Dispersal distance and the pool of taxa, but not barriers, determine the colonisation of restored river reaches by benthic invertebrates

JONATHAN D. TONKIN, STEFAN STOLL, ANDREA SUNDERMANN AND PETER HAASE

Department of River Ecology and Conservation, Senckenberg Research Institute and Natural History Museum Frankfurt, Gelnhausen, Germany

SUMMARY

1. Restoration is an increasingly central theme in river ecology. Recent studies have highlighted the importance of the species pool in the surrounding river network for determining colonisation of restored river reaches by both invertebrates and fish.

2. Using a comprehensive data set of 21 river restoration sites and 292 sites in the immediate surroundings, we tested the influence of distance to nearest colonist source on invertebrate colonisation based on a comparison of river network distances and Euclidean distances, expecting river network distances would better align with colonisation rates. We then assessed the importance of dispersal distance in relation to several other parameters, such as the number and intensity of barriers along the river network, surrounding taxon pool occupancy rate, physical characteristics of the restored sites and restoration techniques used in determining colonisation of commonly occurring benthic invertebrates.

3. We hypothesised that (i) distance would be critical, with colonisation of restored sites declining with increasing distance; (ii) barriers between these sites would be a minor, but taxon-specific, influence on the colonisation; and (iii) the higher the regional pool occupancy rate of a certain taxon, the higher its probability of presence at a restored site.

4. Overall, taxon pool occupancy rate was the most important driver of colonisation likelihood, followed by distance to nearest source, with the first kilometre particularly important. The effect of barriers was minor but significant, and taxon identity had no effect on the predictive ability of the model. Factors associated with the restoration projects such as techniques used and physical characteristics had minor influences, being completely outweighed by taxon pool and dispersal-related factors.

5. To gauge the likelihood of successful outcomes of habitat restoration projects, we suggest it is important to assess regional taxon pools and ensure distances between healthy populations are minimised. These results clearly emphasise the importance of spatial planning for restoration projects.

Keywords: dispersal, macroinvertebrates, restoration, species pool, streams

Introduction

River restoration has increasingly been in the public and scientific spotlight in recent decades (Bernhardt *et al.*, 2005; Palmer *et al.*, 2005; Dudgeon *et al.*, 2006), yet, despite significant expenditures, projects often do not meet their goals of restoring biodiversity (Palmer, Menninger & Bernhardt, 2010; Louhi *et al.*, 2011; Haase *et al.*, 2013). Lack of attention to ecological theory (Lake, Bond & Reich, 2007) and proper evaluation in the years following restoration (Bernhardt *et al.*, 2005) have been identified as key issues associated with restoration failures. For these

Correspondence: Jonathan D. Tonkin, Department of River Ecology and Conservation, Senckenberg Research Institute and Natural History Museum Frankfurt, Clamecystrasse 12, 63571 Gelnhausen, Germany. E-mail: jonathan.tonkin@senckenberg.de

reasons, it is important to formulate clear and focused goals prior to restoration projects taking place (Jähnig *et al.*, 2011).

The large-scale channelisation of rivers worldwide has led to much of the focus of restoration effort being on improving local-scale physical habitat condition (Palmer, 2009). Although widely applied, this focus on improving habitat to drive ecological response is based on a myth that has underpinned much of river restoration ecology (Hilderbrand, Watts & Randle, 2005) and has not led to positive outcomes if the stressor causing reduced ecological health has itself not been addressed (Palmer *et al.*, 2010). In conjunction with this focus on restoration of physical habitats is the fact that river restorations often occur in isolated patches at small spatial scales (Bond & Lake, 2003; Bernhardt *et al.*, 2005), and usually within highly disturbed environments. The dominant influence of catchment-scale processes such as land-use degradation often swamps small-scale restoration projects (Bohn & Kershner, 2002; Bond & Lake, 2003; Pretty *et al.*, 2003), yet there remains a paucity of catchment-scale restoration projects (Holl, Crone & Schultz, 2003).

Along with inappropriate scale, constraints on colonisation from the surrounding species pool, such as distance to source and in-stream barriers, are key areas requiring attention for restoration projects (Bond & Lake, 2003; Sundermann, Stoll & Haase, 2011b; Stoll *et al.*, 2013). In fact, dispersal is a critical factor driving responses of stream invertebrates to restoration (Riley & Fausch, 1995; Parkyn & Smith, 2011), which is not surprising given the importance of dispersal in metapopulation persistence within landscapes (Hanski & Gilpin, 1997; Grönroos *et al.*, 2013). In isolated restored sites, when surrounding source populations are scarce or distant, species' dispersal capacity is likely to be crucial (Cushing & Gaines, 1989; Milner *et al.*, 2000). Nonetheless, a large proportion of stream invertebrates have a terrestrial adult stage, which assists with dispersal, and should theoretically increase the chances of dispersal overland between adjacent channels. In fact, active dispersers with terrestrial adult stages can track environmental variability better than more passive dispersers (Grönroos *et al.*, 2013). However, many aquatic insects are poor fliers and thus are dependent on navigating along river channels (e.g. Petersen *et al.*, 2004; Winterbourn *et al.*, 2007) with some indication that wind may assist dispersal (Briers *et al.*, 2004) and can be hampered by in-stream barriers such as culverts and bridges narrowing flight paths in smaller streams (e.g. Blakely *et al.*, 2006).

