Environment and biogeography drive aquatic plant and cladoceran species richness across Europe

DUARTE S. VIANA*, L. SANTAMARÍA*,†, K. SCHWENK†,‡,§, M. MANCA¶, A. HOBÆK**,††, M. MJELDE‡‡,
C. D. PRESTON§§, R. J. GORNALL¶¶, J. M. CROFT§§, R. A. KING***, A. J. GREEN* AND J. FIGUEROLA*

Wetland Ecology, Estación Biológica de Doñana (EBD-CSIC), Sevilla, Spain
†*Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, The Netherlands*
‡*Institute for Environmental Sciences, University of Koblenz-Landau, Landau/Pfalz, Germany*
§*Biodiversity and Climate Research Centre (BiK-F), Frankfurt am Main, Germany*
¶*CNR – Insitute of Ecosystem Study, Verbania, Italy*
**Norwegian Institute for Water Research (NIVA), Region West, Bergen, Norway*
††*Department of Biology, University of Bergen, Bergen, Norway*
‡‡*Norwegian Institute for Water Research (NIVA), Oslo, Norway*
§§*Centre for Ecology and Hydrology, Wallingford, U.K.*
¶¶*Department of Biology, University of Leicester, Leicester, U.K.*
***College of Life and Environmental Sciences, University of Exeter, Exeter, U.K.*

SUMMARY

1. The relative role of environmental and regional processes in determining patterns of biodiversity may vary across spatial scales. In inland aquatic communities, macrophytes and zooplankton have the potential to disperse over large spatial scales (hundreds of kilometres), but the role of environmental and regional processes in determining species richness (SR) at different spatial scales is poorly understood.

2. To assess the relative importance of environmental and regional processes acting on both regional (within-region) and continental (among-region) scales, we surveyed 139 waterbodies in five geographic regions (Spain, Italy, the Netherlands, Scotland and Norway) and in each region measured environmental variables, descriptors of connectivity and SR of two aquatic taxa, namely aquatic plants and cladocerans.

3. Analyses using boosted regression trees and generalised linear models, integrated with multiple imputation of missing data, indicated that both environmental and regional processes were related to local SR.

4. Total phosphorus was the main environmental correlate of SR, showing a unimodal relationship with both taxonomic groups. Conductivity and lake depth were also related to cladoceran SR, and Secchi depth was related to aquatic plant richness.

5. The SR of aquatic plants and cladocerans were significantly correlated, but this relationship was mainly indirect (i.e. determined by the effect of common environmental variables).

6. The limited role of connectivity and space (lack of spatial autocorrelation in model residuals) relative to environment suggests that dispersal limitation is not as important as the environment in explaining richness patterns for both groups within regions. However, region identity strongly influenced SR, which suggests an important effect of biogeographic factors (e.g. the level of endemicity; the number of unique species in each region) and/or dispersal limitation at the continental scale.

7. We conclude that environmental factors and biogeographic processes largely determine the patterns of local and regional SR in aquatic plants and zooplankton.

Keywords: aquatic plants, spatial scale, species distribution, species richness, zooplankton

Correspondence: Duarte S. Viana, Wetland Ecology, Estación Biológica de Doñana (EBD-CSIC), c/Américo Vespucio, s/n, 41092, Isla de la Cartuja, Sevilla, Spain. E-mail: dviana@ebd.csic.es

Introduction

Local species richness (SR) is determined by environmental factors and biotic interactions (local processes), as well as the dispersal of organisms from the regional pool of species into the local habitat patch (regional process; Ricklefs, 1987). Dispersal is especially important in explaining SR in fragmented and naturally isolated habitats (e.g. islands, wetlands and mountain tops) where metacommunity dynamics prevail (Davies, Melbourne & Margules, 2001; Cottenie & De Meester, 2004; Leibold *et al.*, 2004). In inland aquatic habitats, dispersal has been generally assumed to be one of the key determinants of local SR (Cottenie & De Meester, 2004). Indeed, several studies have provided evidence of dispersal limitation (e.g. Capers, Selsky & Bugbee, 2010; Frisch *et al.*, 2012) and its influence on SR in aquatic taxa (e.g. Dodson, 1992). However, other observational and experimental studies have shown that local, rather than regional, processes (i.e. local environment instead of dispersal limitation) determine community diversity in zooplankton (Shurin, 2000; Forbes & Chase, 2002; Cottenie & De Meester, 2003, 2004) and macrophyte assemblages (e.g. Alahuhta & Heino, 2013).

Numerous field and experimental observations attest to the strong dispersal abilities of many aquatic organisms (e.g. high colonisation rates of zooplankton, Louette & De Meester, 2005; low spatial autocorrelation of passively dispersed organisms, Shurin, Cottenie & Hillebrand, 2009), largely through the passive dispersal of dormant propagules (resting eggs and seeds) by abiotic vectors, such as water and wind (Cáceres & Soluk, 2002), and biotic vectors, usually vertebrate animals such as amphibians (Bohanak & Whiteman, 1999), birds (Figuerola & Green, 2002; Brochet *et al.*, 2010; Van Leeuwen *et al.*, 2012) and mammals (Vanschoenwinkel *et al.*, 2008). While abiotic vectors can mediate dispersal over short scales, vertebrates, particularly water birds, are able to disperse the propagules of aquatic plants and zooplankton over hundreds of kilometres on a regular basis (Viana *et al.*, 2013).

In addition to the high vagility of aquatic plants and zooplankton, which facilitates their frequent dispersal among waterbodies, successful interchange of species or genotypes requires that immigrants are able to establish in the recipient community (i.e. germinate or hatch and reproduce in the recipient community). Local environmental conditions are generally controlled by climatic factors, hydroperiod, water chemistry, habitat size and spatial/structural heterogeneity (Williams, 2001; Lacoul & Freedman, 2006; Bornette & Puijalon,

2011 for aquatic plants; Hessen *et al.*, 2006 for zooplankton). Among these factors, productivity is the main driver of species diversity and community composition in lentic systems (lakes, ponds and wetlands; hereafter 'lakes'). For example, nutrient availability determines the productivity of each waterbody, which in turn influences the SR of both aquatic plants and zooplankton (Dodson, Arnott & Cottingham, 2000; Jeppesen *et al.*, 2000). Such environmental factors both influence and are influenced by community composition, especially keystone taxa known to have strong effects on other elements of the aquatic community, such as aquatic plants and zooplankton, which, through their effects on algal growth and water turbidity, may influence the structure and productivity of the whole community (Scheffer *et al.*, 1993).

