Biodiversity and Conservation
https://doi.org/10.1007/s10531-025-03046-5

ORIGINAL RESEARCH

Rare fish species in European lakes – patterns and processes

Thomas Mehner[1] · **Christine Argillier**[2] · **Teresa Ferreira**[3] · **Kerstin Holmgren**[4] · **Erik Jeppesen**[5,6,7,8] · **Fiona Kelly**[9] · **Teet Krause**[10] · **Mikko Olin**[11] · **Pietro Volta**[12] · **Ian J. Winfield**[13] · **Sandra Brucet**[14,15]

Received: 11 November 2024 / Revised: 21 February 2025 / Accepted: 24 February 2025
© The Author(s) 2025

Abstract

According to differences in life history and species niches, community assembly processes are predicted to differ between common and rare species. While neutral processes, in particular dispersal, should contribute strongly to assembly of rare species, environmental filtering (species sorting) should dominate the community assembly of common species. We analysed commonness and rarity by occurrence, abundance and geographic range among 82 fish species in 1871 European lakes and reservoirs. The common 25% (21 species) were excluded, and structure and assembly processes of the 75% rare (61 species) species occurring in 348 lakes were evaluated. General linear latent variable models indicated that environmental variables related to lake type (lakes vs. reservoirs), climate and ecosystem size predicted a large proportion of variance for both rare species presence/absence and abundance models, while spatial variables (co-occurrence in watersheds) contributed little to the models. To link community structure with assembly processes, the contribution of seven fish traits related to life history, feeding and habitat preference to predicting the geographic range of the rare species was analysed by boosted regression trees. Intermediate average fecundity and high maximum body length of species predicted smaller geographic range and hence a higher level of rarity, but the response curves of the variables were mostly non-linear and difficult to interpret. In contrast to our assumption, the results of the dominant assembly processes and predictors for rare species were very similar to those predicted for the common fish species in lakes at the continental scale.

Keywords Dispersal · Environmental filtering · Community assembly · Stochasticity · Fish traits

---

Communicated by Michael Joy.

Extended author information available on the last page of the article

Published online: 01 March 2025

Introduction

Macroecology as a discipline investigates the relationship between patterns of species distributions and the underlying processes forming the respective communities (McGill 2019). As an example, geographic patterns in community composition can be caused by several processes, including dispersal, environmental filtering and species interactions (Stegen et al. 2013; Chang and HilleRisLambers 2016). Another class of process contributing to forming local communities is contingency, for example evolutionary history, stochasticity or past glaciations that affect patterns of species occurrences. By adopting the terminology from evolutionary biology, a conceptual synthesis suggested that the community composition of species can be influenced by selection, drift, speciation and dispersal (Vellend 2010). Selection represents deterministic differences in fitness among species, drift represents stochastic changes in species abundance, speciation creates new species, and dispersal is the movement of organisms across space. However, since these processes can usually not be observed directly in natural ecosystems, the relative importance of the processes has to be inferred from patterns of community composition along environmental, spatial and temporal gradients (Crisfield et al. 2024).

Many studies on community assembly processes focus on the occurrence and distribution of common species (Cottenie and De Meester 2003; Mehner et al. 2021a). These highly frequent species can be found in many locations, often have high local abundances and are easy to detect and capture by monitoring approaches, such that mapping their occurrences creates robust patterns (Gaston 2011). Accordingly, the spatial patterns of species richness are predominantly caused by common species (Lennon et al. 2004). In contrast, frequently, the majority of species in a given region can be considered rare, having low occurrence, narrow range and low local abundances (Magurran and Henderson 2003; Callaghan et al. 2023). However, rare species make significant contributions to ecosystem functions and processes (Mouillot et al. 2013; Dee et al. 2019) and are often the focus of conservation activities (Eken et al. 2004) because they are particularly vulnerable to rapid declines or even extirpations in response to anthropogenic stressors. If rare species can be identified and classified properly (Maciel 2021), effective measures targeted at protecting these species from extirpation can be introduced (Kunin and Gaston 1993; Prendergast et al. 1993; Pritt and Frimpong 2010).

A few studies have tested the link between the patterns of commonness and rarity of species and the dominant processes contributing to community assembly. General hypotheses are that the occurrence of common species is driven by niche-based processes and high dispersal and colonisation rates, while rare species are more affected by neutral processes (stochastic drift of population parameters) and a limited colonisation potential due to the low number of propagules (Santoul et al. 2005; Siqueira et al. 2012; van Galen et al. 2023; Iop et al. 2024). Furthermore, large-scale spatial distribution often differs between rare and common species (Prendergast et al. 1993; Jetz and Rahbek 2002), presumably caused by a need for specialised habitats and an overall lower reproductive capacity or poorer dispersal abilities of rare species (Gaston and Lawton 1990; Kunin and Gaston 1993; Magurran and Henderson 2003; Lennon et al. 2004).

However, the definition of rarity is not straightforward. One of the most comprehensive classification schemes considering rarity as a multidimensional phenomenon was provided by Rabinowitz (1981), who distinguished between three rarity dimensions – local abundance, habitat specificity and geographic range – and classified species by combining the three dichotomised dimensions in eight possible groups. According to this scheme, only generalists with large local abundance and wide geographic range would be considered common, while the seven other combinations of the three dimensions would indicate varying forms of rarity. Although the scheme has been applied in several studies (Harnik et al. 2012; Reed et al. 2020; Sykes et al. 2020), it has also been criticised because it classifies rarity based on patterns but does not elucidate the underlying processes generating these patterns (Crisfield et al. 2024). To overcome this limitation, a conceptual improvement of rarity dimensions has recently been suggested, which addresses several limitations of previous concepts (Crisfield et al. 2024). By considering the three rarity dimensions, local abundance, occupancy and geographic range, as patterns along an increasing spatial scale, inference of underlying processes is built on the link between the functional traits of species and community assembly processes. In particular, occupancy and range size are assumed to be reflected by traits related to movement and niche breadth (tolerance) (Crisfield et al. 2024).

We apply this recent concept to define rarity based on the dimensions of local abundance, occupancy and geographic range to study the rarity patterns of fish species in European lakes. We evaluate the patterns of community composition and species distribution in 1871 lakes and define commonness or rarity of the species according to rankings along these three dimensions. Model-based inference on niche-based or dispersal-related assembly processes based on the distribution patterns of all fish species would reflect the processes driving the assembly of common species because of their strong numerical dominance in the dataset. Therefore, the subsequent analytical steps excluded the distribution data of the most common species and focused on local abundance, occupancy and geographic range of only the rare species. To infer processes from the patterns, we used multivariate latent variable models to identify environmental and spatial predictors for rare species occurrence and local abundance that would suggest environmental filtering and dispersal limitation. Finally, we evaluated the contribution of traits related to movement, tolerance and life history to the rank of geographic range and abundance patterns among common and rare fish species combined. By this approach, we intend to link processes of community assembly to patterns of rarity and to compare, in the discussion, the patterns and drivers of rarity of fishes with those found for 19 common fish species in the same set of European lakes (Mehner et al. 2021a).

Materials and methods

Field sampling and abiotic predictors

We used a dataset of fish communities in 1943 European lakes and reservoirs accumulated from standardised fishing by multi-mesh gillnets used for ecological classification according to the European Water Framework Directive (Brucet et al. 2013; Mehner et al. 2017). Details of background, methods and basic fish community structure have been summarised earlier (Argillier et al. 2013; Brucet et al. 2013; Mehner et al. 2017). The lakes were sampled between 1995 and 2012. If more than one survey was conducted in a lake, we used the most recent survey. In short, the fishing effort (number of nets per lake) was standardised according to lake area and depth, and fishing was performed using 30 m long and 1.5 m high benthic multi-mesh gillnets with 12 mesh sizes in a geometric row between 5 mm and 55 mm (CEN 2015). The nets were placed at the lake bottom. In lakes with a maximum depth larger than 6 to 10 m, additional pelagic gillnets were used, being either 1.5 m, 3 m or 6 m high and with the same mesh sizes as the benthic ones (the 5 mm mesh section being omitted from the 3 m and 6 m nets). The catch data were converted into a presence/absence (p/a) matrix per fish species and lake. Relative abundances per fish species and lake were expressed as catch per unit effort (CPUE, individuals per net and night). If both benthic and pelagic nets were applied in a lake, CPUE was the arithmetic average of catches from both habitats. Arithmetic averaging was applied to weight the abundances of pelagic species appropriately, which are often underestimated by the low fishing effort by pelagic in comparison with benthic gillnets (Alexander et al. 2015).

Complete information useful for the purpose of our study was available from a total of 1871 locations in 13 European countries, i.e. Denmark, Estonia, Finland, France, Germany, Italy, Norway, Portugal, Republic of Ireland, Slovenia, Spain, Sweden and United Kingdom. Among these lakes, 1755 were natural and 116 were artificial (mostly reservoirs, some gravel pits, Fig. 1). Therefore, we included the lake type (natural or artificial) as a predictor in our analyses, to evaluate whether naturally assembled communities differ from those assembled by the artificial creation of lakes. We added another five environmental predictors, namely altitude (m above sea level), minimum and maximum annual air temperature (°C) at the lake location, lake area (km^2) and lake maximum depth (m) (overview Supplementary Information Table S1), taken from national databases. Air temperature was calculated using the Climatic Research Unit (CRU) model (New et al. 2002) based on temperature records for the years before 2008, thus matching the period when the lakes were sampled. The model we applied had a spatial resolution of 10’ latitude and (or) longitude and considered elevational differences between stations (New et al. 2002). The dataset contained only lakes with $pH>6$ to exclude the potential effect of anthropogenically induced acidification on fish communities. Originally acidified lakes with current $pH>6$ because of mitigation by regular liming were included because the first liming generally started 5–30 years before the present fish sampling (Holmgren et al. 2016). Information on the productivity of lakes estimated as total phosphorus (TP) concentration (mg m^{-3}, obtained from at least four samples across the seasons per year) was available only for 1109 out of the 1871 lakes (Supplementary Information Table S1). Accordingly, we did not include TP concentration as a predictor in the subsequent analyses.