The dendritic nature of stream networks (Campbell Grant, Lowe & Fagan, 2007) means dispersal along a river network typically involves much larger distances than simply the overland distance between sites. Given the dependence on the river network for dispersal, hydrologic connectivity has been recognised as one of the key areas requiring restoration effort (Bernhardt & Palmer, 2011) and consideration for restoration projects to have a chance of success (Bond & Lake, 2003). Fragmentation, through the installation of physical barriers, is one of the key issues facing river managers worldwide and has received intensive research focus (Nilsson *et al.*, 2005; Fullerton *et al.*, 2010; Liermann *et al.*, 2012). Indeed, barrier removal and reconnecting fragmented rivers have been a strong focus of research for some time (Bednarek, 2001; Doyle *et al.*, 2005; DeGraff & Evans, 2013).

To determine colonisation rates of restored river reaches by benthic macroinvertebrates, we focus on two main approaches incorporating dispersal constraints. Using a comprehensive data set of 21 river restoration sites and 292 sites in the surroundings, we first tested the influence of distance to nearest colonist source on invertebrate colonisation based on a comparison of river network distances and Euclidean distances. Given the dendritic nature of river networks and the fact that many species follow these networks during dispersal (e.g. Petersen *et al.*, 2004), we hypothesise that river network distances will much better align with colonisation rates than Euclidean distances. Second, we tested the importance of dispersal distance in relation to several other parameters, such as the number and intensity of barriers along the river network, surrounding species pool occupancy rate, physical characteristics of the restored sites, habitat restoration techniques used and restoration cost in determining colonisation of commonly occurring benthic invertebrates. Using a modelling approach, we test the hypothesis that the distance to nearest colonist source will be a critical factor driving colonisation of restored reaches, with colonisation rates declining with increasing distance. Due to the high mobility of many stream invertebrates, particularly those with aerial adult stages, we hypothesise that the number of barriers between these sites will be a minor, but taxon-specific, influence on the colonisation of individual restored sites. Further, we hypothesise that the higher the regional pool occupancy rate of a certain taxon, the higher its probability of presence at a restored site. The results of this study will provide quantitative information for the design and implementation of river restoration projects across landscapes seeking to maximise the

potential of restoration projects, as well as for maintenance of healthy metapopulations across landscapes affected by disturbances.

Methods

Restoration projects

We assessed the probability of benthic macroinvertebrate colonisation of reach-scale restoration projects in 21 second- to seventh-order rivers in Hesse, Germany. The restoration projects assessed in this study are the same sites assessed by Sundermann *et al.* (2011b), with the exclusion of the three North Rhine-Westphalia sites. However, Sundermann *et al.* (2011b) only assessed Euclidian distances and did not account for in-stream passage barriers, leaving room for a more detailed assessment of these dispersal-related constraints. The projects we investigated included a diverse range of combined habitat improvements ranging from placement of woody material to reconnection with backwaters (Table 1). The goal of these restoration projects was improvement of local habitat diversity and hydromorphological conditions to restore biota to more natural

Table 1 Variables used in the analyses. Goals represent the goals associated with the restoration project, and measures are the specific measures used to perform the restoration

Category	Subcategory	Variable
Restoration sites	Parameter	Size of catchment (km^2)
		Length of restored section (km)
		Invested costs (Euros m^{-1})
		Time since restoration (years)
		Stream order
		Land use
	Goals	Increasing physical heterogeneity
		Flood prevention
		Longitudinal connectivity
		Lowering of entrenchment depth
	Measures	Removal of bank fixation
		Wood placement
		Installation of flow deflectors
		Elongation of river length
		Creating a new water course
		Creation of multiple channels
		Extensification of land use
		Re-connection of back waters
Surrounding sites	Route	Network distance
		Euclidian distance
		Barriers (upstream)
		Barriers (downstream)
		Barriers total
	Taxa	Taxon ID
		Taxon pool occupancy rate

conditions. Restorations were intensive, representing a major physical alteration to the habitat, thus denuding local habitats and resetting communities in the process. More details about each restoration project can be found in Sundermann *et al.* (2011b).

Mean catchment size of these restoration sites was 453.2 km^2, ranging from 11.3 to 2373 km^2. The average length of restoration site was 0.99 km, ranging from 0.2 to 2.6 km. Project costs varied between €12 000 and €1 700 000, averaging €351 900, and cost per km ranged between €22 500 and €1 141 750, with a mean of €386 602. The dominant catchment land use was calculated using settlements, arable land, meadows and pastures, forest and waterbodies as categories. We determined stream order of the restoration sites with 1 : 50 000 topographic maps (TOP 50 Hessen; Geogrid-Viewer 1.1, EADS Deutschland GmbH, Munich, Germany).

Surrounding site selection and barrier quantification

Using Euclidian distances and grouping surrounding monitoring sites into 5-km classes, Sundermann *et al.* (2011b) found only the monitoring sites within the first 5-km radius were important with regard to restoration success. However, this does not account for the dendritic nature of river systems (Fagan, 2002), and this 5-km radius may equate to much larger in-stream distances. To assess this in more detail and focus on network distances, we selected all monitoring sites within a 10-km distance of river network.

The distance along the river network between each restoration site and the surrounding monitoring sites was assessed using the Network Analyst tool in ArcMap 10.0 (ESRI Inc., Redlands, CA, U.S.A.). Initially, all monitoring sites within 10 km, both up- and downstream, from each restoration site were selected. Using the routes created with Network Analyst, we used the Linear Referencing tool in ArcMap to quantify the number of barriers along each route and the direction of passing each barrier from the monitoring site to the restored site.