Although aquatic plants and zooplankton may respond to common environmental drivers, a direct effect of one group on the other might also underlie covariation in their SR. Disentangling direct and indirect (environment-mediated) effects allows us to discriminate whether one group represents an appropriate surrogate of the other (a phenomenon often referred to as cross-taxon congruence) or whether the common environment explains the SR of both groups. Reported examples of direct interactions between aquatic plants and zooplankton include (i) refuge effects, whereby aquatic plant stands protect zooplankton against predation (e.g. Jeppesen *et al.*, 1997), (ii) habitat heterogeneity effects, whereby aquatic plants provide substrate and/or a diverse spatial template for other aquatic organisms (periphyton, zooplankton and benthic invertebrates; De Meester *et al.*, 2006) and (iii) light-climate effects, whereby the filtering activity of zooplankton removes phytoplanktonic algae, reducing water turbidity and improving conditions for aquatic plant growth (Jeppesen *et al.*, 1999).

We used a survey method specifically designed to document the SR of both aquatic plants (angiosperms and ferns) and cladocerans (as representatives of zooplankton assemblages) in waterbodies across a wide geographic range (five regions in Europe) and spanning a broad range of environmental conditions to evaluate: (i) whether there are generic environmental drivers of aquatic SR, applicable to both functional groups; (ii) whether the SR of these groups is affected by regional processes acting over regional (dispersal limitation) and continental scales (biogeography and/or dispersal limitation); and (iii) whether aquatic plant and cladoceran SR influence each other, through direct or indirect effects.

Methods

Sampling design and data collection

Aquatic plant and cladoceran assemblages were surveyed, respectively, in 128 and 124 of 139 lakes located in five European regions: (i) the lower Guadalquivir catchment (SW Spain; $N = 20$); (ii) the upper and middle Po catchment (NE Italy; $N = 29$); (iii) the lower Rhine and Maas catchments (the Netherlands; $N = 30$); (iv) the Fife area (E Scotland; $N = 30$); and (v) several small, adjacent catchments in central Norway ($N = 30$; see list of sites in Appendix S1, Table S1, in Supporting information). Sampling took place during the summers of 1998 (Spain, the Netherlands and Norway) and 1999 (Scotland and Italy). In each region, a circular area of approximately 150 km radius was defined and a subsample of 30 lakes was randomly selected among those contained within it (based on available cartographic information). Because we aimed at surveying wide environmental gradients, 'lakes' included all sorts of lentic waterbodies from large oligotrophic lakes to shallow ponds and from natural wetlands to water reservoirs (see environmental variation in each region in Appendix S3, Figure S2). Some of the selected lakes in Spain and Italy were not sampled because they were inaccessible or did not allow for proper sampling procedures.

Floating and submerged angiosperms, plus two fern species (Lycopodiophyta), were sampled along four orthogonal transects perpendicular to the lake shore and situated approximately at its intersection with the four cardinal points. In addition, the rest of the lake and its edges were visited, and any additional species recorded. In none of the sampled lakes did we find species at the edges that were not also present within the flooded area. The surveys were carried out during the peak of the growing season in each area, so that all angiosperm species present in the lake through the season could be detected. Furthermore, the composition of aquatic plant assemblages is usually consistent across years over small time scales (a few years), until major environmental changes occur (Sayer *et al.*, 2010). Duplicate herbarium specimens were prepared for each species and transect, and one of them submitted to a central team of taxonomists (C. D. Preston and J. M. Croft) to confirm the identifications based on a common checklist of European aquatic angiosperms agreed among the different teams prior to sampling. Hybrid species were considered different taxonomic units and therefore included in measures of SR (see the list of species in Appendix S1, Table S2).

To obtain a fully representative sample of both littoral and pelagic cladocerans across the recent history of the lake, the composition of cladoceran assemblages was based on a combination of 'live samples', using plankton nets (90 μm mesh) and preserved in ethanol, and 'remains samples', isolated from sediment samples collected at the deepest part of each lake using corers or small van Veen grabs. Analysis of cladoceran remains from sediment samples has been shown to ensure an adequate representation of all species occurring throughout the year while also accounting for interannual variation in species composition (Brendonck & De Meester, 2003; Catalan *et al.*, 2009); hence, they were considered more reliable than single snapshot samples of cladocerans using plankton nets (which were, however, used to aid in the identification of cryptic taxa, by means of genetic analysis). Species were identified largely following Flössner (2000), except for Bosminidae, which were identified according to Flössner (1972) and Lieder (1996). Putative hybrids within the *Daphnia longispina* complex and those resulting from *Bosmina* (*Eubosmina*) species were identified after Wolf & Mort (1986) and Lieder (1983, 1996), respectively. Cladoceran remains were analysed according to Frey (1960). When a full species identification was not possible, the individual(s) was ascribed to the lowest identifiable taxonomic level and included in measures of SR. As with aquatic plants, hybrid species were also considered different taxonomic units and therefore included in measures of SR (see the list of species in Appendix S1, Table S2).

For all the sampled lakes, several geographic and physicochemical variables were recorded: altitude (m), area (km^2), maximum lake depth (m), Secchi depth (m), conductivity (at 20 °C; μS cm^{-2}), pH, calcium (mg L^{-1}) and total phosphorus (TP) (μg P L^{-1}) (except for the Scottish lakes, where lake depth and Secchi depth were not measured). In addition, we recorded the number and area (km^2) of all neighbour lakes within two different radii (10 and 20 km) of each surveyed lake (obtained from cartographic information). These two latter variables measure the availability of nearby immigration sources and were incorporated in the analyses as an estimator of connectivity, similar to other studies in terrestrial (e.g. forest fragments; Magrach, Larrinaga & Santamaría, 2012) and aquatic systems (e.g. lake zooplankton; Dodson, 1992).

Data analyses

Based on data obtained from the 139 lakes, we modelled SR of each taxonomic group as a function of eight environmental variables, four connectivity variables and

the SR of the other group. Dispersal limitation at the regional scale was evaluated by checking the spatial correlation in SR (correcting for spatially structured environment, i.e. spatial autocorrelation in model residuals) and the effect of connectivity variables, whereas biogeographic patterns and dispersal limitation at the continental scale were evaluated by including region identity as a factor variable in the statistical models and investigating regional, local and endemic richness patterns.

Determinations of SR

We calculated local SR (number of species in local assemblages) and regional SR, as well as the number of unique species found in each region (endemic SR). In addition, because sampling effort varied among regions, we calculated corrected regional SR using an asymptotic estimator, Chao 2 (Chao, 1987; Colwell, 2013), and a Michaelis–Menten richness estimator (MMMean), where richness is obtained by functional extrapolation with a fitted analytical rarefaction curve (Colwell, Mao & Chang, 2004; Colwell, 2013). Chao 2 and MMMean were calculated using the software EstimateS (Colwell, 2013).