Rarity ranking of species

We calculated three metrics of rarity for the fish species in the 1871 lakes. Rarity by occurrence was evaluated by counting the number of lakes in which a certain species was caught (Jeliazkov et al. 2022). We used this metric as an approximation to occupancy defined as the occupied proportion of habitat patches within a species’ range (Crisfield et al. 2024). This definition of occupancy based on observed species ranges creates problems for species occurring only once since they would have an occupancy of 100% and a geographic range of zero. Because the real (instead of observed) ranges of all species are unknown, we consider the geographic extent covered by the sampled lakes as an identical range for all species; hence, the species occurrences in the set of 1871 lakes was considered an estimate of proportional occupancy. Rarity by abundance was calculated from the species-specific

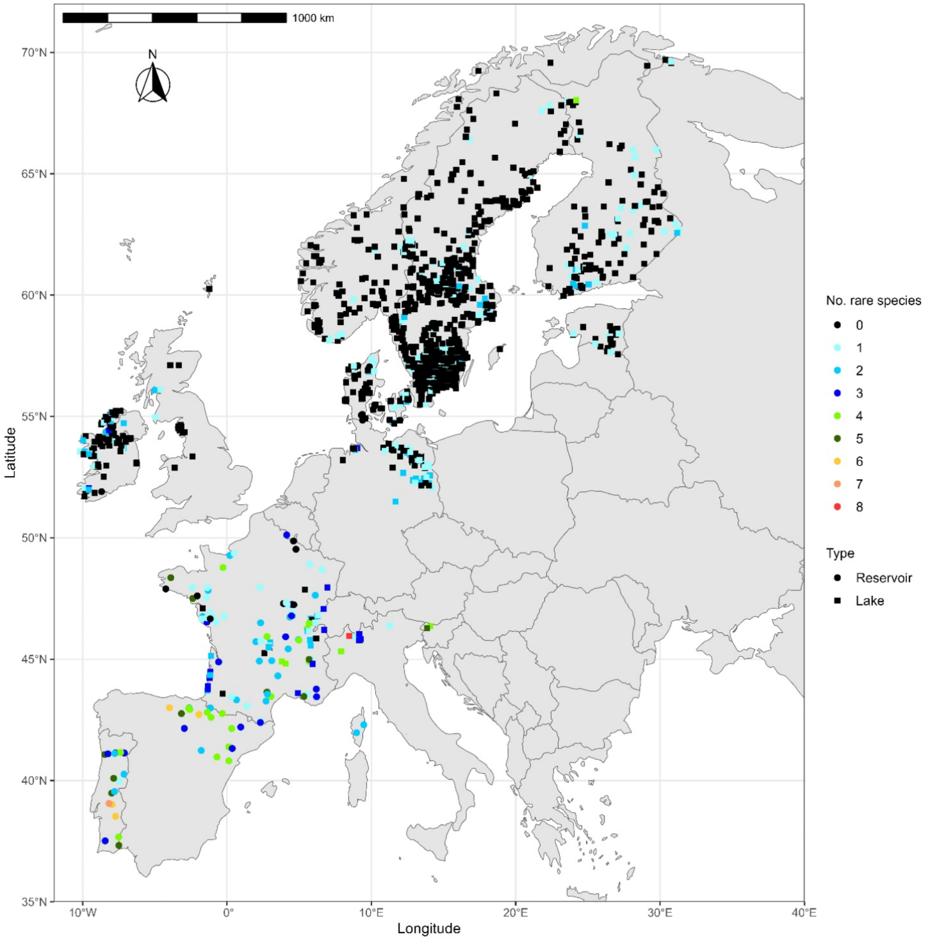

Fig. 1 Map of Europe with an overview of the geographic location of the 1871 sampled natural lakes (quadrat, $n = 1755$) and reservoirs (dots, $n = 116$) and the number of rare fish species per location (occurrence in $n = 348$ locations) indicated by the colour scale

CPUE per lake by summing CPUE per species per lake across all lakes. This approach gives the same ranking among species as when calculating the arithmetic average CPUE among all lakes in which the species was present. Rarity by geographic range was evaluated from the convex hull (km²) of decimal latitude and longitude of occurrence data per species by using the R-package 'GeoRange' version 0.1.0. (https://CRAN.R-project.org/package=GeoRange). In this way, the geographic range estimated per species is smaller than the geographic extent of the sampled lakes because it is based on observed occurrences.

We ranked the 82 species (see results) in the three rarity matrices (occurrence, abundance, range) by giving the most common species the highest ranks, such that rare species had low ranks. We then summed the ranks per species across the three matrices and ranked the 82 species to obtain a combined rarity rank. The 21 (25% of 82 species) species with the highest rank sums were considered common (Magurran 2004), while the remaining 61 species (75%) were considered rare in the subsequent analyses (Siqueira et al. 2012; Iop et al. 2024). More strict boundaries to define rare species have been applied and discussed (e.g., the 25% least frequent species) (Gaston and Lawton 1990); however, the frequency distribution of species occurrences (see results) indicates that of the 21 least frequent species in our dataset each occurred in only one or two lakes, and hence these patterns would be inappropriate for the intended inference of underlying community assembly processes.

Exploring environmental and spatial effects by latent variable models

We created matrices of occurrence and abundance for the 61 rare species and excluded the 1523 lakes in which only common species were found. We analysed the effect of the six environmental predictors on the occurrence and abundance of the 61 rare species in the remaining 348 lakes (290 natural lakes, 98 artificial lakes, mainly reservoirs, Supplementary Information Table S1) by multivariate generalized linear latent variable models (gllvm) using the R-package 'gllvm', which is based on a maximum likelihood framework (Niku et al. 2019). The five continuous predictors were centred and scaled. Our previous work on abiotic predictors indicated that (log)linear relationships properly describe the effects on fish communities (Mehner et al. 2005, 2007, 2016); hence, we did not include unimodal (quadratic) terms. For the occurrence data, we applied a binomial distribution with three latent variables and Laplace approximation. For the abundance data, we used tweedie distributions with power = 1.4 and two latent variables by applying the enhanced variation approximation method. The number of latent variables and the underlying distributions including the power exponent of tweedie distributions were optimised by testing several models and by checking convergence, Akaike Information Criteria (AIC) and residual diagnostics among the model results.

To account for the effects of spatial distribution of lakes within watersheds and thus to address potential dispersal limitation, we assumed that communities of rare species in lakes within the same river basin were more similar than rare fish communities in lakes from different river basins, reflecting stronger dispersal within than among river basins. We recorded the river basins to which the lakes belong based on the geographical coordinates of lakes and river basins. These data were obtained from Catchment Characterization and Modelling (CCM) data version 2.1 provided by EU Joint Research Centre (JRC, http://ccm.jrc.ec.europa.eu/php/index.php?action=view%26id=23). The spatial correlation within river basins for the gllvm was accounted for by adding random row effects per river basin to the model, analogous to a random intercept in mixed models. This approach was similar to the methods we applied when studying the common fish species in European lakes (Mehner et al. 2021a).

For both occurrence and abundance matrices, we created four hierarchical gllv models. For basic comparison, we ran a model without environmental or spatial predictors, equivalent to an unconstrained ordination. Then we added either environmental or spatial predictors separately to the unconstrained ordination. The fourth model was run by combining spatial and environmental predictors. The most parsimonious model was selected according to the lowest Akaike Information Criterion (AIC). Furthermore, we compared the reduction in trace of the residual covariance matrix of the models with predictors relative to the trace of the unconstrained ordination models to evaluate the amount of variation accounted for by the environmental and spatial predictors ('gllvm' vignette, https://cran.r-project.org/web/packages/gllvm/vignettes/vignette1.html). We plotted the residual diagnostics of the finally selected models (residuals against linear predictors of fitted values, a Normal Q-Q plot of residuals with a simulated point-wise 95% confidence interval envelope, residuals against row index and column index and scale-location plot) to inspect how appropriately the models fit the datasets. Plots of environmental model coefficients were created to check which predictor had a significant positive or negative effect on the 61 rare fish species. All analyses were conducted in R version 4.2.2 (R Development Core Team 2022).

Effect of fish traits on rarity dimensions

We compiled information on maximum fish length ($\log_{10}$, in cm), maximum fecundity ($\log_{10}$ number of eggs per female), age at maturity (years) and average trophic level (levels 2–5 reflecting the position in the trophic pyramid, with high levels indicative of top predators) from FishBase (Froese and Pauly 2022) by the ‘rfishbase’ package (Boettiger et al. 2012). If more than one value was available, arithmetic averages were calculated. We added habitat preference for either estuaries or rivers and streams as binary variables (1 = preferred, 0 = non-preferred) to find out whether the geographical range of species was related to preference of freshwater habitats other than lakes. Finally, we added the native status for the 82 species, coding non-native = 1 for all species alien (but not translocated) to the countries included, with coding = 0 otherwise (Trochine et al. 2018). These seven traits refer to tolerance (habitat specificity, trophic level, native status), movement (size) and life history (fecundity, age at maturity) and are assumed to link processes that form rarity with rarity patterns (Crisfield et al. 2024).