The barriers in these GIS layers range from small culverts and cascades through to large-scale hydroelectric dams. Therefore, rather than tallying all barriers between each monitoring site and restoration site without regard to their type or biotic influence, we fuzzy coded barriers in different manners, including their absence along a route, depending on pre-defined assessments (provided by Hessisches Landesamt für Umwelt und Geologie [HLUG]) of their influences on biota using four different coding systems (Table 2). Barrier presence/absence and

Table 2 Coding system used to enumerate barrier types (including barrier absence), ranging from no barrier to fully impassable barrier used in the analysis. Barrier score was calculated by summing the scores of each barrier crossed. Categories were developed using expert knowledge through visual inspection of the potential effects of barriers in the field (provided by Hessisches Landesamt für Umwelt und Geologie)

Category	A	B	C	D
No barrier	0	0	0	0
Fully passable barrier	0	0	1	1
Somewhat passable barrier	1	1	1	2
Mostly impassable barrier	1	2	1	3
Fully impassable barrier	1	3	1	4

differing levels of passability were classified by HLUG in both upstream and downstream directions in five categories: (i) no barrier; (ii) fully passable barrier; (iii) somewhat passable barrier; (iv) mostly impassable barrier and (v) fully impassable barrier (Table 2). Coding system A counts only barriers that are considered in some way an issue for passage, whereas system C simply tallies all barriers regardless of passability (i.e. 0 or 1; Table 2). System B and D grade barriers regarding their severity of impact on passage (i.e. 0–3 or 0–4; Table 2). Using this fuzzy coding system, we summed all the barriers along a route based on the combined score; thus, the combined total barrier score of each route differed, enabling us to compare their effects.

Benthic macroinvertebrate data

The time between the restoration and sampling date ranged from 1 to 12 years, with a mean of 4.1 years. Benthic macroinvertebrates were sampled at the 21 restoration sites following the official EU Water Framework Directive (WFD) compliant sampling protocols for German streams (Haase *et al.*, 2004a). This involved a multihabitat sampling approach with 20 sample units being taken from each site. Following river monitoring protocols for sampling in Germany, all sampling occurred between March and July in 2007 and 2008. See Sundermann *et al.* (2011b) for more specific details on the sampling and sample processing procedures.

Samples were stored in 70% ethanol and, upon return to the laboratory, were identified to the level proposed by Haase *et al.* (2004a,b), to ensure taxonomic comparability and consistency with EU WFD protocols. Only taxa that were identified to genus or species level were included in the analysis.

To assess the importance of the potential pool of colonist taxa for the restored reaches, we evaluated benthic macroinvertebrate monitoring data in the surrounding

river network. This consisted of 292 sites within 10 km of the 21 restoration sites. Monitoring data were collected by German governmental environmental agencies between 2004 and 2008 and followed the same protocols used at the restoration sites (Haase *et al.*, 2004a Haase *et al.*, 2004b). The surrounding taxon pool occurrence rate was calculated as the number of monitoring sites in the surrounding 10-km area with each taxon found at the restoration sites present, divided by the total number of sites in each surrounding monitoring network.

To ensure adequate regional representation of each taxon, we reduced the original list of taxa occupying either or both restoration or surrounding monitoring sites to 77 taxa. We did this by selecting only those that were found in at least 11 of the 21 10-km buffer zones around restoration sites, regardless of restoration site presence. Only where taxa were present at monitoring sites were routes between restored and surrounding sites considered. Of this final set of taxa, 45 were merolimnic (both aquatic and terrestrial life stages) and 32 hololimnic (fully aquatic lifecycle).

Route determination

To ensure dispersal routes were not duplicated, using the previously selected river network routes, we selected only one occurrence of each taxon in the surrounding network for analysis using three selection methods, giving a total of three routes per taxon per restoration site. First, we selected the route to the nearest colonist pool from the surrounding monitoring sites using river network distance (i.e. the closest site along the river network). Second, we selected the route to the nearest colonist pool along the river network with the lowest number of barriers regardless of distance and used the secondary criteria of nearest source if there were multiple routes with the same number of barriers. Third, we used Euclidian distance (i.e. direct line) to select the nearest colonist source across land, disregarding the river network distance. This produced a final number of routes of 1196 (as not all taxa were found in all 21 surrounding pools) and allowed for interpretation of the importance of network cf. Euclidian distance, as well as barriers. Nonetheless, it is important to consider that the nearest site in relation to restored sites with a taxon present was not necessarily the true nearest source, as the monitoring sites with sample data represent a small fraction of the length of river network. Thus, working only with common taxa ensures the maximum accuracy of these sites, given their widespread occurrence rate.

Statistical analyses

All statistical analyses were carried out in R 3.0.2 (R Core Team, 2013). We converted all taxa data for both the monitoring network and restoration sites into presence or absence.

Generalised Linear Modelling (GLM). To assess the influence of distance and number of upstream and downstream barriers between the surrounding source populations, we used multiple logistic regression (GLM) using the 'GLM' function in the R 'stats' package. We used the 'binomial' family due to predicting binomial presence absence data and performed these regressions on each of the three data sets used to select the single colonist sites (i.e. minimum river network distance, minimum Euclidian distance and minimum total barriers). For the minimum river network distance and minimum barrier data, we predicted the probability of presence at a restored site from distance to source, total barrier score for each of the coding systems and their interaction. As using Euclidian distance assumes flight dispersal across land and thus a lowered importance of river barriers, we predicted restoration site presence from the distance value alone when using Euclidian distance data.

Boosted Regression Trees (BRTs). To further explore the effects of distance and barriers in the presence of other factors driving invertebrate colonisation (i.e. presence of a particular taxon) at these sites (shown in Table 1), we used boosted regression trees in the R package 'dismo' (Elith, Leathwick & Hastie, 2008). BRTs originated from machine learning techniques and are considered an advanced form of regression. Rather than producing a single 'best' model as in traditional regression approaches, BRTs use boosting to combine multiple simple models, enhancing their predictive ability (Elith *et al.*, 2008; Buston & Elith, 2011).