Statistical analyses

Although a considerable sampling effort was made ($N = 128$ and $N = 124$ for plants and cladocerans, respectively), the high number of measured explanatory variables limited our ability to test for interactions and nonlinear relationships that could be important for explaining SR patterns. We therefore decided to use a machine learning method, boosted regression trees (BRT; Elith, Leathwick & Hastie, 2008), to learn from the data how each explanatory variable affects SR. This method allowed us to model interactions (as do regression trees) and to explore nonlinear relationships that are otherwise difficult to hypothesise. We then used generalised linear models (GLM), fitted using Poisson error distributions and log links, to model linear and quadratic relationships (those observed in the BRT analysis plus those identified through the analysis of GLM residuals) and test which variables contributed significantly to explain SR. GLMs also allowed us to obtain simpler, more general models with fewer parameters. Moreover, the implementation of two types of models (BRT and GLM, explained below) provided a way to check the consistency of results, which is a desirable property when investigating general drivers for a given biological pattern. BRT and GLM models were fitted independently to SR of aquatic plants and cladocerans.

The data set had a few missing values in some of the explanatory variables (4.4 and 4.1% of data values for plants and cladocerans, respectively; SR was measured in all lakes). Since removing all cases with missing values would require removing the data from an entire region (Scotland), we resorted to multiple imputation (MI) of missing data, using multivariate imputation by chained equations (MICE; implemented with the R package *mice*; Van Buuren & Groothuis-Oudshoorn, 2011), before modelling SR. This technique provides a way of estimating unbiased model parameters by accounting for the uncertainty derived from the imputation process.

For both BRT and GLM, the following explanatory variables were considered: altitude, area, depth, Secchi depth, conductivity, pH, TP, connectivity (area and number of lakes within 10 or 20 km, with separate models fitted for each of the two radii to test which one produced the best fit), SR of the other group (either of plants or cladocerans; SR-cross) and region (as a fixed factor with five levels: Spain, Italy, the Netherlands, U.K. and Norway). Calcium was excluded due to co-linearity with conductivity, as indicated by variation inflation factors (VIF <3 for all variables except calcium). All statistical analyses (explained below) were performed using the R program (R Development Core Team, 2012; see detailed statistical methods in Appendix S2).

Results

Regional, local and endemic SR differed among regions even when accounting for sample sizes (Table 1; see species accumulation curves in Appendix S3, Fig. S3).

The relationships between plant and cladoceran SR were weak and depended on the group used as predictor. When cladoceran richness was used as predictor of aquatic plant richness, we observed a unimodal (i.e. a downward parabola) relationship (Fig. 1; $SR_P = 1.715 + 0.046*SR_Z + 0.006*SR_Z^2$, $F_{1,28458} = 21.79$, $P < 0.01$, $R^2 = 0.11$, note the high degrees of freedom resulting from hypothesis testing on data subjected to MI, as explained in Appendix S2). When plant richness was used as predictor of cladoceran richness, a significant but weak linear relationship was found ($SR_Z = 2.209 + 0.267 * \log_{10}(SR_P + 1)$, $F_{1,48629} = 7.39$, $P = 0.01$, $R^2 = 0.02$; Fig. 1). Both relationships became non-significant when environmental variables were included in the models, largely owing to the effect of region and TP, as observed by adding one variable at a time (data not shown).

Table 1 Summary information on the regional distribution of species richness (SR)

Group	Region	N	Total SR	Chao 2	MMMean	Mean local SR	Endemic SR
Plants	Spain	19	14	20.1	17.77	2.10	5
	Italy	19	22	29.1	24.96	6.05	4
	the Netherlands	30	24	22.42	32.3	2.27	3
	Scotland	30	25	22.94	27.47	4.87	3
	Norway	30	31	28.56	32.58	8.37	13
Cladocera	Spain	12	33	42.26	50.59	6.75	5
	Italy	29	55	65.98	57.96	14.34	7
	the Netherlands	25	52	61.05	61.41	10.40	6
	Scotland	28	45	66.85	53.32	7.29	1
	Norway	30	39	41.41	40.48	13.57	7

Chao 2, corrected regional SR using an asymptotic estimator; MMMean, Michaelis–Menten richness estimator (richness is obtained by functional extrapolation with a fitted analytical rarefaction curve).
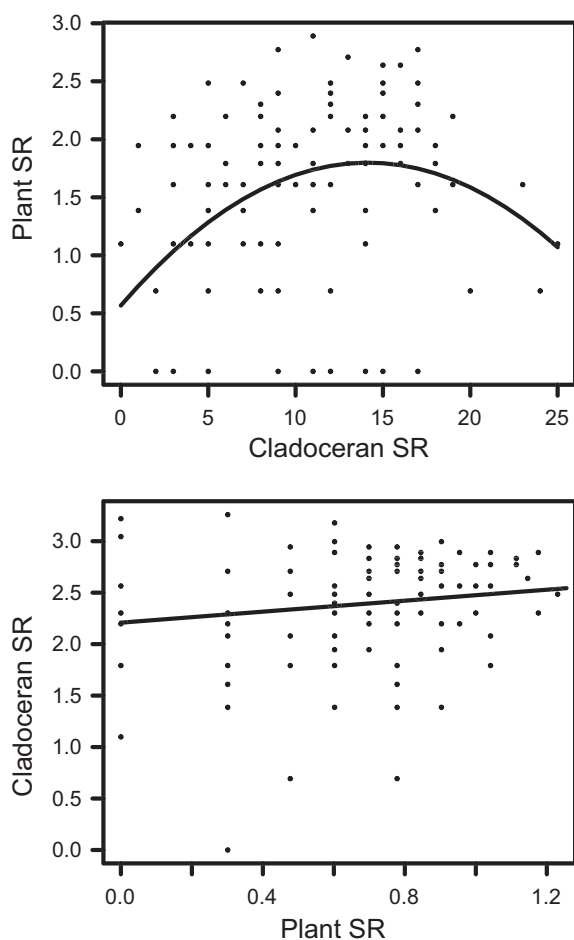

Fig. 1 Cross-taxon relationship between plant and cladoceran species richness (SR). The response variable (*y*-axis) is shown on a log scale (log-SR) in both plots, and the respective models are
$SR_P = 1.715 + 0.046 * SR_Z + 0.006 * SR_Z^2$ for aquatic plants and
$SR_Z = 2.209 + 0.267 * \log_{10}(SR_P + 1)$ for cladocerans.