To infer whether these traits predict range size (Crisfield et al. 2024), the rank of all common and rare species with respect to the convex hull area of geographical range was used as the response variable in boosted regression trees (BRTs), while the seven variables characterising trait variability of fish species were used as predictors. Boosted regression trees are robust to combinations of categorical and continuous data and interactions between variables, and they are not affected by the distribution of the data (De’ath 2007). The approach of Elith et al. (2008) and the stepwise procedure provided in the tutorial by Elith & Leathwick (https://rspatial.org/raster/sdm/9_sdm_brt.html) in combination with the r-packages ‘gbm’ (Ridgeway 2006) and ‘dismo’ (http://rspatial.org/sdm/) were employed to find the optimal number of trees. Tree complexity was set at three with learning rates of 0.001, and with the bag fraction set at 0.75, meaning that each individual tree was constructed using 75% of the data (Elith et al. 2008). Several other combinations of bag fraction, learning rate and tree complexity were tested, but they had a poorer performance than the ones combined here. Partial dependence plots of fitted function versus observed values for variables significantly predicting the response variables were prepared, seeking to present the influence uniquely attributable to a single predictor. This procedure was repeated with the rank of 82 fish species abundances as response variable.

Graphical outputs of results was created by the R packages ‘ggplot 2’ (Wickham 2016), ‘corrplot 0.92’ (Wei and Simko 2021) and ‘rnaturalearth 0.3.1’ (Massicotte and South 2023) in R version 4.2.2 (R Development Core Team 2022).

Results

In total, we found 85 fish species or genera in the 1871 lakes. We excluded cyprinid hybrids and removed a few entries where only genus level was mentioned (*Abramis* spp., *Cottus* spp.), the total dataset thus encompassing 82 species (Supplementary Information Table S2). *Salmo trutta* included both potamodromous and anadromous forms of brown trout. *Coregonus lavaretus* was used as a replacement species to combine all lacustrine whitefish species of the Baltic and North Sea basins, in part because of unclear taxonomy (Kottelat and Freyhof 2007; Mehner et al. 2018), which often leads to incorrect or incomplete species determination in fish surveys. Fish grouped as Arctic charr (*Salvelinus alpinus*) may, in some Alpine lakes, correctly be determined as Alpine charr (*Salvelinus umbla*), but in the vast majority of lakes included here, occurrence of Alpine charr can be excluded. *Alosa fallax* from the Irish Loch Leane is named either as subspecies *Alosa fallax killarnensis* or as *Alosa killarnensis*; we use *A. killarnensis* in all graphical output here.

Among the 82 species, only six, i.e. Eurasian perch (*Perca fluviatilis*), roach (*Rutilus rutilus*), pike (*Esox lucius*), ruffe (*Gymnocephalus cernuus*), bream (*Abramis brama*) and rudd (*Scardinius erythrophthalmus*), occurred in >25% of the lakes (Supplementary Information Figs. S3a). 21 species (25% of the 82 species) were considered common by occurrence, and hence the other 61 (75%) species were considered rare. The range of occurrence for each rare species was one to a maximum of 44 lakes, suggesting that none of the 61 rare species occurred in more than 2.3% of the 1871 lakes.

Species with common occurrence also had the highest overall abundance (Supplementary Information Figs. S3b). The most abundant species, perch, had a mean abundance sum of 124,600 individuals per net and night (catch per unit effort per lake summed over all 1871 lakes), while the 61 rare species had sums of abundances (numbers per net and night) <~200. The geographical range, expressed as convex hull (CH, km²) of occurrences in the 1871 lakes, grouped almost the same 21 species (25%) as common (Supplementary Information Figs. S3c) as found for commonness by occurrence and by abundance (Supplementary Information Figs. S3a, b). The 61 rare species had CH-values $<1.7\times10^6$ km², while brown trout (*Salmo trutta*) was the species with the largest $\mathrm{CH}=4.7\times10^6$ km². Accordingly, the ranks of the 82 species in the three matrices of commonness or rarity were strongly correlated to each other (Spearman rank correlation, all $P<0.001$; Occurrence-Abundance $r=0.86$; Occurrence-Geographical Range: $r=0.96$; Abundance-Geographical Range: $r=0.78$). After summing the rank numbers per species from the three matrices (Fig. 2), we identified 21 common (rank sum > 182) and 61 rare species with rank sums < 182. These 61 species formed the matrices of rare species for the subsequent calculations. The 61 rare species occurred only in 348 of 1871 lakes (19%), with the numbers of rare species per lake varying between 1 and 8 (Fig. 1). The lakes with rare species were more evenly distributed over the sampled area than the lakes of the full dataset, which were dominantly located in Sweden (Fig. 1). However, there was a tendency for higher numbers of rare species in lakes of the southern part of the continent.

The 348 lakes with rare species were distributed among 135 European watersheds. Only a few watersheds contained more than 10 lakes, i.e. Rivers Elbe ($n=31$), Loire ($n=21$), Rhone ($n=22$), Ebro ($n=18$), Neva ($n=11$) and Dalälven ($n=18$). In contrast, in 89 watersheds only one lake was present.

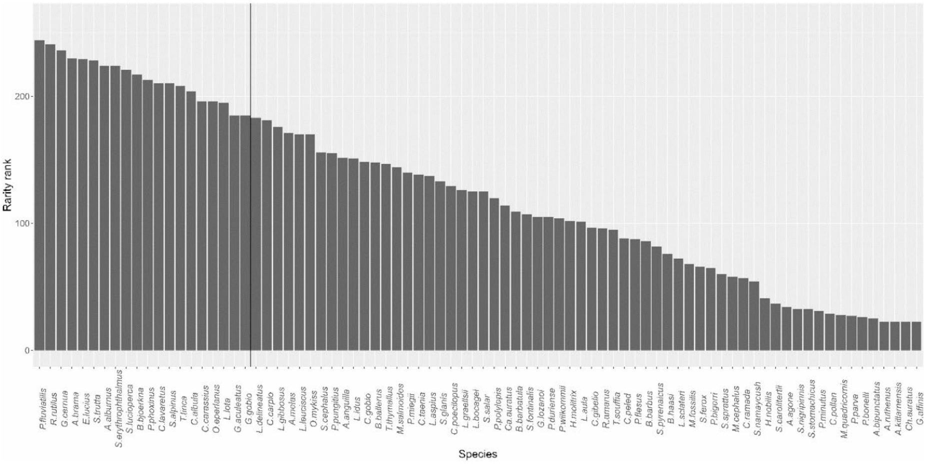

Fig. 2 Frequency distributions for 82 fish species in 1871 European lakes, indicating the rank sum obtained from the three frequency distributions of occurrence, abundance and range (Supplementary Information Figs. S3). Species with high rank sums are common according to all three metrics. The vertical line separates the 25% common species ($n=21$, left) from the 75% ($n=61$, right) rare species, according to the rank sums

Table 1 Overview of generalized linear latent variables models (gllvm) predicting the occurrence or abundance of 61 rare fish species in 348 European lakes. For both datasets, unconstrained models without predictors are compared with models including six environmental predictors only, including spatial predictors (common occurrence in watersheds) only or combining environmental and spatial predictors. AIC = Akaike information criterion, trace = trace of the residual covariance matrix, logll = log likelihood, resdf = residual degrees of freedom

Model	AIC	Trace	logLL	ResDF
Occurrence				
No predictors	4742	730	-2129	20,987
Environmental	4444	274	-1615	20,621
Spatial	4745	1166	-2131	20,986
Environmental + spatial	4446	277	-1615	20,620
Abundance				
No predictors	5479	6244	-2497	20,985
Environmental	4813	527	-1798	20,619
Spatial	5488	5479	-2500	20,984
Environmental + spatial	4996	1297	-1888	20,618

The hierarchical generalized latent variable models for rare species occurrence and abundance in the 348 lakes showed similar patterns. The most parsimonious models according to the lowest AIC were the one including only environmental predictors, while the contribution of spatial predictors alone and with environmental predictors resulted in poorer performance (Table 1). For the occurrence models, the trace of the residual correlation matrix was lowest for the models with environmental predictors only (Table 1), predicting 62% of the variation relative to the unconstrained ordination without predictors (1-300/797). For rare species abundance, the reduction of trace relative to the unconstrained model was likewise highest for the model including only environmental predictors (1-555/6244 = 91.1%). Accordingly, we show only the results of the models with environmental predictors.

The model fit was appropriate for both occurrence and abundance models, as indicated by the residual diagnostic plots (Supplementary Information Figs. S4, S5). Among the six environmental predictors for rare species occurrences, lake type (lake vs. reservoir) was the strongest one by significantly predicting the occurrence of 41 of the 61 rare species (Fig. 3). Altitude (significant for 20 species), minimum (20 species) and maximum (17 species) temperature and lake maximum depth (15 species) were likewise partially important (Fig. 3), while lake area (10 species) was only marginally important. The same six environmental predictors were significant for the abundances of a substantially larger number of species (Fig. 4) as those found in the occurrence patterns (Fig. 3) and had similar rank of importance. Lake type was significant for 55 species, followed by maximum (55 species) and minimum temperature (49 species) and altitude (47 species), while maximum lake depth (35 species) and lake area (40 species) were slightly less important (Fig. 4).

The boosted regression model (number of trees = 10,000, interaction depth = 1, learning rate = 0.001, bag fraction = 0.75, tree complexity = 3, Poisson error distribution) included all seven fish traits to predict the rank of geographical range size (convex hull) among the total of 82 fish species (Supplementary Information Fig. S6). The overall model performance with respect to reduction of deviance by adding predictors was 70.7%, calculated from residual deviance (rd = 4.14) and total deviance (td = 14.1) as 1-(rd/td) (Yu et al. 2020). Fish fecundity was the most dominant predictor (31.5%) for species range, followed by maximum length (26.8%) and trophic level (21.3%). The other four predictors contributed < 10% (Fig. 5). Intermediate fecundity, large size, and low trophic level decreased the rank of the species (equivalent to smaller geographic range and higher rarity) (Fig. 5). However, the response curves to continuous predictors were, in part, non-linear and did not provide clear tendencies of the effect directions (Fig. 5). When running identical boosted regressions trees with the rank of species-specific abundances (Supplementary Information Fig. S3 b), the reduction of deviance by the seven predictors relative to a null model was marginal (< 1%), even with wide variations in the parameter of the gbm.step function.