As we wanted to select the most important factors, we took a stagewise approach using the gbm.step procedure, and, given we were predicting presence–absence data, we used the 'bernoulli' (i.e. binomial) family loss function. This approach uses cross-validation to estimate the optimal number of trees. The final models were calculated with a bag fraction of 0.5 (introduces stochasticity into the model), learning rate of 0.001 to ensure at least 1000 trees were used, tree complexity of five, and the maximum number of trees set to 10 000. We used 10-fold cross-validation for all the models. Essentially,

the explanatory power of the training model details the model's fit to the training data (that used to build the model), whereas the cross-validated model details its predictive ability on data not used in building the model.

As we kept all taxa as individual entries, we expected taxon identity would dominate the models. Therefore, we ran four different models to compare the importance of taxon identity and location, as well as dispersal mode. Model 1 included taxon identity, restoration location, distance of route, taxon pool occupancy rate (proportion of surrounding sites a taxon was present) and both upstream and downstream barriers. In addition to unmeasured variables, the factor 'location' represented all of the specific variables associated with each restoration site, such as the techniques used, catchment size, and stream order (Table 1). Therefore, given restoration location was synonymous with the restoration techniques, we could not include these variables in addition to 'location'. To assess the importance of restoration methods and other site-specific parameters, we removed the factor 'location' and included all of the characteristics of the restoration project and site in model 2 (Table 1). Model 3 included the same variables as model 2 with taxon identity removed, enabling an assessment of the importance of taxon identity in predictive ability. To assess whether dispersal mode of invertebrates (either aquatic or aerial) was important in determining colonisation probability, we also ran a model with this included. Model 4 was thus identical to model 3 with the addition of whether a taxon had the capability of dispersing aerially or not.

To further assess the importance of regional pool occupancy rate, we used simple linear regression using the 'lm' function to predict the proportion of total restoration sites from the mean proportion of surrounding sites in each regional pool that each taxon was present.

Results

The mean number of barriers along the routes was 3.71 ± 0.13 (SE), ranging from 0 to 34, with more downstream barriers on average (2.43 ± 0.12) than upstream (1.28 ± 0.07). Length along the stream network and Euclidian distance were highly correlated for both the minimum barrier data set (Pearson's $r = 0.94$, $P < 0.0001$) and for the minimum distance data set (Pearson's $r = 0.95$, $P < 0.0001$).

GLM

Probability of occurrence of a taxon at a restoration site declined with both network and Euclidian distance from source and combined total barrier score, as well as both upstream and downstream barrier score individually (Fig. 1; Table 3; Table S1). However, the influence of distance was stronger than barriers (Table 3; Table S1), and both upstream and downstream barriers appeared to have a similar effect (note the different *x*-axis scales; Fig. 1).

There was little difference in the approaches taken for predicting taxa presence from the surrounding taxon pool (Table 3; Table S1). The models using data selected from the nearest source of colonists along the river network slightly outperformed those using the least number of barriers along the route (Table 3; Table S1). The model predicting taxa presence using direct-line distance gave a better (smaller) AIC score than the barrier data

set but not than the river network distance models (Table 3; Table S1).

Furthermore, the four different fuzzy coding systems had negligible influence on the outcome of the models, with coding system D performing the best (Table 3; Table S1). For these reasons, the remainder of analyses were performed with the river network nearest source data, using the barrier coding system D.

Boosted regression trees

Model 1: Taxon identity and location included. Model 1, which included both taxon identity and location but not restoration or other site-specific measures, had a correlation of 0.704 between raw and fitted values of taxon presence at a restored site in the training data and $0.445 \pm 0.026\%$ using cross-validation (Table 4). Taxon identity was overwhelmingly the most important variable in the model with a relative influence of
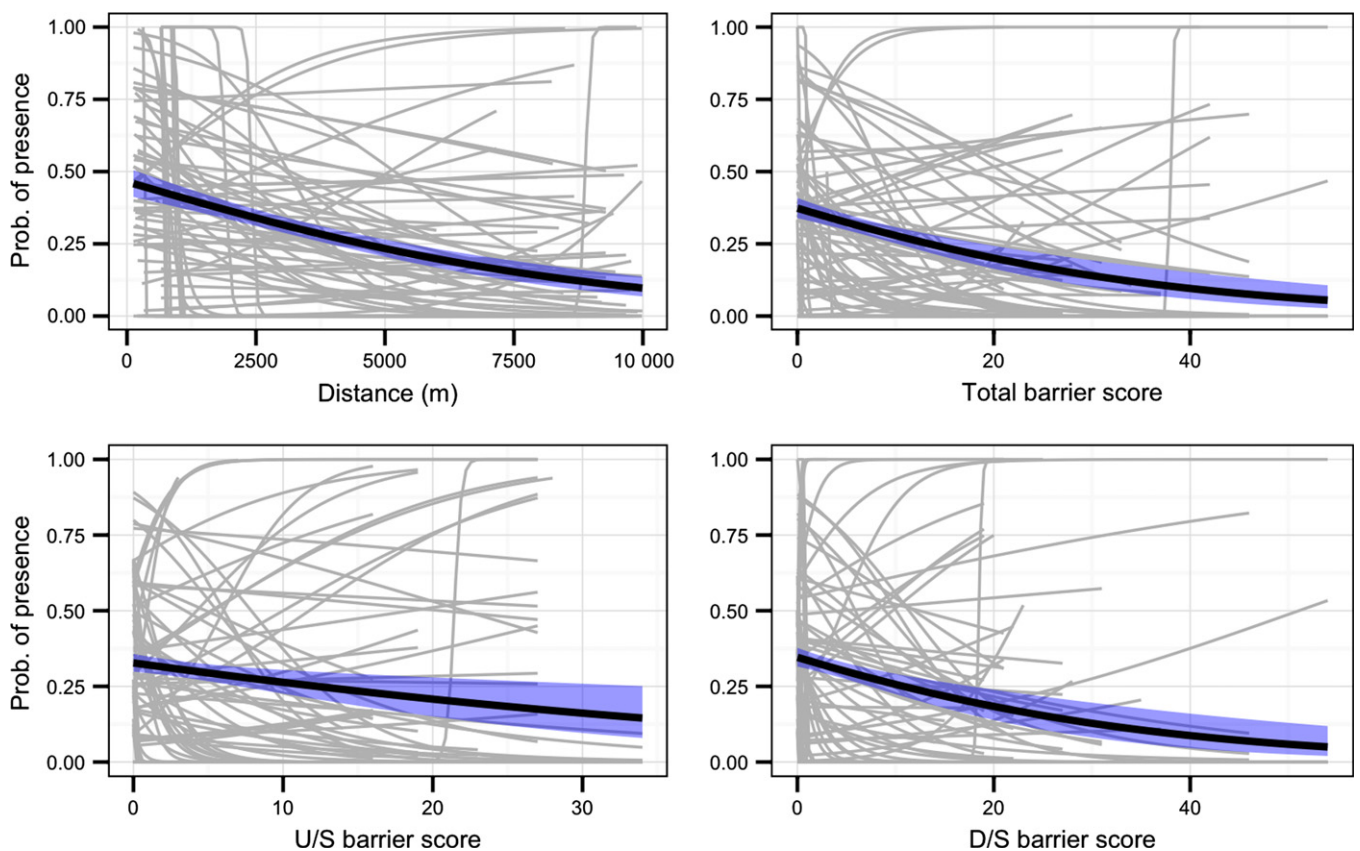