The results from the BRT models (see Appendix S3, Figures S4 & S5) were overall consistent with those from the GLMs (Figs 2 & 3). Region was the most important variable explaining local SR of both aquatic plants and cladocerans, showing the highest explanatory power in GLMs (Figs 2 & 3; see Appendix S3, Table S3 for GLM results). However, regional variation in local SR differed between aquatic plants and cladocerans, even when accounting for the effect of environmental variation across such different regions. Plant SR was higher in Norway, Italy and Scotland, and lower in Spain and the Netherlands (Fig. 2), whereas cladoceran SR was higher in Italy, the Netherlands and Norway, and lower in Scotland and Spain (Fig. 3).

Independently of the region, only a reduced subset of significant environmental variables determined SR. For both plants and cladocerans, TP and conductivity were the only predictors consistent between the two types of model (GLM and BRT). TP showed a significant unimodal relationship (downward parabola) with both plant (Fig. 2) and cladoceran richness (Fig. 3), plus a negative linear relationship with cladoceran richness (Fig. 3). Conductivity had a quadratic relationship with plant richness (Fig. 2) and a negative linear relationship with cladoceran richness (Fig. 3). On the other hand, pH was an important common predictor in GLMs but not in BRT (see Appendix S3, Figures S4 & S5), probably because only two extreme values (of acidic lakes in the Netherlands) drove the relationship in the GLM but were insufficient to be important in the BRT. In GLMs, pH had a positive relationship with plant richness (Fig. 2) and a unimodal relationship with the cladoceran richness (Fig. 3).

Other important predictors of SR were specific to one of the two groups. Secchi depth was the third most important predictor of plant SR, showing a positive relationship (Fig. 2). In the GLMs, when the Secchi depth was included along with lake depth, none showed a significant relationship with SR, which suggests that the two variables have redundant effects (due to collinearity;
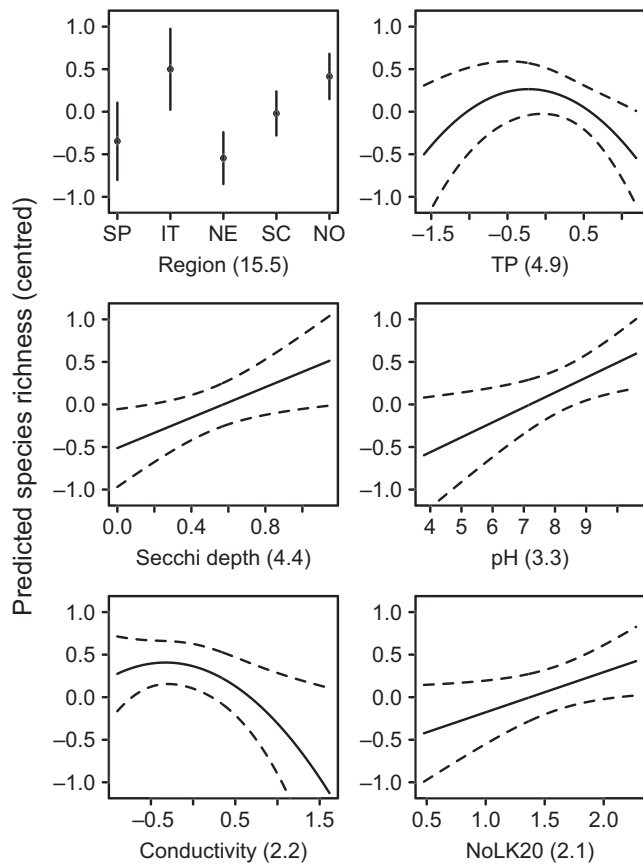
Fig. 2 Partial effect plots for all variables included in the generalised linear model of plant species richness (SR). The *y*-axes represent the marginal effect of the respective explanatory variable on the predicted SR (centred to have zero mean). Dashed lines represent the 95% confidence interval. The relative contribution (%; percentage of explained deviance) of each variable is shown between parentheses in the *x*-axes labels. TP, total phosphorus; NoLK20, number of neighbour lakes within 20 km.

Fig. 3 Partial effect plots for all variables included in the generalised linear model of cladoceran species richness (SR). The *y*-axes represent the marginal effect of the respective explanatory variable on the predicted SR (centred to have zero mean). Dashed lines represent the 95% confidence interval. The relative contribution (%; percentage of explained deviance) of each variable is shown between parentheses in the *x*-axes labels. TP, total phosphorus; AreaLK10, area of neighbour lakes within 10 km.

c. rho = 0.6). These similar relationships can also be observed with the BRT analysis (see Appendix S3, Figures S4 & S5). One connectivity variable (the number of lakes within 20 km) showed a significant relationship with plant SR (Fig. 2), although it explained a relatively low proportion of the variation in SR (only 2.1% of explained deviance, as compared to 15.5% explained by region and 14.8% by the four environmental variables included in the GLM model; Fig. 2).

Lake depth showed a negative relationship with cladoceran SR, although it was only observed with GLMs (Fig. 3). Cladoceran SR increased significantly with one measure of connectivity, the area of surrounding lakes within 10 km, although the relationship only explained 2.2% of the deviance, as compared to 21.3% explained by region and 34.2% by the four environmental variables included in the GLM model (Fig. 3). All variables identi-
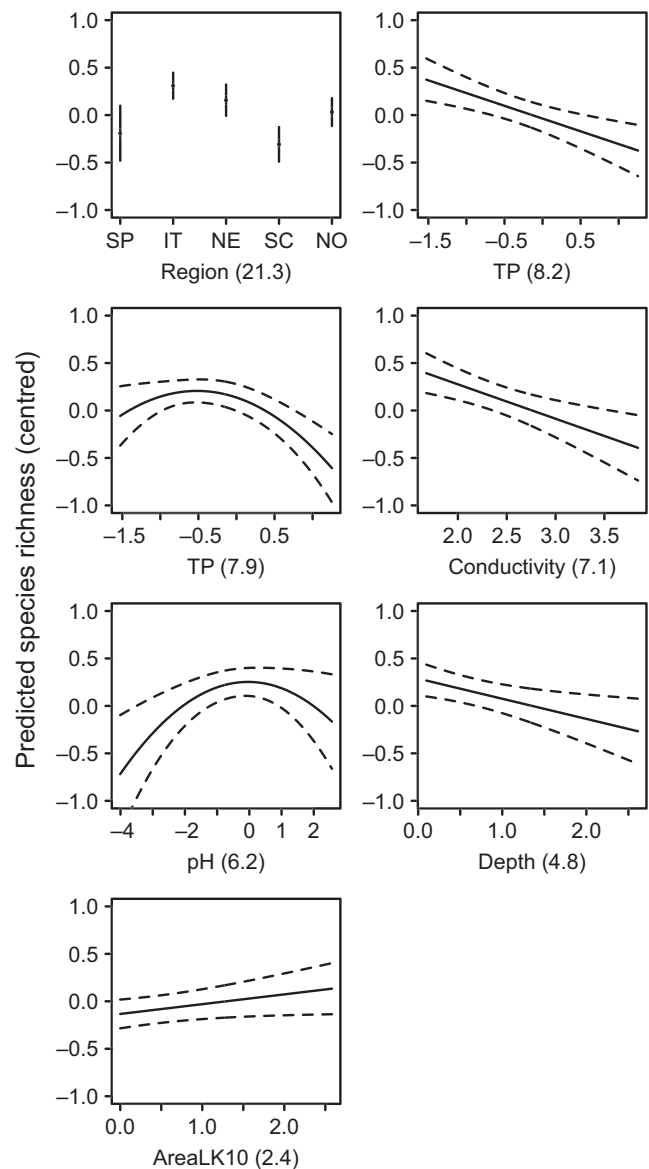
fied as important after the first model selection were kept in the final model.