Discussion

We aimed at inferring the processes driving the rarity of fish species from their distribution and abundance patterns in 1871 European lakes. We found strong correspondence between occurrence, abundance and geographic range of 82 fish species, suggesting that these three axes of rarity are driven by similar processes, with little among-species variability. The occurrence and abundance of the 61 rare species (75% of total richness) in 348 lakes and reservoirs were primarily predicted by the lake type (natural vs. artificial) and by climatic variables, in particular by altitude and minimum and maximum temperature at lake location. We did not find strong evidence for the effect of dispersal limitation being a driving process because models including random terms for spatial similarity (occurrence in the same watersheds) did not perform better than models with environmental predictors only. The geographic range of all species was primarily related to life history traits, but many responses were non-linear and therefore not easy to interpret. In contrast, we did not find any trait that predicted the rank of average abundance of fish species in the lakes. In com-

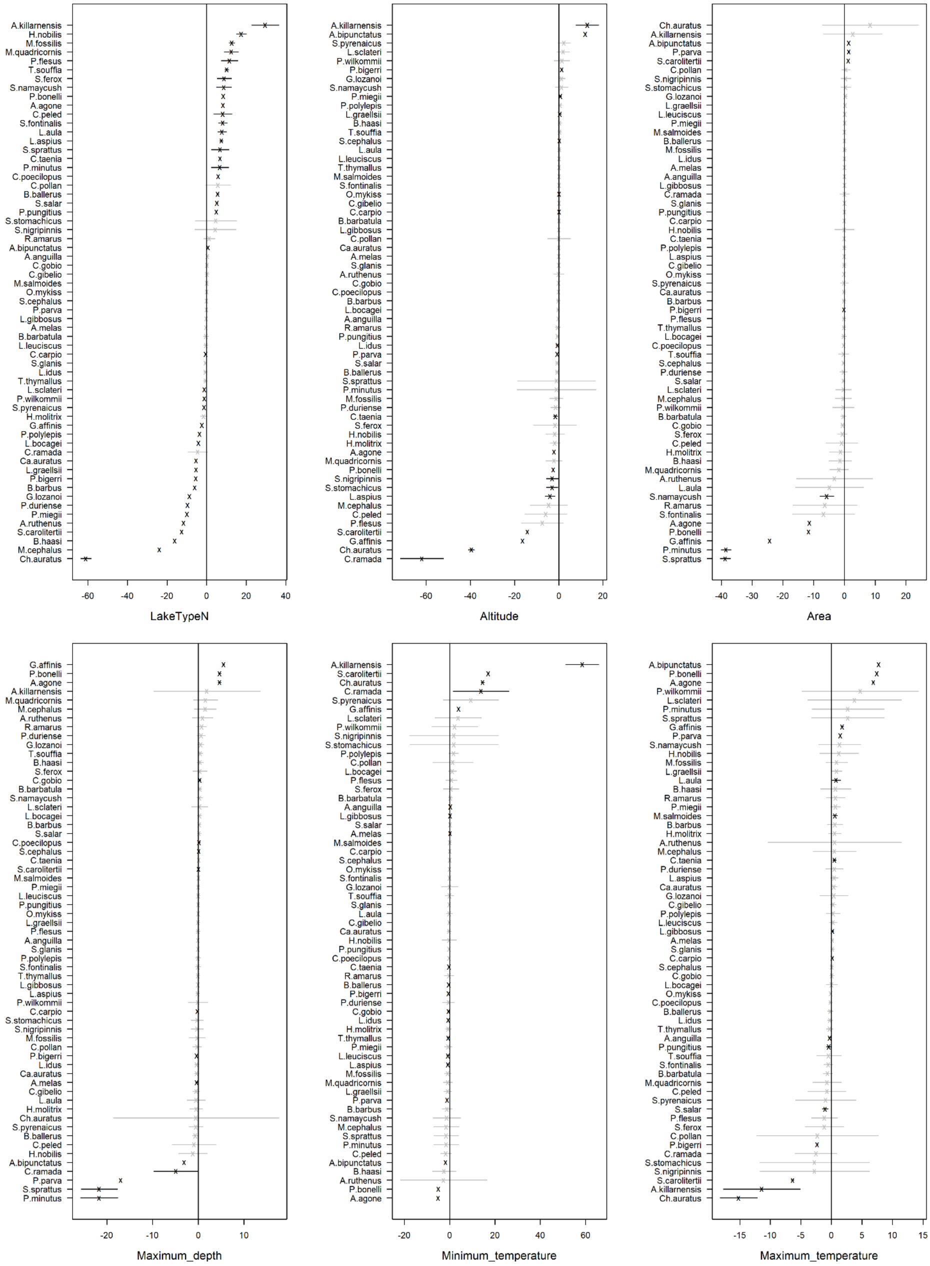

Fig. 3 Coefficient plot ± 95% confidence intervals of a generalized linear latent variables model to predict the occurrence of 61 rare fish species in 348 European lakes from six environmental predictors (lake type with natural lakes as reference, altitude, lake area, lake maximum depth, minimum temperature, maximum temperature). The order of species follows the strength of positive and negative predictions by the predictor. Black crosses and horizontal lines indicate significant predictions, while grey crosses with horizontal lines crossing the zero vertical line indicate insignificant predictors

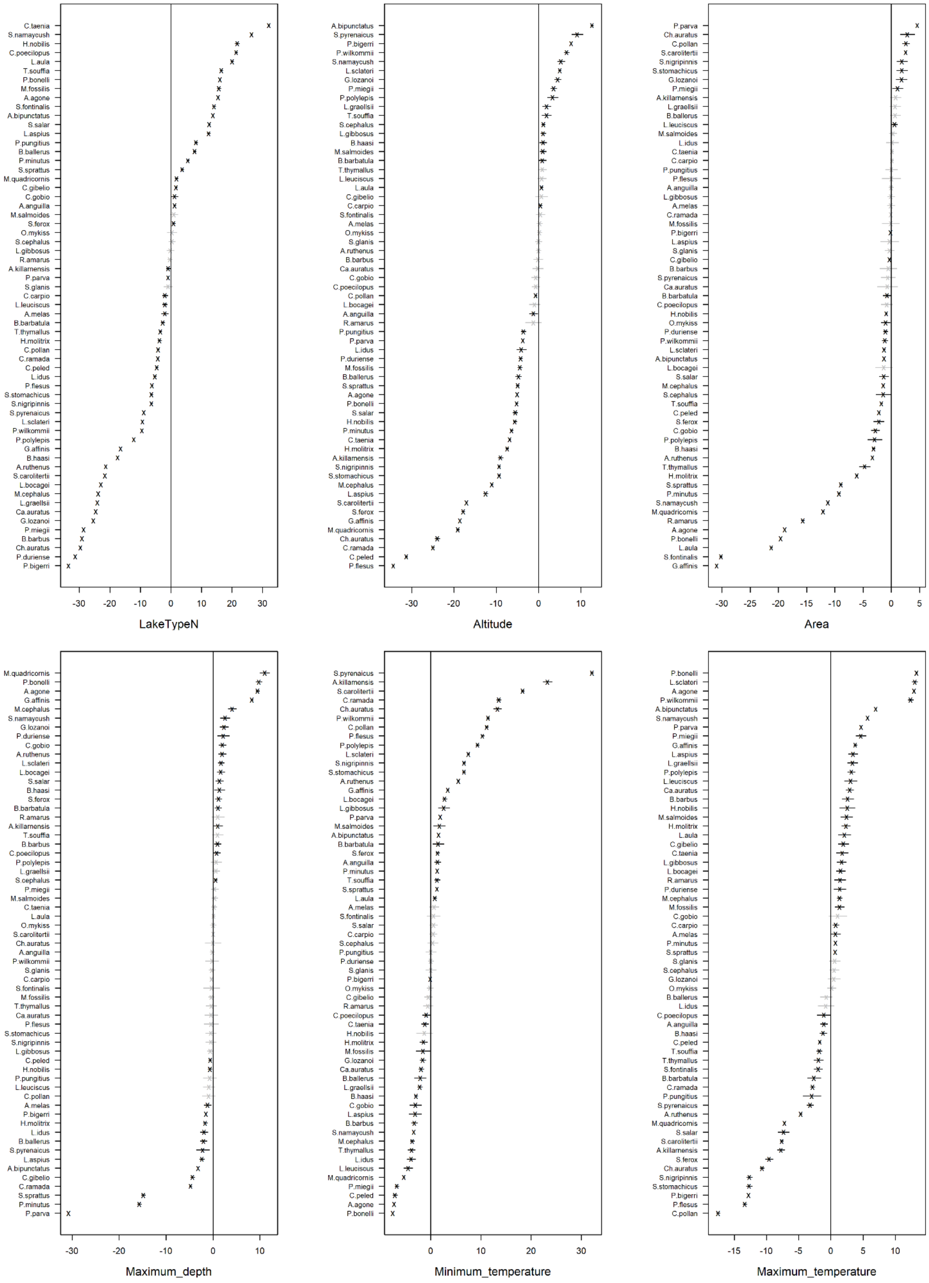

Fig. 4 Coefficient plot ± 95% confidence intervals of a generalized linear latent variables model to predict the abundance (catch per unit effort, individuals per net and night) of 61 rare fish species in 348 European lakes from six environmental predictors (lake type with natural lakes as reference, altitude, lake area, lake maximum depth, minimum temperature, maximum temperature). The order of species follows the strength of positive and negative predictions by the predictor. Black crosses and horizontal lines indicate significant predictions, while grey crosses with horizontal lines crossing the zero vertical line indicate insignificant predictions