Fig. 1 Logistic regression plots predicting the probability of taxon presence at restored sites as a function of distance to nearest source along the river network, the total barrier score, upstream barrier score and downstream barrier score. Statistical results are presented in Table 3, and more detailed results can be found in Table S1. The central black line is the overall response of all taxa, with the shaded area around the line representing the 95% confidence interval of the prediction. Each of the 77 grey lines represents a single taxon. Taxa data used in these plots are using the nearest colonist source along the river network data outlined in the methods, and barrier score is based on the 'D' barrier coding system outlined in Table 2. U/S, upstream; D/S, downstream.

Table 3 Summary table of the nine individual logistic regression models predicting taxon presence at a restored site as a function of distance and total barrier score. The three data categories represent the three different data sets with regard to selection of the surrounding taxon pool locations. The Euclidian distance model only included distance as it considered across-land dispersal, and thus, in-stream barriers were irrelevant. Code is the scoring method used to assess barriers. The five numbers represent five categories assigned to the barrier: no barrier, passable barrier, somewhat passable barrier, mostly impassable barrier, and fully impassable barrier. * = $P < 0.05$, ** = $P < 0.01$, *** = $P < 0.0001$. 'dev.' = deviance; AIC = Akaike's information criterion. Full results can be found in Table S1

Data	Code	Distance	Barrier	Distance × Barrier	Null dev.	Residual dev.	AIC
Minimum number of barriers	A: 0-0-1-1-1	***	**	*	1479.7	1399.5	1407.5
	B: 0-0-1-2-3	***	**	*	1479.7	1400.3	1408.3
	C: 0-1-1-1-1	***	**	**	1479.7	1401.3	1409.3
	D: 0-1-2-3-4	***	**	**	1479.7	1399.4	1407.4
Minimum river network distance	A: 0-0-1-1-1	***	**	*	1479.7	1393	1401
	B: 0-0-1-2-3	***	**	**	1479.7	1391.7	1399.7
	C: 0-1-1-1-1	***	**	**	1479.7	1391.7	1399.7
	D: 0-1-2-3-4	***	**	**	1479.7	1390.6	1398.6
Minimum Euclidian distance		***	–	–	1479.7	1399.7	1403.7

Table 4 Summary information of the four models developed using boosted regression trees to predict restoration site colonisation probability based on surrounding taxon pool data

Data		Model 1	Model 2	Model 3	Model 4
Training	Number of trees	2700	2550	3500	3650
	Correlation	0.704	0.665	0.546	0.549
	Mean total deviance	1.237	1.237	1.237	1.237
	Mean residual deviance	0.767	0.81	0.947	0.942
	ROC score	0.918	0.896	0.821	0.823
Cross-validated	Correlation	0.445 ± 0.026	0.445 ± 0.013	0.445 ± 0.021	0.441 ± 0.037
	Estimated deviance	1.032 ± 0.024	1.033 ± 0.015	1.047 ± 0.021	1.044 ± 0.033
	ROC score	0.769 ± 0.014	0.775 ± 0.010	0.758 ± 0.010	0.756 ± 0.022

69.9%, followed by location with 17.6%, distance (8.0%) and taxon pool occurrence rate (4.1%; Fig. 2). Upstream barriers and downstream barriers contributed little with relative contributions of 0.3 and 0.2%, respectively (Fig. 2). Given the dominance of taxon identity and location in this model, the only interactions evident were between taxon identity and location, and taxon identity and distance (Table 5).

