. (1989) West-African rivers as biogeographic islands - species richness of fish communities. *Oecologia*, **79**, 236–243.
- Hugueny B., Movellan A. & Belliard J. (2011) Habitat fragmentation and extinction rates within freshwater fish communities: a faunal relaxation approach. *Global Ecology and Biogeography*, **20**, 449–463.
- Ives A.R. & Helmus M.R. (2011) Generalized linear mixed models for phylogenetic analyses of community structure. *Ecological Monographs*, **81**, 511–525.
- Keith P. & Allardi J. (2007) *Atlas des poissons d'eau douce de France. Patrimoines Naturels*, vol 47. SPN/IEGB/MNHN, Paris.
- Kembel S.W., Cowan P.D., Helmus M.R., Cornwell W.K., Morlon H., Ackerly D.D. et al. (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, **26**, 1463–1464.
- Laliberté E. & Legendre P. (2010) A distance-based framework for measuring functional diversity from multiple traits. *Ecology*, **91**, 299–305.
- Lavin S.R., Karasov W.H., Ives A.R., Middleton K.M. & Garland T. (2008) Morphometrics of the avian small intestine compared with that of nonflying mammals: a phylogenetic approach. *Physiological and Biomedical Zoology*, **81**, 526–550.
- Leibold M.A. & McPeck M.A. (2006) Coexistence of the niche and neutral perspectives in community ecology. *Ecology*, **87**, 1399–1410.
- Losos J.B. (2008) Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology Letters*, **11**, 995–1003.

- Mayfield M.M. & Levine J.M. (2010) Opposing effects of competitive exclusion on the phylogenetic structure of competition. *Ecology Letters*, **13**, 1085–1093.
- Mohseni O. & Stephan H.G. (1999) Stream temperature - air temperature relationship: a physical interpretation. *Journal of Hydrology*, **218**, 128–141.
- Montgomery D.R. (1999) Process domains and the river continuum. *Journal of the American Water Resources Association*, **35**, 397–410.
- Murray K. & Conner M.M. (2009) Methods to quantify variable importance: implications for the analysis of noisy ecological data. *Ecology*, **90**, 348–355.
- New M., Lister D., Hulme M. & Makin I. (2002) A high-resolution data set of surface climate over global land areas. *Climate Research*, **21**, 1–25.
- Newton R.J., Jones S.E., Helmus M.R. & McMahon K.D. (2007) Phylogenetic ecology of the freshwater Actinobacteria acI lineage. *Applied and Environmental Microbiology*, **73**, 7169–7176.
- Oberdorff T., Huguény B., Compagnon A. & Belkessam D. (1998) Non-interactive fish communities in the coastal streams of north-western France. *Journal of Animal Ecology*, **67**, 472–484.
- Peres-Neto P. (2004) Patterns in the co-occurrence of fish species in streams: the role of site suitability, morphology and phylogeny versus species interactions. *Oecologia*, **140**, 352–360.
- Persat H. & Keith P. (1997) La répartition des poissons d'eau douce en France: qui est autochtone et qui ne l'est pas? *Bulletin Français de la Pêche et de la Pisciculture*, **344** (345), 15–32.
- Posada D. & Buckley T.R. (2004) Model selection and model averaging in phylogenetics: advantages of akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology*, **53**, 793–808.
- Rahel F.J. & Hubert W.A. (1991) Fish assemblages and habitat gradients in a rocky-mountain great-plains stream - biotic zonation and additive patterns of community change. *Transactions of the American Fisheries Society*, **120**, 319–332.
- Reyjol Y., Huguény B., Pont D., Bianco P.G., Beier U., Caiola N. *et al.* (2007) Patterns in species richness and endemism of European freshwater fish. *Global Ecology and Biogeography*, **16**, 65–75.
- Ricklefs R.E. (1987) Community diversity - relative roles of local and regional processes. *Science*, **235**, 167–171.
- Ricklefs R.E. (2004) A comprehensive framework for global patterns in biodiversity. *Ecology Letters*, **7**, 1–15.
- Ricklefs R.E. & Miles B. (1994) *Ecological and Evolutionary Inferences from Morphology: An Ecological Perspective*. *Ecological Morphology*, pp. 13–41. University of Chicago Press, Chicago, IL.
- Rosenzweig M.L. & Ziv Y. (1999) The echo pattern of species diversity: pattern and processes. *Ecography*, **22**, 614–628.
- Stubbs W.J. & Bastow Wilson J. (2004) Evidence for limiting similarity in a sand dune community. *Journal of Ecology*, **92**, 557–567.
- Swofford D.L. (1993) Paup - a computer-program for phylogenetic inference using maximum parsimony. *Journal of General Physiology*, **102**, A9.
- Thompson J.D., Higgins D.G. & Gibson T.J. (1994) Clustal-W - improving the sensitivity of progressive multiple sequences alignment through sequences weighting position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.
- Valiente-Banuet A. & Verdu M. (2007) Facilitation can increase the phylogenetic diversity of plant communities. *Ecology Letters*, **10**, 1029–1036.
- Vamosi S.M., Heard S.B., Vamosi J.C. & Webb C.O. (2009) Emerging patterns in the comparative analysis of phylogenetic community structure. *Molecular Ecology*, **18**, 572–592.
- Vannote R.L., Minshall G.W., Cummins K.W., Sedell J.R. & Cushing C.E. (1980) River continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences*, **37**, 130–137.
- Villéger S., Novack-Gottshall P.M. & Moullot D. (2011) The multidimensionality of the niche reveals functional diversity changes in benthic marine biotas across geological time. *Ecology Letters*, **14**, 561–568.
- Webb C.O., Ackerly D.D., McPeck M.A. & Donoghue M.J. (2002) Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, **33**, 475–505.
- Wiens J.J., Graham C.H., Moen D.S., Smith S.A. & Reeder T.W. (2006) Evolutionary and ecological causes of the latitudinal diversity gradient in hylid frogs: treefrog trees unearth the roots of high tropical diversity. *American Naturalist*, **168**, 579–596.

## Supporting Information

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

**Table S1** Description of the PCA used to summarise the 17 ecological/life history traits.

**Table S2** Phylogenetic signal in ecological traits.

**Appendix S1** Description of the standardised measures of trait evenness.

(Manuscript accepted 29 October 2013)