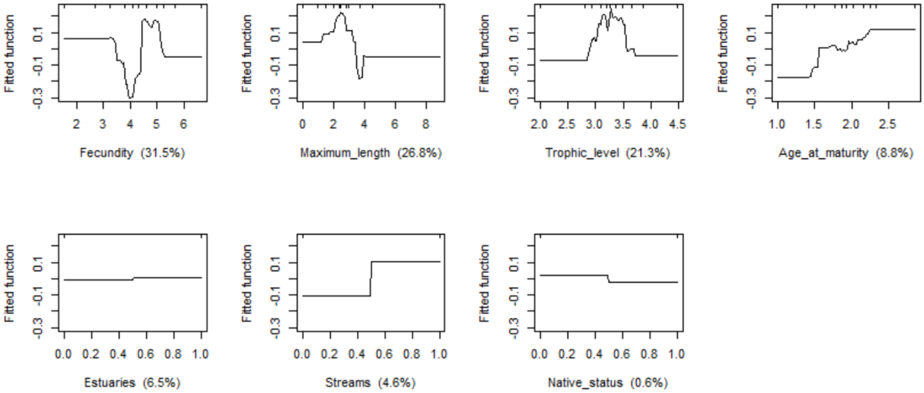

Fig. 5 Partial dependence plots showing the response of the rank of geographical range (based on convex hulls in km^2) among 82 fish species in European lakes to seven fish traits related to life history and ecology as fitted by boosted regression trees, namely $\log_{10}$ maximum length (mm), $\log_{10}$ fish fecundity (number of eggs), age at maturity (years), primary occurrence in streams (binary with 1 = presence, 0 = absence), fish trophic level (2–5), primary occurrence in estuaries (binary with 1 = presence, 0 = absence) and native status (1 = non-native, 0 = native). Higher levels of fitted functions indicate higher rank and hence greater convex hull areas of the respective species, suggesting a lower level of rarity with respect to geographical range. The lines per trait show proportional deviations in response to the trait above or below the average rank (zero) if the other traits are kept constant. For example, trophic level in the range of 3.0 to 3.5 increases the average rank (= 41) of convex hull by about 20% if all other traits are constant

bination with our earlier analyses of the community assembly processes of common fish species (Mehner et al. 2021a), we suggest that the drivers of commonness and rarity of fish species in European lakes at continental scale are largely similar and primarily deterministic. Historical contingencies such as glaciations may interact with contemporary effects from environmental filtering.

Methodological caveats

We used the three axes of rarity recently suggested by Crisfield et al. (2024), i.e. occupancy, abundance and geographic range of 82 fish species. Fish surveys to evaluate the ecological integrity of lakes by their fish communities for the European Water Framework Directive (Argillier et al. 2013; Ritterbusch et al. 2022) have been conducted in a highly standardised way among the countries included in this study. Despite some biases with respect to size and species selectivity of the gillnets used in the fishing surveys (Diekmann et al. 2005; Prchalova et al. 2009; Smejkal et al. 2015), occurrence and abundances matrices of species in the 1871 lakes can be considered as reliable for among-lake comparisons. Some uncertainty with respect to local occurrences is introduced by the fact that Anguilliformes (eel-like fishes) and small, strictly littoral species are often not representatively caught by gillnets, but their occurrence in the lake can be demonstrated by e.g. electrofishing (Diekmann et al. 2005; Smejkal et al. 2015). Electrofishing has not been applied in all countries, and the species list per lake is therefore only based on gillnet catches.

However, because of a completely different focus, the design of the fishing surveys according to the European Water Framework Directive does not match the conceptual guidelines for sampling rare species (Jeliazkov et al. 2022). Accordingly, we had to adapt suitable metrics for rarity from the available fishing survey data. We estimated occupancy by counting the occurrence of species in all 1871 lakes. This is not a perfect solution because the approach assumes that all species have identical geographic ranges. Alternatively, counting occurrences within the observed geographic range per species would have strongly biased the dataset because then many species occurring in a few lakes only would have achieved 100% occupancy if the lakes formed exact spatial polygons on the map. It was tempting to consider European bioregions (Abell et al. 2008) as potential ranges for the species dominantly occurring in only one of the regions. However, integrating occupancy as percentage of sites with the species present among regions would be possible only if lakes within the bioregions had been sampled in a systematic and stratified way, accounting for differing areas of the bioregions. Therefore, the definition of rarity by proportional occupancy recently suggested by Crisfield et al. (2024) requires datasets that are currently not yet freely accessible for fish species in European lakes.

The geographic range per species was calculated applying the occurrence data, and strong correlations among the rarity dimensions were found. This suggests that slightly variable rankings in one dimension would have a minor effect on the combined rank, and hence the list of rare species is likely very robust. To combine the three dimensions of rarity for defining rare species, we used the rank sums from the three matrices and defined common species as those belonging to the 25% quartile of the summed ranks, while the other 75% of the species were considered rare (Siqueira et al. 2012; Iop et al. 2024). Modifying this threshold would have reduced the total number of species considered rare. For example, using only the lower 25% quartile of rarity dimensions (21 species) to define species as rare (Gaston and Lawton 1990) would have led to an occurrence matrix with an extreme overdominance of zeroes because each of the 25% rarest species were observed only in one or two lakes. Even expanding the inclusion of species to the 50% threshold would still limit the occupancy of rare species to a maximum of 10 among the 1871 lakes (0.53%), with little power to infer any process from these sketchy patterns. Therefore, we considered all 61 species outside of the 25% quartile of the common species rare, still resulting in only 348 lakes with occurrence of any rare species and with occurrences per species that did not exceed 44 lakes. The remaining 1523 lakes without rare species were occupied by at least one of the 21 common species, which reflects the extreme dominance of a few common freshwater fish species in European lakes (Mehner et al. 2007; Argillier et al. 2013; Brucet et al. 2013; Martinsen et al. 2023). Admittedly, the list of the 61 rare species in lakes at the continental scale of Europe given here can only provide coarse information on the abundance and geographic range of species of low occurrence. Intercontinental comparisons of rare species' records to support conservation actions would require sampling surveys specifically designed to capture commonness and rarity of fishes (Jeliazkov et al. 2022), ideally based on entire watersheds (see discussion below). However, the fish species determined as rare according to the three rarity dimensions can be used to infer potentially similar environmental predictors and assembly processes.

Prediction of rarity by environment and spatial factors

The latent variable models on occurrence and abundance patterns of the 61 rare species defined similar predictors. Above all, lake type (natural vs. artificial) was the main predictor of the occurrence and abundance of rare species. However, the coefficients per species suggested that the response was not uniform. The occurrence and abundance of some species were positively predicted by lake type = artificial (mainly reservoirs; for example, the genera *Luciobarbus*, *Pseudochondrostoma*, *Barbus* or *Squalius*); however, other species were positively predicted by lake type = natural (for example, the genera *Coregonus*, *Salmo* and *Salvelinus*). Similarly, climatic variables such as minimum or maximum air temperature and altitude of lake location as proxies of lake thermal conditions had almost equal amounts of significant positive or negative predictions, suggesting that rare species cannot be defined by a certain thermal guild. Rarity was found both among stenotherm coldwater species, such as the genera *Coregonus*, *Salmo* and *Salvelinus*, and species of the warmwater guild, for instance many carp-like Cypriniformes. We can also exclude that rarity was caused by the size of the lake ecosystem. Neither lake area nor maximum depth were strong predictors of the abundance or occurrence of many rare species, and the significant coefficients were also a mix of positive and negative ones, suggesting again that there was no unique effect of ecosystem size on rarity. Accordingly, we can exclude richness and sampling effects (richness increases with lake area and hence more rare species can be expected and caught in larger lakes) on the rarity of fish species in lakes. The dominant environmental predictors found here correspond to those identified in previous studies on large-scale drivers of fish occurrences and abundances in lakes in Europe and North America (Jeppesen et al. 2000; Mehner et al. 2007; Brucet et al. 2013; MacDougall et al. 2018; Kadoya et al. 2024).

The spatial distribution and aggregation of lakes in European watersheds appeared to have only a marginal effect on the occurrence or abundance of rare species. We used random intercepts in the generalized linear models to test for greater similarity in rare species community composition within than among watersheds. Neither AIC nor the reduction in trace of the models suggested that spatial information was an important contributor to prediction of rare species abundance or occurrence. An alternative approach to account for spatial similarity is a geographic distance matrix among lakes. By using Euclidean distance matrices, areal dispersal would be assumed, which is not appropriate for fishes (Beisner et al. 2006; De Bie et al. 2012; Mehner et al. 2014). Waterway distances via hydraulic connections among lakes would be the superior approach, but such maps are not yet available at European spatial scale. Furthermore, with a few exceptions, the number of lakes per watershed was relatively low, which may have limited statistical inference on the comparison of within- and among-watershed similarities, i.e. we may have underestimated the effects of potential dispersal limitation on the rarity of fish species in lakes. Future developments in satellite-based geographic data extraction and algorithm development may provide maps allowing calculation of waterway-based distance matrices among lakes (Domisch et al. 2024).