Model 2: Location excluded. Model 2, which removed location but included restoration site measures, had slightly lower correlation coefficient than model 1 using the training data (0.665) but no difference in cross-validated predictive ability (0.445 ± 0.013; Table 4). Little changed in model two with regard to taxon identity's influence (75.0%; Fig. 2). The next most important drivers were distance (9.6%) and taxon occurrence rate (6.2%; Fig. 2). The most important restoration site factor was cost with a contribution of 4.9%; none of the other restoration site variables registered, and barriers contributed little to the model (upstream: 0.6%; downstream: 0.4%; Fig. 2). Only three interactions in this model had a strength greater than 1, and these all involved taxon

identity (Table 5). The most important interaction was between taxon identity and distance, followed closely by cost of restoration (Table 5).

Model 3: Location and taxon identity excluded. The correlation coefficient of model 3, which excluded both location and taxon identity, was substantially lower for the training data (0.546), but there was little to no difference in its predictive ability according to the cross-validated model (0.445 ± 0.021; Table 4). With the removal of taxon identity, taxon pool occurrence rate became the most important variable with a contribution of 40.8%. Distance followed, contributing 25.4%, with the initial 1 km appearing crucial to colonisation (Figs 2 & 3). Cost was again the most important restoration site factor (8.4%) followed by the time between restoration and sampling (3.8%; Figs 2 & 3). Downstream barriers contributed 3.5% and upstream barriers 2.9% (Figs 2 & 3).

Clearly, the taxon pool occurrence rate improved colonisation chances, whereas distance between nearest source and restoration site decreased them (Fig. 3). In fact, the mean taxon pool occupancy rate across all regions was an excellent predictor of the proportion of
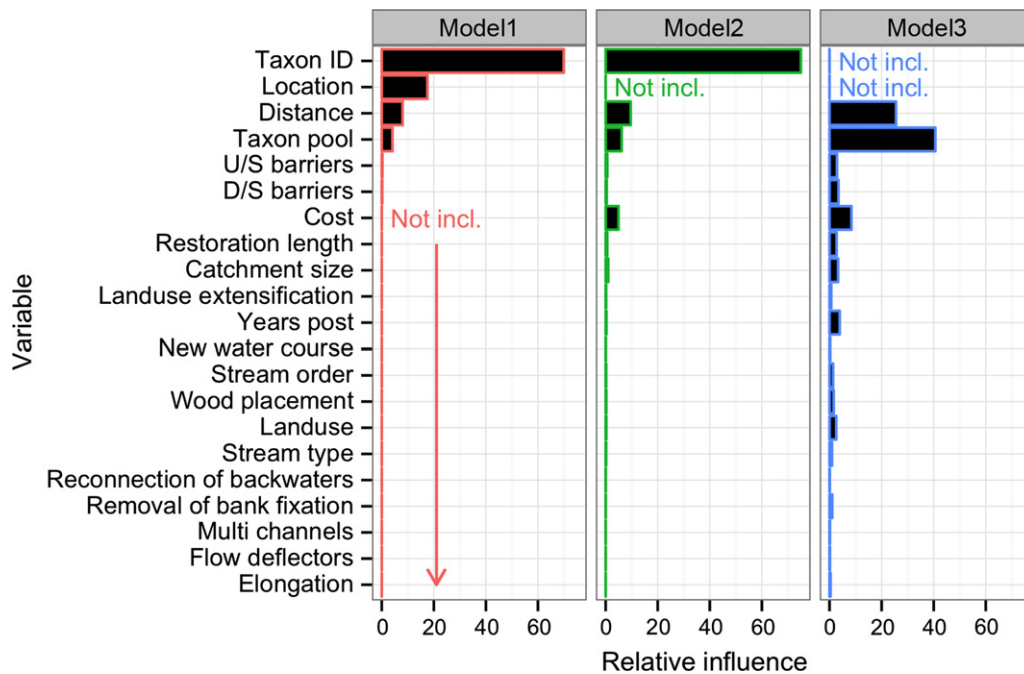
Fig. 2 Relative influence of each variable on each of the first three boosted regression tree models predicting restoration site presence from surrounding taxon pool data. Model 4 is not included due to its similarity to model 3. Not incl., not included in the model. Taxon pool, taxon pool occupancy rate; U/S, upstream; D/S, downstream.

Table 5 Summary of the main interactions between variables in each of the first three boosted regression tree models. Model 4 is not included due to its similarity to model 3. Only interaction sizes greater than 1 are shown. Taxon pool, taxon pool occupancy rate

Rank	Variable 1	Variable 2	Interaction size
Model 1			
1	Location	Taxon ID	32.62
2	Distance	Taxon ID	8.22
Model 2			
1	Distance	Taxon ID	14.96
2	Cost	Taxon ID	14.03
3	Taxon pool	Taxon ID	4.33
Model 3			
1	Taxon pool	Cost	3.22
2	Years post	Distance	2.04
3	Taxon pool	Removal of bank fixation	1.75
4	Wood placement	Distance	1.21
5	Cost	Distance	1.2
6	Taxon pool	Years post	1.05

restoration sites that a taxon had colonised ($r^2 = 0.65$, $F_{1,75} = 140.4$, $P < 0.0001$, $y = -0.10 + 1.26x$; Fig. 4). Cost spent on the restoration improved restoration success, but influence of the remaining factors was too minor to be clearly discernable (Fig. 2). Six interactions in model 3 had an interaction size greater than 1, including years post-restoration and dispersal distance, but these were all relatively minor (Table 5).

Including dispersal mode in the model (model 4) did not improve the prediction rate or alter variable contribution substantially, and dispersal mode only contributed 1.19% to the model; thus, we have not displayed this model.