The total proportion of variance explained by the GLM models was 55 and 44% for aquatic plants and cladocerans, respectively. Correlograms and Moran's I statistics (Fig. 4) showed that, although dependent (SR) and most independent (environment and connectivity) variables were spatially autocorrelated, model residuals
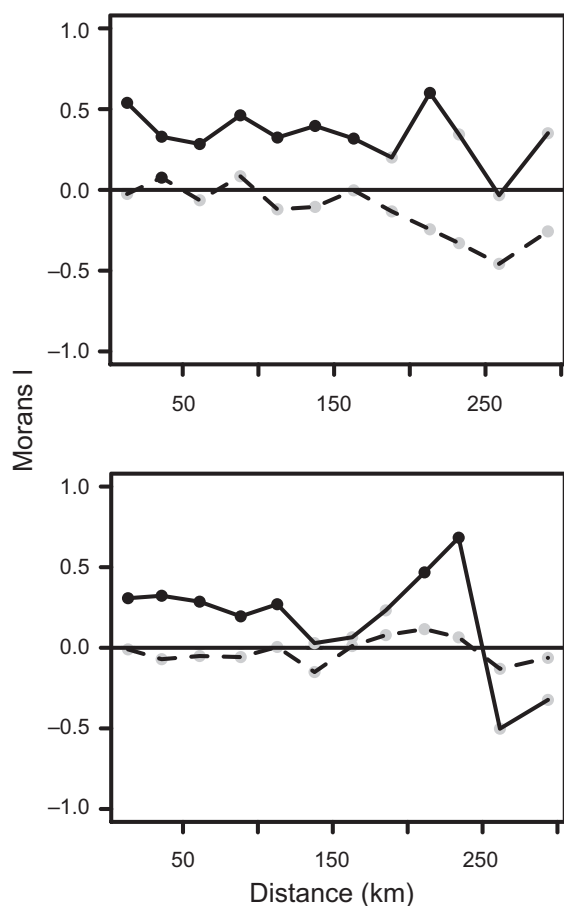
Fig. 4 Spatial correlograms for plant (upper plot) and cladoceran (lower plot) species richness (SR), showing Moran's I index for each distance class (25 km). The solid and dashed lines represent the spatial autocorrelation for the raw and residual (resulting from generalised linear models) SR, respectively.

were not. This means that the relationships with environmental and connectivity variables reflected in the model sufficed to explain the spatial autocorrelation in SR.

Discussion

We present a cross-scale analysis in lentic habitats that identified TP (a key determinant of lake productivity) as the main environmental driver of SR, which peaked at intermediate TP concentration for both aquatic plants and cladocerans. Conductivity was also a common predictor of SR of both groups, although it was more important for cladocerans (accounting for 7.1% of total variance; negative relationship) than for aquatic plants (2.2% of total variance; unimodal relationship). Other environmental variables were also related to SR: aquatic plant richness increased with increasing water transparency, and cladoceran richness decreased with increasing lake depth. The positive relationship of habitat connec-

tivity with SR, on the other hand, was significant but less important in explaining SR. The limited importance of connectivity on this small spatial scale and the lack of spatial autocorrelation in SR (after correcting for the environmental factors) provided evidence for little dispersal limitation within regions in both aquatic plants and cladocerans. However, at a larger scale, the relatively strong relationship of region identity with SR (unexplained by environmental variables) indicates that biogeographic factors and dispersal limitation may act at the continental scale. Moreover, the particular environments of each region also drive regional differences in SR and contribute to reinforce biogeographic patterns.

The concentration of phosphorus, the main environmental driver of SR in this study, is an essential nutrient in aquatic habitats and considered to be a surrogate for productivity (Dodson *et al.*, 2000; Jeppesen *et al.*, 2000; Chase & Leibold, 2002). Indeed, unimodal and negative relationships between TP and SR of macrophytes and zooplankton have been usually shown to be dependent on primary productivity, specifically phytoplankton biomass (Declerck *et al.*, 2005, 2007). However, high TP levels might also promote the dominance of non-edible phytoplankton, limiting the production of large cladocerans and causing a concomitant reduction in cladoceran SR (e.g. Ghadouani, Pinel-Alloul & Prepas, 2006).