In the present study, the reduction in trace of the gllvm was already substantial between the unconstrained model and the model including only environmental predictors, suggesting that spatial effects or effects from predictors not accounted for contributed only weakly to the prediction of rare species presence/absence and abundance patterns. Accordingly, we did not interpret the residual covariance matrices of the models, which summarize effects from drivers unaccounted for and potential species-interaction effects (Blanchet et al. 2020; Münkemüller et al. 2020). In contrast, in our previous study on common fish species (Mehner et al. 2021a), the biomasses and mean sizes of the six most common fish species in European lakes showed some meaningful and interpretable covariance patterns with respect to predator-prey interactions (Mehner et al. 2021a). Therefore, abundance or occurrence of rare species in the lakes studied here may have been affected by co-occurring common species. However, we excluded the numerically strongly dominant common species by purpose, to ensure that the models predict the environmental and spatial drivers of only the rare species occurrence and abundance. Accordingly, we cannot draw conclusions on potential structuring effects from negative interspecific interactions by common on rare species in the lakes.

Inferring assembly processes by species traits

To link the patterns of commonness or rarity of all 82 fish species with processes of community assembly (Crisfield et al. 2024), we tried to predict species geographic ranges by seven fish traits. We found strong predictors of geographic range related to the life history of the fish species, i.e. maximum size, fecundity and age at maturity. However, there was no consistent pattern showing rarity to be linked to a certain life history type. Higher rarity by smaller geographic range was connected with large size, intermediate fecundity and early age at maturity, a trait combination that does not correspond to any life history type in the triangular continuum of freshwater fish species (Winemiller and Rose 1992). It is more likely that these predictors of rarity by geographic range include many different fish species that are either small, fecund and have low age at maturity (such as genera *Gobio*, *Phoxinus*, *Barbatula*, *Cobitis*) or those that are large and grow slowly (such as some species of the genera *Leuciscus* or *Luciobarbus*), which would indicate that life history traits act independently from each other in predicting geographic range. In contrast, several other studies have found a strong correspondence between life history and rare species range size, occupancy or abundance (Böhning-Gaese et al. 2006; Laube et al. 2013; Heino and Grönroos 2014; Heino and Tolonen 2018). We also tested whether non-native fish species were disproportionally frequently represented among the rare species because they might still be at the beginning of spread and colonisation and thus potentially rare with respect to occupancy, range and abundance. However, the number of non-native fish species in European lakes at the time of the surveys has still been low, and many of these non-native species occurred in a few lakes only (Trochine et al. 2018). Therefore, native status was found to be just one among several traits that correlates weakly with rarity. Accordingly, the hypothesised strong link between rarity patterns and underlying processes realised through certain functional traits (Crisfield et al. 2024) could not be found in our analyses. However, the information on traits for the species included in our study was not complete, with substantial gaps, in particular, for some rare species (Supplementary Table S2), and the BRT results may accordingly be partly biased towards the traits of the more common species.

Ultimately, it is also obvious that the list of rare species encompassed a rather large number of species for which lakes are not the most preferred ecosystem type. The list of rare fish species includes oligohaline (hence primarily marine) species occasionally found in coastal brackish lakes such as goby (*Pomatoschistus minutus*), sprat (*Sprattus sprattus*) and mullets (genera *Chelon* and *Mugil*). Other species that are usually considered riverine, such as the genera *Luciobarbus*, *Pseudochondrostoma*, *Barbus* and *Squalius*, often occurred in reservoirs originating from dammed rivers, in particular, on the Iberian Peninsula. The strong effect of the dammed rivers on the reservoir fauna is obvious from the geographic map of rare species distribution and from the importance of lake type (natural vs. artificial) as a predictor in the latent variable models. Therefore, these originally riverine species have rare occurrences in lentic waters (reservoirs), while they may be more common in their preferred lotic habitats. Many southern European species also have a small geographic distribution due to their paleohistoric evolution and isolation, and colonisation of newly formed lentic habitats and population build-up are slow. Similar causes of rarity in lakes were found also for several cold-water riverine salmonids such as salmon (*Salmo salar*) and grayling (*Thymallus thymallus*). These were often caught in lakes connected to rivers but are not part of the typical lake fish fauna because they reproduce only in lotic systems. Accordingly, the reasoning that the rarity of many fish species in our dataset was caused by the fact that they do not reproduce in the habitat in which they were caught resembles the earlier distinction between diffusive and suffusive rarity (Schoener 1987). Species with reproduction outside of lakes, including anadromous and catadromous migrants, may be considered as the diffusive part of rarity in the lake fish fauna.

Similarity in community assembly processes between rare and common species

According to the results of this and our previous study (Mehner et al. 2021a), we did not find strong evidence that the dominant community assembly processes differed between common and rare species. Several other studies have hypothesised that commonness is caused by deterministic processes like environmental filtering, while rarity is predominantly caused by neutral processes, in particular those related to dispersal and spatial arrangement of locations (Siqueira et al. 2012; Iop et al. 2024). In contrast, we found strong correspondence of processes inferred from patterns between common and rare fish species, suggesting that the fish community assembly in European lakes is primarily driven by environmental filtering. A similar conclusion on the dominance of environmental drivers for the community assembly of fishes in Canadian, Japanese and Swedish lakes, mediated by spatial processes and species interactions at lake level, was drawn from joint species distribution models (Kadoya et al. 2024). Large-scale (i.e., continental) studies often demonstrate strong effects of climatic variables on species occurrences or size, while locally important processes, for example anthropogenic stressors or the effect of species interactions, emerge only from spatially restricted analyses (Brucet et al. 2013; Emmrich et al. 2014). Therefore, it would not be justified to develop general conservation actions for rare fish species in European lakes, as based on our study, without conducting more detailed species-specific analyses for the lakes, in which a certain rare species occurs.

It should be considered that historical contingency, a stochastic process, may also have contributed to the overall fish distribution patterns. Large parts of northern and central Europe were glaciated repeatedly during the last about 2.6 million years, and the last retreat of the ice shield dates back only 12,000 to 15,000 years. These climatic effects led to a major depauperation of the European fish fauna (Lèvêque et al. 2008) with the consequence that the overall richness is lower here than in comparable areas nearer to glacial refuges (Griffiths 2006) such as Mediterranean basins (Casal-López and Doadrio 2018) or on other continents (Tonn et al. 1990). Furthermore, few eurytopic species such as Eurasian perch, pike, roach, rudd, ruffe and bream dominate in numerous lakes, while many rare species have limited distribution ranges, which may still reflect the effects of the last glaciation and the subsequent constrained recolonisation from few refugial areas (Hewitt 1999). A particular phenomenon is also a few (often endemic) species pairs or species flocks with narrow distribution ranges in the order Salmoniformes (salmon-like fishes) or Clupeiformes (herring-like fishes) including trout (*Salmo)*, whitefish (*Coregonus*), charr (*Salvelinus*) and shad (*Alosa*), originating from incipient speciation after the last glaciation (Kottelat and Freyhof 2007; Hendry 2009; Blain et al. 2023; Tiddy et al. 2024). Speciation is another stochastic process contributing to community assembly, and the emergence of endemic species suggests that local fish richness in postglacial lakes may increase substantially after a sufficient passage of time. These incipient species can be considered truly rare in European lakes because they occur in only few locations and appear in relatively low abundances even at these few localities. These species match all definitions of rarity by Rabinowitz (1981) and correspond to the suffusive rarity (Schoener 1987). If rarity of occurrence is used to develop conservation actions for lake fish species, lakes with locally endemic salmonids are the most likely candidates for strong protection measures to prevent loss of unique biodiversity (Mehner et al. 2021b; Ferguson and Prodohl 2022).

Conclusions

We demonstrate here that the vast majority of fish species in European lakes can be considered rare, while only very few common species dominate the communities. The major determinants of rare species occurrences and abundances at the European scale are climatic factors, but stochastic processes such as historical contingencies may substantially contribute to the patterns. Furthermore, many rare species were found to not being typical lake-dwelling species, but occurring occasionally in lakes because there is connectivity among lakes, rivers and estuaries. A dataset unifying occurrence and abundance data on freshwater fishes in lakes and rivers in Europe, based on a hydraulic connectivity grid including potential migration barriers (for an example with intermediate spatial resolution, see https://amber.international/european-barrier-atlas/), would be a substantial step towards understand ing fish community assembly mechanisms. General conservation actions for rare species in lakes cannot be developed based on this European dataset, but lakes with postglacial incipient fish speciation events can be suggested for explicit conservation surveillance of populations of endemic species.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10531-025-03046-5.

Acknowledgements We thank Samo Podgornik, Thrond Haugen, Trygve Hesthagen, David Ritterbusch and Lluis Benejam for help with fish sampling and raw data analyses. Many thanks to three anonymous reviewers, whose comments helped improving the text.

Author contributions Study conception, design and formal analysis were done by Thomas Mehner. Material preparation, data collection and curation were performed by all authors. The first draft of the manuscript was written by Thomas Mehner, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was funded by WISER (Water bodies in Europe: Integrative Systems to assess Ecological status and Recovery) funded by the European Union under the 7th Framework Programme, Theme 6 (Environment including Climate Change, contract No. 226273).