Discussion

While the regional species pool has been highlighted as an important driver of benthic invertebrate colonisation at restored sites (Lake *et al.*, 2007; Sundermann *et al.*, 2011b), it has not been quantified specifically. Rather than assessing rates of post-hydromorphological restoration colonisation in relation to pre-restored conditions, we assessed the probability of presence of a set of 77 relatively widespread individual taxa. As hypothesised, we found that the colonisation probability of benthic macroinvertebrates at 21 restored river reaches could be successfully predicted (0.45 correlation between raw and fitted values) largely by factors associated with the surrounding regional taxon pool, such as occupancy rate and dispersal distance.

As recently found with fish (Stoll *et al.*, 2013, 2014), and in line with our final hypothesis, the results of the

Fig. 3 Partial dependence plots showing fitted functions of each of the top six variables contributing to model 3 using boosted regression trees, predicting restoration site presence from surrounding taxon pool data. These plots differ from Fig. 1 by displaying each variable's contribution to the model after accounting for the effects of the other variables. Rugs along the x-axis represent the distribution of data in deciles, and values in parentheses represent the relative contribution of each variable. Taxon pool, taxon pool occupancy rate; distance, river network dispersal distance in m; cost, cost in Euros m^{-1}; years post, number of years after restoration when sampling occurred; D/S barriers, downstream barrier score based on coding system D; restoration length, length of restored section in km.
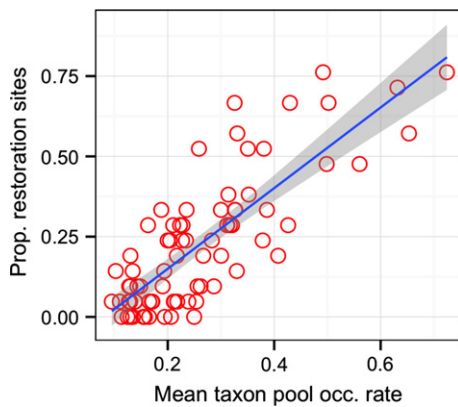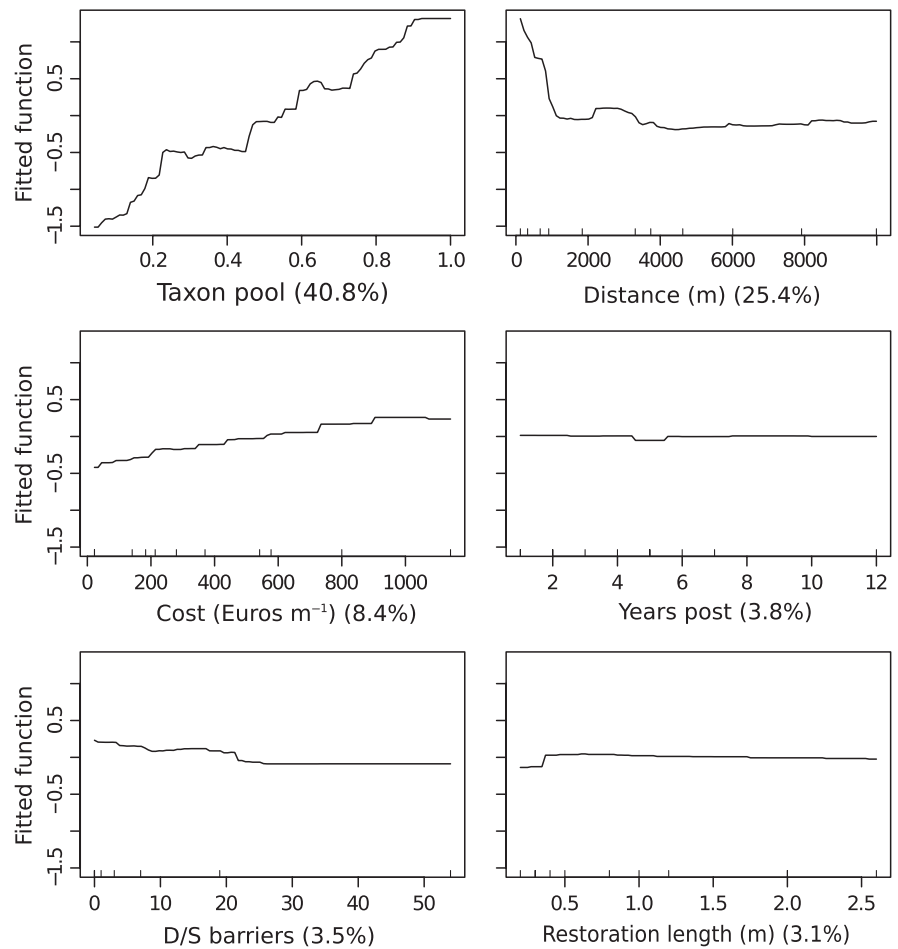

Fig. 4 The proportion of restoration sites colonised by each taxon as a function of the mean occupancy rate across all 21 regional taxon pools. The grey area around the regression line represents the 95% confidence interval of the prediction.

BRT model show that it is not just the presence within the surrounding area but the rate of occurrence of taxa that affects colonisation chances, with density of occurrence in the regional pool being the most important dri-

ver of site presence. Taxon identity and restoration site location were overwhelmingly influential in the BRT models when they were included, but cross-validation highlighted they were not strong predictors of general patterns. No reduction in prediction success was evident for cross-validated models with and without their presence, suggesting their main influence was to overfit their respective models to the training data, and thus reducing their generality outside of the training data set.