The covariation of the SR of both groups, aquatic plants and cladocerans, with its main environmental determinant (TP) is probably mediated by TP's effect on water turbidity, through the chain of effects generally hypothesised in eutrophication processes: increased phytoplankton growth causes increased water turbidity, which results in decreased abundance and/or SR of aquatic plants and, through cascading effects between predatory/zooplanktivorous fish, a reduction in zooplankton abundance and/or SR (Jeppesen *et al.*, 1999; note, however, that abundance and SR might not respond in the same way). This mechanism is supported by the existence of a significant relationship of water transparency (Secchi disc depth) with aquatic plant richness but not cladoceran richness. However, because such an effect would be mediated by the aquatic plants, one should expect a significant effect of plant richness on cladoceran richness (i.e. cross-taxon congruence) even after accounting for the effect of lake productivity (TP) and water turbidity (Secchi disc depth). The weak relationship between cladoceran and plant SR, which disappears when region and TP effects are considered, seems therefore paradoxical. However, the unimodal relationship between TP and cladoceran richness might be driven by the unimodal response of aquatic plant

abundance, rather than its SR, to TP (as also suggested by De Meester *et al.*, 2006). Increased plant abundance (rather than SR) might provide more shelter to zooplankton while maintaining the water column in a clean, transparent state (Strand & Weisner, 2001; Declerck *et al.*, 2007). For example, in eutrophic ponds (high TP), extensive monospecific stands of aquatic plants may host diverse communities of littoral phytophilous zooplankton, living pelagically or lightly attached to the plants, thereby maintaining a high microinvertebrate biodiversity (Van Onsem, De Backer & Triest, 2010). At the other extreme, a diverse community of aquatic plants may thrive in relatively oligotrophic lakes (benefited by the high water transparency and access to nutrients from the sediment), while cladocerans show limited abundance and SR owing to limited food availability in the water column (such as observed in many Italian and Norwegian lakes in our study). In short, while the overall, unimodal response to TP may be comparable for the SR of aquatic plants and cladocerans across different lakes, it does not necessarily need to be consistent between the two taxonomic groups at an individual lake level.

Besides the effect of TP, a few additional predictors (mainly conductivity and pH) explained a considerable fraction of the variation in local SR, although these had different effects on the two functional groups. Conductivity had a quadratic relationship with plant richness, which may reflect the positive effect of inorganic nutrient availability (e.g. calcium, which was highly correlated with conductivity) at low salinity, and the negative effect of osmotic stress at high salinity (Lacoul & Freedman, 2006). On the other hand, conductivity had a strongly negative, linear relationship with cladoceran richness, possibly owing to the progressive dominance of a few species with broader osmotic tolerance (e.g. Green *et al.*, 2005). pH showed the opposite pattern, with cladoceran richness peaking at neutral pH, whereas plant richness increased linearly with increasing pH. The latter relationship probably reflects a dual cause–effect relationship: a decrease in plant SR in acidic-lake communities, dominated by one or a few specialised species (e.g. *Juncus bulbosus*), and an increase in pH caused by strong photosynthetic activity in lakes with abundant and diverse aquatic plant assemblages.

Finally, the decrease in cladoceran SR with lake depth may be explained by the negative relationship between mixing depth and zooplankton biomass (reported, for example, by Berger *et al.*, 2006) and the absence of littoral taxa in deep lakes. In fact, some of the positive relationships between lake depth and SR reported in the literature may be due to sampling biases (Keller & Conlon, 1994), as benthic and littoral zooplankton rarely appear in pelagic zones, but pelagic species frequently appear in littoral zones (Walseng *et al.*, 2006).

Despite the strong effect of several environmental variables (largely TP), regional variation still accounted for a considerable proportion of variance in SR (similar to that accounted for by the whole set of environmental variables). Such an effect could reflect, at least in part, the influence of latitude-related variables that vary at a large scale, which are often used to delimit biogeographic regions and explain broad diversity patterns. For example, climatic variation has been suggested to control a trade-off between dispersal and ecological specialisation that in turn controls species diversity (Jocque *et al.*, 2010). In addition, limited inter-regional dispersal might also determine the observed differences in regional and endemic SR among regions and contribute to explain the variation in local SR. Overall, both the regionally structured environment and limited dispersal at the continental scale (biogeographic processes) might promote the formation of the observed biogeographic patterns.

Dispersal has also been shown to affect the SR of aquatic organisms, including zooplankton (e.g. Dodson, 1992). However, those taxa that rely on the production of propagules for dispersal (such as aquatic plants and cladocerans) often show high rates of dispersal and colonisation (e.g. Santamaría, 2002; Louette & De Meester, 2005). Our results show that habitat connectivity at the small scale (10–20 km) had weak relationships with plant and cladoceran SR ($\leq$2.2% of the total variance). Moreover, after accounting for this relationship, there was no apparent dispersal limitation up to 300 km (maximum distance separating two lakes within each region), because no spatial autocorrelation was detected in the GLM's residuals. Other studies have also suggested that the effect of dispersal depends on scale (Shurin *et al.*, 2000), becoming evident at long rather than short distances (King *et al.*, 2002; Viana *et al.*, 2013). Mechanistic models of propagule dispersal by waterbirds have shown that the propagules of aquatic plants and zooplankton may often be dispersed over distances up to 100 km, and occasional dispersal may take place up to hundreds of kilometres (>200 km for zooplankton and >1000 km for plants; Viana *et al.*, 2013). Insights from studies of community structure also suggest that zooplankton propagules can disperse over distances of the order of 100 km and show uniform dispersal within geographic areas of 300 000 km^2 (Pinel-Alloul, Niyonsenga & Legendre, 1995). Similarly, population genetic studies

of the aquatic plant *Potamogeton pectinatus* suggest that dispersal is more or less uniform at distances of up to about 150–200 km, beyond which geographic separation becomes limiting (King *et al.*, 2002).

Preliminary analyses, in the absence of MI procedures, showed that the deletion of cases would have caused both a significant loss of statistical power (to identify important drivers of taxon richness) and inconsistent results between different groups (plants and zooplankton) and taxonomical ranks (data not shown). We therefore encourage the use of imputation methods that account for the variability of imputed values for analysing cross-scale (or multiregion) data, in which missing data are common features, with the aim of identifying general drivers of biodiversity distribution. Furthermore, for studies on taxon richness for which Poisson regression is often a suitable type of model, likelihood ratio tests for multiple imputed data are useful tools to test the significance and contribution of the predictors' effect. The identification of common important drivers (TP and conductivity) is important for conservation purposes and for understanding the underlying causes of cross-taxon congruence. In our case, cross-taxon congruence was determined by extrinsic factors rather than by direct relationships between the groups. Indeed, the low congruence in SR often reported for lentic aquatic taxa, for example, across trophic levels (Declerck *et al.*, 2005; Longmuir, Shurin & Clasen, 2007; Heino, 2010), suggests the influence of complex interactions and/or specific predictors. The independent and interaction effects accounted for by either the GLM or the BRT analyses presented here may therefore underlie the idiosyncratic variation frequently reported for aquatic SR, particularly when it is surveyed at continental scales.

In conclusion, our work suggests that the SR of aquatic, passively dispersed organisms is primarily determined by local environmental conditions and biogeographic factors. Ecosystem characteristics driven by TP (such as productivity) explained a large proportion of the spatial variation in SR of aquatic plants and cladocerans, and probably drove their covariation via direct and/or indirect effects. On the other hand, dispersal limitation only explained a small fraction of the spatial variation in SR within regions (<300 km). Our results therefore suggest that the environment controls SR at all scales (environmental variables were spatially correlated at all scales and significantly affected SR), while dispersal limitation may become important at large scales (>300 km), strengthening the differences between regions and reinforcing biogeographic patterns.