Data availability Data accessibility is described in: Mehner T, Brucet S, Argillier C, Beklioglu M, Ferreira T, Hesthagen T, Holmgren K, Jeppesen E, Kelly F, Krause T, Rask M, Volta P, Winfield IJ, Podgornik S. 2017. Metadata of European lake fishes dataset. Freshwater Metadata Journal 23: 1–8.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abell R, Thieme ML, Revenga C, Bryer M, Kottelat M, Bogutskaya N, Coad B, Mandrak N, Balderas SC, Bussing W, Stiassny MLJ, Skelton P, Allen GR, Unmack P, Naseka A, Ng R, Sindorf N, Robertson J, Armijo E, Higgins JV, Heibel TJ, Wikramanayake E, Olson D, Lopez HL, Reis RE, Lundberg JG, Perez MHS, Petry P (2008) Freshwater ecoregions of the world: A new map of biogeographic units for freshwater biodiversity conservation. Bioscience 58:403–414

Alexander TJ, Vonlanthen P, Periat G, Degiorgi F, Raymond JC, Seehausen O (2015) Estimating whole-lake fish catch per unit effort. Fish Res 172:287–302

Argillier C, Causse S, Gevrey M, Pedron S, De Bortoli J, Brucet S, Emmrich M, Jeppesen E, Lauridsen T, Mehner T, Olin M, Rask M, Volta P, Winfield IJ, Kelly F, Krause T, Palm A, Holmgren K (2013) Development of a fish-based index to assess the eutrophication status of European lakes. Hydrobiologia 704:193–211

Beisner BE, Peres PR, Lindstrom ES, Barnett A, Longhi ML (2006) The role of environmental and Spatial processes in structuring lake communities from bacteria to fish. Ecology 87:2985–2991

Blain SA, Schluter D, Adams CE, Amundsen PA, Knudsen R, Chavarie L (2023) Patterns and repeatability of multi-ecotype assemblages of sympatric salmonids. Glob Ecol Biogeogr 32:2257–2270

Blanchet FG, Cazelles K, Gravel D (2020) Co-occurrence is not evidence of ecological interactions. Ecol Lett 23:1050–1063

Boettiger C, Lang DT, Wainwright PC (2012) Rfishbase: exploring, manipulating and visualizing fishbase data from R. J Fish Biol 81:2030–2039

Böhning-Gaese K, Caprano T, van Ewijk K, Veith M (2006) Range size: disentangling current traits and phylogenetic and biogeographic factors. Am Nat 167:555–567

Brucet S, Pedron S, Mehner T, Lauridsen TL, Argillier C, Winfield IJ, Volta P, Emmrich M, Hesthagen T, Holmgren K, Benejam L, Kelly F, Krause T, Palm A, Rask M, Jeppesen E (2013) Fish diversity in European lakes: geographical factors dominate over anthropogenic pressures. Freshw Biol 58:1779–1793

Callaghan CT, Borda-de-agua L, van Klink R, Rozzi R, Pereira HM (2023) Unveiling global species abundance distributions. Nat Ecol Evol 7:1600–1609

Casal-López M, Doadrio I (2018) The Messinian imprint on the evolution of freshwater fishes of the genus *Luciobarbus* Heckel, 1843 (Teleostei, Cyprinidae) in the Western mediterranean. J Biogeogr 45:1593–1603

CEN (2015) Water quality - Sampling of fish with multi-mesh gillnets. European Committee for Standardization, Brussels

Chang C, HilleRisLambers J (2016) Integrating succession and community assembly perspectives. F1000Res 5. https://doi.org/10.12688/f11000research.18973.12681

Cottenie K, De Meester L (2003) Connectivity and Cladoceran species richness in a metacommunity of shallow lakes. Freshw Biol 48:823–832

Crisfield VE, Blanchet FG, Raudsepp-Hearne C, Gravel D (2024) How and why species are rare: towards an understanding of the ecological causes of rarity. Ecography 2024: https://doi.org/10.1111/ecog.07037

De Bie T, De Meester L, Brendonck L, Martens K, Goddeeris B, Ercken D, Hampel H, Denys L, Vanhecke L, Van der Gucht K, Van Wichelen J, Vyverman W, Declerck SAJ (2012) Body size and dispersal mode as key traits determining metacommunity structure of aquatic organisms. Ecol Lett 15:740–747

De'ath G (2007) Boosted trees for ecological modeling and prediction. Ecology 88:243–251

Dee LE, Cowles J, Isbell F, Pau S, Gaines SD, Reich PB (2019) When do ecosystem services depend on rare species? Trends Ecol Evol 34:746–758

Diekmann M, Brämick U, Lemcke R, Mehner T (2005) Habitat-specific fishing revealed distinct indicator species in German lowland lake fish communities. J Appl Ecol 42:901–909

Domisch S, Bremerich V, Buurman M, Kaminke B, Tomiczek T, Torres-Cambas Y, Grigoropoulou A, Garcia Marquez JR, Amatulli G, Grossart H-P, Gessner MO, Mehner T, Adrian R, De Meester L (2024) Geo-FRESH– an online platform for freshwater Geospatial data processing. Int J Digit Earth 17:1

Eken G, Bennun L, Brooks TM, Darwall W, Fishpool LDC, Foster M, Knox D, Langhammer P, Matiku P, Radford E, Salaman P, Sechrest W, Smith ML, Spector S, Tordoff A (2004) Key biodiversity areas as site conservation targets. Bioscience 54:1110–1118

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. J Anim Ecol 77:802–813

Emmrich M, Pedron S, Brucet S, Winfield IJ, Jeppesen E, Volta P, Argillier C, Lauridsen TL, Holmgren K, Hesthagen T, Mehner T (2014) Geographical patterns in the body-size structure of European lake fish assemblages along abiotic and biotic gradients. J Biogeogr 41:2221–2233

Ferguson A, Prodohl PA (2022) Identifying and conserving sympatric diversity in trout of the genus *Salmo*, with particular reference to Lough Melvin, Ireland. Ecol Freshw Fish 31:177–207

Froese R, Pauly D (2022) FishBase. World Wide Web electronic publication. www.fishbase.org

Gaston KJ (2011) Common Ecol Bioscience 61:354–362

Gaston KJ, Lawton JH (1990) The population ecology of rare species. J Fish Biol 37:97–104

Griffiths D (2006) Pattern and process in the ecological biogeography of European freshwater fish. J Anim Ecol 75:734–751

Harnik PG, Simpson C, Payne JL (2012) Long-term differences in extinction risk among the seven forms of rarity. Proceedings of the Royal Society B-Biological Sciences 279: 4969–4976

Heino J, Grönroos M (2014) Untangling the relationships among regional occupancy, species traits, and niche characteristics in stream invertebrates. Ecol Evol 4:1931–1942

Heino J, Tolonen KT (2018) Ecological niche features override biological traits and taxonomic relatedness as predictors of occupancy and abundance in lake Littoral macroinvertebrates. Ecography 41:2092–2103

Hendry AP (2009) Ecological speciation! Or the lack thereof? Can J Fish Aquat Sci 66:1383–1398

Hewitt GM (1999) Post-glacial re-colonization of European biota. Biol J Linn Soc 68:87–112

Holmgren K, Degerman E, Petersson E, Bergquist B (2016) Long term trends of fish after liming of Swedish streams and lakes. Atmos Environ 146:245–251

Iop S, Caldart VM, Vélez-Martin E, dos Santos TG, Prado PI, Pillar VD, Cechin SZ (2024) Niche and neutral-based processes differ in importance for common and rare species in a metacommunity of anurans in subtropical grasslands. Hydrobiologia 851:713–732

Jeliazkov A, Gavish Y, Marsh CJ, Geschke J, Brummitt N, Rocchini D, Haase P, Kunin WE, Henle K (2022) Sampling and modelling rare species: conceptual guidelines for the neglected majority. Glob Change Biol 28:3754–3777

Jeppesen E, Jensen JP, Søndergaard M, Lauridsen T, Landkildehus F (2000) Trophic structure, species richness and biodiversity in Danish lakes: changes along a phosphorus gradient. Freshw Biol 45:201–218

Jetz W, Rahbek C (2002) Geographic range size and determinants of avian species richness. Science 297:1548–1551

Kadoya T, Nilsson KA, Kelly J, Bartley TJ, Säterberg T, Guzzo MM, Esch E, Koide D, Matsuzaki SIS, Terui A, Akasaka M, Macdougall AS (2024) Common processes drive metacommunity structure in freshwater fish. Glob Ecol Biogeogr 33. https://doi.org/10.1111/geb.13822

Kottelat M, Freyhof J (2007) Handbook of European freshwater fishes. Cornot and Berlin: Kottelat and Freyhof 1-646p

Kunin WE, Gaston KJ (1993) The biology of rarity - patterns, causes and consequences. Trends Ecol Evol 8:298–301

Laube I, Korntheuer H, Schwager M, Trautmann S, Rahbek C, Böhning-Gaese K (2013) Towards a more mechanistic Understanding of traits and range sizes. Glob Ecol Biogeogr 22:233–241

Lennon JJ, Koleff P, Greenwood JJD, Gaston KJ (2004) Contribution of rarity and commonness to patterns of species richness. Ecol Lett 7:81–87

Lèvêque C, Oberdorff T, Paugy D, Stiassny MLJ, Tedesco PA (2008) Global diversity of fish (Pisces) in freshwater. Hydrobiologia 595:545–567

MacDougall AS, Harvey E, McCune JL, Nilsson KA, Bennett J, Firn J, Bartley T, Grace JB, Kelly J, Tunney TD, McMeans B, Matsuzaki SIS, Kadoya T, Esch E, Cazelles K, Lester N, McCann KS (2018) Context-dependent interactions and the regulation of species richness in freshwater fish. Nature Communications 9: ARTN 973

Maciel EA (2021) An index for assessing the rare species of a community. Ecol Ind 124:ARTN107424

Magurran AE (2004) Measuring biological diversity. Blackwell, Oxford, pp 1–256

Magurran AE, Henderson PA (2003) Explaining the excess of rare species in natural species abundance distributions. Nature 422:714–716

Martinsen KT, Kristensen E, Baastrup-Spohr L, Sondergaard M, Carl H, Jeppesen E, Sand-Jensen K, Kragh T (2023) Environmental predictors of lake fish diversity across gradients in lake age and Spatial scale. Freshw Biol 68:1122–1135