References

Alahuhta J. & Heino J. (2013) Spatial extent, regional specificity and metacommunity structuring in lake macrophytes. *Journal of Biogeography*, **40**, 1572–1582.

Berger S.A., Diehl S., Kunz T.J., Albrecht D., Oucible A.M. & Ritzer S. (2006) Light supply, plankton biomass, and seston stoichiometry in a gradient of lake mixing depths. *Limnology and Oceanography*, **51**, 1898–1905.

Bohanak A.J. & Whiteman H.H. (1999) Dispersal of the fairy shrimp *Branchinecta coloradensis* (Anostraca): effects of hydroperiod and salamanders. *Limnology and Oceanography*, **44**, 487–493.

Bornette G. & Puijalon S. (2011) Response of aquatic plants to abiotic factors: a review. *Aquatic Sciences*, **73**, 1–14.

Brendonck L. & De Meester L. (2003) Egg banks in freshwater zooplankton: evolutionary and ecological archives in the sediment. *Hydrobiologia*, **491**, 65–84.

Brochet A.-L., Gauthier-Clerc M., Guillemain M., Fritz H., Waterkeyn A., Baltanás Á. *et al.* (2010) Field evidence of dispersal of branchiopods, ostracods and bryozoans by teal (Anas crecca) in the Camargue (southern France). *Hydrobiologia*, **637**, 255–261.

Cáceres C.E. & Soluk D.A. (2002) Blowing in the wind: a field test of overland dispersal and colonization by aquatic invertebrates. *Oecologia*, **131**, 402–408.

Capers R.S., Selsky R. & Bugbee G.J. (2010) The relative importance of local conditions and regional processes in structuring aquatic plant communities. *Freshwater Biology*, **55**, 952–966.

Catalan J., Barbieri M.G., Bartumeus F., Bitušík P., Botev I., Brancelj A. *et al.* (2009) Ecological thresholds in European alpine lakes. *Freshwater Biology*, **54**, 2494–2517.

Chao A. (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, **43**, 783–791.

Chase J.M. & Leibold M.A. (2002) Spatial scale dictates the productivity-biodiversity relationship. *Nature*, **416**, 427–430.

Colwell R.K. (2013) *EstimateS: Statistical Estimation of Species Richness and Shared Species from Samples. Version 9. User's*

Guide and Application. Available at: http://purl.oclc.org/estimates.

Colwell R.K., Mao C.X. & Chang J. (2004) Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology*, **85**, 2717–2727.

Cottenie K. & De Meester L. (2003) Connectivity and cladoceran species richness in a metacommunity of shallow lakes. *Freshwater Biology*, **48**, 823–832.

Cottenie K. & De Meester L. (2004) Metacommunity structure: synergy of biotic interactions as selective agents and dispersal as fuel. *Ecology*, **85**, 114–119.

Davies K.F., Melbourne B.A. & Margules C.R. (2001) Effects of within- and between-patch processes on community dynamics in a fragmentation experiment. *Ecology*, **82**, 1830–1846.

De Meester L., Declerck S., Janse J.H., Dagevos J., Portielje R., Lammens E. *et al.* (2006) Biodiversity in European Shallow Lakes: a Multilevel-Multifactorial Field Study. In: *Wetlands: Functioning, Biodiversity Conservation, and Restoration* (Eds R. Bobbink, B. Beltman, J.T.A. Verhoeven & D.F. Whigham), pp. 149–167. Ecological Studies. Springer, Berlin, Heidelberg.

Declerck S., Vandekerkhove J., Johansson L., Muylaert K., Conde-Porcuna J.M., Van Der Gucht K. *et al.* (2005) Multi-group biodiversity in shallow lakes along gradients of phosphorus and water plant cover. *Ecology*, **86**, 1905–1915.

Declerck S., Vanderstukken M., Pals A., Muylaert K. & De Meester L. (2007) Plankton biodiversity along a gradient of productivity and its mediation by macrophytes. *Ecology*, **88**, 2199–2210.

Dodson S. (1992) Predicting crustacean zooplankton species richness. *Limnology and Oceanography*, **37**, 848–856.

Dodson S.I., Arnott S.E. & Cottingham K.L. (2000) The relationship in lake communities between primary productivity and species richness. *Ecology*, **81**, 2662–2679.

Elith J., Leathwick J.R. & Hastie T. (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*, **77**, 802–813.

Figuerola J. & Green A.J. (2002) Dispersal of aquatic organisms by waterbirds: a review of past research and priorities for future studies. *Freshwater Biology*, **47**, 483–494.

Flössner D. (1972) *Krebstiere, Crustacea. Kiemen- und Blattfüsser, Branchiopoda. Fischläuse, Branchiura*. Tierwelt Deutschlands 60, Jena.

Flössner D. (2000) *Die Haplopoda und Cladocera (ohne Bosminidae) Mitteleuropas*. Backhuys Publishers, Leiden.

Forbes A.E. & Chase J.M. (2002) The role of habitat connectivity and landscape geometry in experimental zooplankton metacommunities. *Oikos*, **96**, 433–440.

Frey D.G. (1960) The ecological significance of cladoceran remains in lake sediments. *Ecology*, **41**, 684–699.

Frisch D., Cottenie K., Badosa A. & Green A.J. (2012) Strong spatial Influence on colonization rates in a pioneer zooplankton metacommunity. *PLoS One*, **7**, e40205.

Ghadouani A., Pinel-Alloul B. & Prepas E.E. (2006) Could increased cyanobacterial biomass following forest harvesting cause a reduction in zooplankton body size structure? *Canadian Journal of Fisheries and Aquatic Sciences*, **63**, 2308–2317.

Green A.J., Fuentes C., Moreno-Ostos E. & Rodrigues Da Silva S.L. (2005) Factors influencing cladoceran abundance and species richness in brackish lakes in Eastern Spain. *Annales de Limnologie – International Journal of Limnology*, **41**, 73–81.

Heino J. (2010) Are indicator groups and cross-taxon congruence useful for predicting biodiversity in aquatic ecosystems? *Ecological Indicators*, **10**, 112–117.

Hessen D.O., Faafeng B.A., Smith V.H., Bakkestuen V. & Walseng B. (2006) Extrinsic and intrinsic controls of zooplankton diversity in lakes. *Ecology*, **87**, 433–443.