Massicotte P, South A (2023) rnaturalearth: World Map Data from Natural Earth. https://docs.ropensci.org/rnaturalearth/https://github.com/ropensci/rnaturalearth

McGill BJ (2019) The what, how and why of doing macroecology. Glob Ecol Biogeogr 28:6–17

Mehner T, Diekmann M, Brämick U, Lemcke R (2005) Composition of fish communities in German lakes as related to lake morphology, trophic State, shore structure and human use intensity. Freshw Biol 50:70–85

Mehner T, Holmgren K, Lauridsen TL, Jeppesen E, Diekmann M (2007) Lake depth and geographical position modify lake fish assemblages of the European 'central plains' ecoregion. Freshw Biol 52:2285–2297

Mehner T, Emmrich M, Hartwig S (2014) Spatial predictors of fish species composition in European lowland lakes. Ecography 37:73–79

Mehner T, Keeling C, Emmrich M, Holmgren K, Argillier C, Volta P, Winfield IJ, Brucet S (2016) Effects of fish predation on density and size spectra of prey fish communities in lakes. Can J Fish Aquat Sci 73:506–518

Mehner T, Brucet S, Argillier C, Beklioglu M, Ferreira T, Hesthagen T, Holmgren K, Jeppesen E, Kelly F, Krause T, Rask M, Volta P, Winfield IJ, Podgornik S (2017) Metadata of European lake fishes dataset. Freshw Metadata J 23:1–8

Mehner T, Pohlmann K, Bittner D, Freyhof J (2018) Testing the Devil's impact on Southern Baltic and North sea basins Whitefish (*Coregonus* spp.) diversity. BMC Evol Biol 18:208

Mehner T, Argillier C, Hesthagen T, Holmgren K, Jeppesen E, Kelly F, Krause T, Olin M, Volta P, Winfield IJ, Brucet S (2021a) Model-based decomposition of environmental, Spatial and species-interaction effects on the community structure of common fish species in 772 European lakes. Glob Ecol Biogeogr 30:1558–1571

Mehner T, Palm S, Delling B, Karjalainen J, Kielpinska J, Vogt A, Freyhof J (2021b) Genetic relationships between sympatric and allopatric Coregonus ciscoes in North and Central Europe. Bmc Ecology and Evolution 21: ARTN 186

Mouillot D, Graham NAJ, Villeger S, Mason NWH, Bellwood DR (2013) A functional approach reveals community responses to disturbances. Trends Ecol Evol 28:167–177

Münkemüller T, Gallien L, Pollock LJ, Barros C, Carboni M, Chalmandrier L, Mazel F, Mokany K, Roquet C, Smycka J, Talluto MV, Thuiller W (2020) Dos and Don'ts when inferring assembly rules from diversity patterns. Glob Ecol Biogeogr 29:1212–1229

New M, Lister D, Hulme M, Makin I (2002) A high-resolution data set of surface climate over global land areas. Climate Res 21:1–25

Niku J, Hui FKC, Taskinen S, Warton DI (2019) Gllvm: fast analysis of multivariate abundance data with generalized linear latent variable models in R. Methods Ecol Evol 10:2173–2182

Prchalova M, Kubecka J, Riha M, Mrkvicka T, Vasek M, Juza T, Kratochvil M, Peterka J, Drastik V, Krizek J (2009) Size selectivity of standardized multimesh gillnets in sampling coarse European species. Fish Res 96:51–57

Prendergast JR, Quinn RM, Lawton JH, Eversham BC, Gibbons DW (1993) Rare species, the coincidence of diversity hotspots and conservation strategies. Nature 365:335–337

Pritt JJ, Frimpong EA (2010) Quantitative determination of rarity of freshwater fishes and implications for imperiled-species designations. Conserv Biol 24:1249–1258

R Development Core Team (2022) R: a Language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Rabinowitz D (1981) Seven forms of rarity. Synge H editor. The biological aspects of rare plants conservation. Wiley, Somerset, pp 205–217

Reed JM, Nguyen O, Owens ACS, Lewis SM (2020) Linking the seven forms of rarity to extinction threats and risk factors: an assessment of North American fireflies. Biodivers Conserv 29:57–75

Ridgeway G (2006) Generalized boosted regression models. Documentation on the R Package 'gbm', version 1·5–7

Ritterbusch D, Blabolil P, Breine J, Erös T, Mehner T, Olin M, Peirson G, Volta P, Poikane S (2022) European fish-based assessment reveals high diversity of systems for determining ecological status of lakes. Sci Total Environ 802:ARTN149620

Santoul F, Figuerola J, Mastrorillo S, Cereghino R (2005) Patterns of rare fish and aquatic insects in a Southwestern French river catchment in relation to simple physical variables. Ecography 28:307–314

Schoener TW (1987) The geographical distribution of rarity. Oecologia 74:161–173

Siqueira T, Bini LM, Roque FO, Couceiro SRM, Trivinho-Strixino S, Cottenie K (2012) Common and rare species respond to similar niche processes in macroinvertebrate metacommunities. Ecography 35:183–192

Smejkal M, Ricard D, Prchalova M, Riha M, Muska M, Blabolil P, Cech M, Vasek M, Juza T, Herreras AM, Encina L, Peterka J, Kubecka J (2015) Biomass and abundance biases in European standard gillnet sampling. Plos One 10: ARTN e0122437

Stegen JC, Lin XJ, Fredrickson JK, Chen XY, Kennedy DW, Murray CJ, Rockhold ML, Konopka A (2013) Quantifying community assembly processes and identifying features that impose them. ISME J 7:2069–2079

Sykes L, Santini L, Etard A, Newbold T (2020) Effects of rarity form on species' responses to land use. Conserv Biol 34:688–696

Tiddy IC, Schneier K, Elmer KR (2024) Environmental correlates of adaptive diversification in postglacial freshwater fishes. J Fish Biol 104:517–535

Tonn WM, Magnuson JJ, Rask M, Toivonen J (1990) Intercontinental comparison of small lake fish assemblages: the balance between local and regional processes. Am Nat 136:345–375

Trochine C, Brucet S, Argillier C, Arranz I, Beklioglu M, Benejam L, Ferreira T, Hesthagen T, Holmgren K, Jeppesen E, Kelly F, Krause T, Rask M, Volta P, Winfield IJ, Mehner T (2018) Non-native fish occurrence and biomass in 1943 Western Palearctic lakes and reservoirs and their abiotic and biotic correlates. Ecosystems 21:395–409

van Galen LG, Orlovich DA, Lord JM, Bohorquez J, Nilsen AR, Summerfield TC, Larcombe MJ (2023) Zeta diversity differentiates factors driving community assembly of rare and common ectomycorrhizal fungi. Mol Ecol 32:2092–2109

Vellend M (2010) Conceptual synthesis in community ecology. Q Rev Biology 85:183–206

Wei T, Simko V (2021) R package 'corrplot': Visualization of a Correlation Matrix. https://github.com/taiyun/corrplot

Wickham H (2016) ggplot2: elegant graphics for data analysis. Springer, New York

Winemiller KO, Rose KA (1992) Patterns of life-history diversification in north-American fishes - implications for population regulation. Can J Fish Aquat Sci 49:2196–2218

Yu H, Cooper AR, Infante DM (2020) Improving species distribution model predictive accuracy using species abundance: Application with boosted regression trees. Ecological Modelling 432: ARTN 109202

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Thomas Mehner[1] · **Christine Argillier**[2] · **Teresa Ferreira**[3] · **Kerstin Holmgren**[4] · **Erik Jeppesen**[5,6,7,8] · **Fiona Kelly**[9] · **Teet Krause**[10] · **Mikko Olin**[11] · **Pietro Volta**[12] · **Ian J. Winfield**[13] · **Sandra Brucet**[14,15]

✉ Thomas Mehner
thomas.mehner@igb-berlin.de

[1] Leibniz Institute of Freshwater Ecology and Inland Fisheries, Müggelseedamm 310, 12587 Berlin, Germany

[2] INRAE, Aix Marseille Univ, RECOVER, Aix-en-Provence 13182, France

[3] Forest Research Centre, Associate Laboratory TERRA, University of Lisbon, Tapada da Ajuda, Lisbon 1349-017, Portugal

[4] Department of Aquatic Resources, Swedish University of Agricultural Sciences, Institute of Freshwater Research, Drottningholm, Sweden

[5] Department of Ecoscience, Aarhus University, Aarhus 8000, Denmark

[6] Sino-Danish Centre for Education and Research, Beijing 100049, China

[7] Limnology Laboratory, Department of Biological Sciences and Centre for Ecosystem Research and Implementation, Middle East Technical University, Ankara 06800, Turkey

[8] Institute for Ecological Research and Pollution Control of Plateau Lakes, School of Ecology and Environmental Sciences, Yunnan University, Kunming 650500, China

[9] Inland Fisheries Ireland, 3044 Lake Drive, Citywest Business Campus, Dublin 24, Dublin, Ireland

[10] Institute of Agriculture and Environment, Estonian University of Life Sciences, Tartu, Estonia

[11] Natural Resources Institute Finland, P.O. Box 2, Helsinki FI-00791, Finland

[12] IRSA-CNR, Water Research Institute of the National Research Council, L.go Tonolli 50, Verbania 28922, Italy

[13] Lake Ecosystems Group, UK Centre for Ecology & Hydrology, Lancaster Environment Centre, Library Avenue, Bailrigg, Lancaster LA1 4AP, UK

[14] Aquatic Ecology Group, University of Vic - Central University of Catalonia, Vic, Catalonia, Spain

[15] CREA, Catalan Institution for Research and Advanced Studies, Barcelona, Spain