Jeppesen E., Jensen J., Søndergaard M. & Lauridsen T. (1999) Trophic dynamics in turbid and clearwater lakes with special emphasis on the role of zooplankton for water clarity. *Hydrobiologia*, **408–409**, 217–231.

Jeppesen E., Jensen J.P., Søndergaard M., Lauridsen T., Pedersen L.J. & Jensen L. (1997) Top-down control in freshwater lakes: the role of nutrient state, submerged macrophytes and water depth. *Hydrobiologia*, **342–343**, 151–164.

Jeppesen E., Peder Jensen J., Søndergaard M., Lauridsen T. & Landkildehus F. (2000) Trophic structure, species richness and biodiversity in Danish lakes: changes along a phosphorus gradient. *Freshwater Biology*, **45**, 201–218.

Jocque M., Field R., Brendonck L. & De Meester L. (2010) Climatic control of dispersal–ecological specialization trade-offs: a metacommunity process at the heart of the latitudinal diversity gradient? *Global Ecology and Biogeography*, **19**, 244–252.

Keller W. & Conlon M. (1994) Crustacean Zooplankton Communities and Lake Morphometry in Precambrian Shield Lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, **51**, 2424–2434.

King R.A., Gornall R.J., Preston C.D. & Croft J.M. (2002) Population differentiation of *Potamogeton pectinatus* in the Baltic Sea with reference to waterfowl dispersal. *Molecular Ecology*, **11**, 1947–1956.

Lacoul P. & Freedman B. (2006) Environmental influences on aquatic plants in freshwater ecosystems. *Environmental Reviews*, **14**, 89–136.

Leibold M.A., Holyoak M., Mouquet N., Amarasekare P., Chase J.M., Hoopes M.F. *et al.* (2004) The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters*, **7**, 601–613.

Lieder U. (1983) Revision of the Genus *Bosmina* Baird, 1845 (Crustacea, Cladocera). *Internationale Revue der gesamten Hydrobiologie und Hydrographie*, **68**, 121–139.

Lieder U. (1996) *Crustacea Cladocera Bosminidae. Susswasser Fauna von Mitteleuropa. Band 8/Heft 2-3*. Fisher Verlag, Stuttgart.

Longmuir A., Shurin J.B. & Clasen J.L. (2007) Independent gradients of producer, consumer, and microbial diversity in lake plankton. *Ecology*, **88**, 1663–1674.

Louette G. & De Meester L. (2005) High dispersal capacity of cladoceran zooplankton in newly founded communities. *Ecology*, **86**, 353–359.

Magrach A., Larrinaga A.R. & Santamaría L. (2012) Effects of matrix characteristics and interpatch distance on functional connectivity in fragmented temperate rainforests. *Conservation Biology*, **26**, 238–247.

Pinel-Alloul B., Niyonsenga T. & Legendre P. (1995) Spatial and environmental components of freshwater zooplankton structure. *Ecoscience*, **2**, 1–19.

R Development Core Team (2012) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available at: http://www.R-project.org.

Ricklefs R.E. (1987) Community diversity: relative roles of local and regional processes. *Science*, **235**, 167–171.

Santamaría L. (2002) Why are most aquatic plants widely distributed? Dispersal, clonal growth and small-scale heterogeneity in a stressful environment. *Acta Oecologica*, **23**, 137–154.

Sayer C.D., Burgess A.M.Y., Kari K., Davidson T.A., Peglar S., Yang H. *et al.* (2010) Long-term dynamics of submerged macrophytes and algae in a small and shallow, eutrophic lake: implications for the stability of macrophyte-dominance. *Freshwater Biology*, **55**, 565–583.

Scheffer M., Hosper S.H., Meijer M.L., Moss B. & Jeppesen E. (1993) Alternative equilibria in shallow lakes. *Trends in Ecology & Evolution*, **8**, 275–279.

Shurin J. (2000) Dispersal limitation, invasion resistance and the structure of pond zooplankton communities. *Ecology*, **81**, 3074–3086.

Shurin J., Cottenie K. & Hillebrand H. (2009) Spatial autocorrelation and dispersal limitation in freshwater organisms. *Oecologia*, **159**, 151–159.

Shurin J.B., Havel J.E., Leibold M.A. & Pinel-Alloul B. (2000) Local and regional zooplankton species richness: a scale-independent test for saturation. *Ecology*, **81**, 3062–3073.

Strand J.A. & Weisner S.E.B. (2001) Dynamics of submerged macrophyte populations in response to biomanipulation. *Freshwater Biology*, **46**, 1397–1408.

Van Buuren S. & Groothuis-Oudshoorn K. (2011) mice: multivariate imputation by chained equations in R. *Journal of Statistical Software*, **45**, 67.

Van Leeuwen C.H.A., Van Der Velde G., Van Groenendael J.M. & Klaassen M. (2012) Gut travellers: internal dispersal of aquatic organisms by waterfowl. *Journal of Biogeography*, **39**, 2031–2040.

Van Onsem S., De Backer S. & Triest L. (2010) Microhabitat-zooplankton relationship in extensive macrophyte vegetations of eutrophic clear-water ponds. *Hydrobiologia*, **656**, 67–81.

Vanschoenwinkel B., Waterkeyn A., Vandecaetsbeek T.I.M., Pineau O., Grillas P. & Brendonck L.U.C. (2008) Dispersal of freshwater invertebrates by large terrestrial mammals: a case study with wild boar (Sus scrofa) in Mediterranean wetlands. *Freshwater Biology*, **53**, 2264–2273.

Viana D.S., Santamaría L., Michot T.C. & Figuerola J. (2013) Migratory strategies of waterbirds shape the continental-scale dispersal of aquatic organisms. *Ecography*, **36**, 430–438.

Walseng B.R., Hessen D.O., Halvorsen G. & Schartau A.K. (2006) Major contribution from littoral crustaceans to zooplankton species richness in lakes. *Limnology and Oceanography*, **51**, 2600–2606.

Williams D. (2001) *The ecology of temporary waters*. Blackburn Press, Caldwell, NJ.

Wolf H. & Mort M. (1986) Inter-specific hybridization underlies phenotypic variability in Daphnia populations. *Oecologia*, **68**, 507–511.

Supporting Information

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

Appendix S1. Sites and species lists.

Appendix S2. Detailed description of statistical methods.

Appendix S3. Supporting tables and figures.

Appendix S4. R code to perform likelihood ratio tests to generalised linear models (GLM) with Poisson error distributions fitted to multiply imputed data.

(Manuscript accepted 11 June 2014)