The river restoration projects studied here achieved their physical habitat restoration objectives of restoring natural or semi-natural habitat conditions (Jähnig *et al.*, 2011). Of the variables associated with the particular restoration projects, cost and the number of years following restoration were the most important predictors in the model, but their influence was minor and far outweighed by regional taxon pool factors. Other factors associated with the restoration projects, such as the methods used, had little influence on the model. Thus, failure to colonise restored reaches was less a matter of lack of habitat improvement than a function of dispersal

constraints operating between regional taxon pools and restored reaches. Restoration often achieves the goals of restoring more natural hydromorphological conditions, but corresponding improvements in the benthic invertebrates and fish are often not apparent (Palmer *et al.*, 2010; Louhi *et al.*, 2011; Sundermann *et al.*, 2011a). The colonisation of organisms at restored sites is determined firstly by their presence in the regional species pool (Cornell & Lawton, 1992; Lake *et al.*, 2007) and secondly by dispersal constraints such as distance and in-stream barriers (Bond & Lake, 2003; Lake *et al.*, 2007). Only once these filters are overcome do local-scale factors such as habitat structure and biotic interactions matter. Within highly disturbed landscapes with diminished regional species pools, local improvements are likely to be limited by catchment-scale press disturbances such as land-use change.

As per our primary hypothesis regarding the drivers of colonisation rates, distance to colonist source was a key factor driving taxon presence at restored sites, with a clear decline in taxon probability of occurrence at a restored site with increasing distance to source. However, this response was highly taxon specific, as demonstrated by the importance of taxon identity in the BRT training model, although including taxon identity as a factor did not improve model prediction. Given the lack of importance of dispersal mode in model 4, this taxon-specific response was not a clear result of whether a species has an aerial dispersal mode. Identifying thresholds is a key aspect of managing lotic systems (Clements, Vieira & Sonderegger, 2010; Dodds *et al.*, 2010), but despite evidence of thresholds for individual taxa, the mean response of all taxa using GLM was a gradual decline over the 10-km distance. Nonetheless, by holding other variables constant, the BRT model suggested the most important range was that in the first km from the restored site, and the trend of individual taxa was highly variable with thresholds more apparent. To ensure adequate regional representation (i.e. removing rare species), we considered only taxa present in more than half of the restoration sites' surrounding taxon pools. By excluding rare taxa, we may in fact be restricting these patterns and the effect of distance may be more severe than we have shown. Recent work has demonstrated the importance of rare species in maintaining ecological functions, especially vulnerable functions (Mouillot *et al.*, 2013). Thus, assessing the importance of rare species as vectors of more vulnerable, non-redundant traits at restored rivers would be a fruitful area to explore, especially in relation to various thresholds such as dispersal distance.

Many colonist taxa were not present at sites nearby restored sites, and thus, dispersal was often over large distances, and in fact, many regions did not have certain taxa. Previous research on both invertebrates and fish suggests most species colonise from within 5 km of the restoration sites (Sundermann *et al.*, 2011b; Stoll *et al.*, 2013), and our results suggest that the first km of river network may in fact be the most important. The average time post-restoration for these sites was 4.1 years, ranging up to 12 years, and although this only had a minor role to play in the models, there was an interaction with dispersal distance, potentially indicating distance loses importance with time. With the requirement of long-distance dispersal, restoration may be a slow and stochastic process (Bond & Lake, 2003) as dispersal is typically more difficult between isolated populations in lotic than other systems due to their unidirectional and dendritic nature (Fagan, 2002). In fact, recolonisation of restored sites may indeed be continuing 30–50 years after habitat restoration (Langford *et al.*, 2009). Thus, we may be measuring just short-distance colonisers in the present study as the first stage in the colonisation process. Nonetheless, this time is dependent on the surrounding species pool and proximity to a viable source, and colonisation can be rapid in restored streams with natural colonist sources nearby and on small spatial scales (Gore & Milner, 1990). Therefore, placement of restoration sites near pre-existing populations can help to ensure successful colonisation (Tilman, Lehman & Kareiva, 1997; Huxel & Hastings, 1999).

A clear interaction was present between the distance to a viable source and the density of these sources in the surrounding area in predicting colonisation of restored sites. Thus, with appropriate restoration planning, maintenance of viable metapopulations could be ensured if dispersal between restored patches can be enhanced (Lake *et al.*, 2007). Distance between patches is likely to be a critical factor enabling this dispersal as indicated by the low likelihood of restoration site presence with nearest source pool at 10-km distance. It may be more beneficial to create many smaller restored patches within a degraded region to allow stepping stones across the landscape, which are a critical component of long-distance dispersal (Saura, Bodin & Fortin, 2014). In this instance, and with the support of prior research (Sundermann *et al.*, 2011b), restored reaches < 5 km apart could provide adequate stepping stones for many of these taxa. Nonetheless, while maintenance of restored patches could ensure population viability under conditions with underlying large-scale disturbances, the reverse could also be true, with greater res-

toration success where regional species pools are more intact. It might even be that restoration has a higher chance of succeeding within regions or catchments with lower rates of overall disturbance such as land-use change (Bernhardt & Palmer, 2011).

While increasing barrier score reduced the overall probability of colonisation in the GLM regression models, in accordance with our expectation, the impact of barriers was minor and clearly taxon specific. As a substantial proportion of the 77 taxa in the present study were merolimnic and thus had an adult aerial dispersal mode, we expected in-stream barriers would be a minor influence on colonisation success at restored reaches. Depending on their life-history traits, barriers such as culverts and weirs can impede upstream passage for stream invertebrates in many ways including direct physical impedance of hololimnic non-insect taxa, especially those with a migratory phase (Resh, 2005; David et al., 2014). Adult insect flight dispersal can both follow stream corridors or across-land between stream sections (Macneale, Peckarsky & Likens, 2005), and, although mostly along stream corridors (Petersen et al., 2004), the importance of lateral movement across land is becoming more apparent (Didham et al., 2012). For flight dispersers, many other potential reasons exist for in-stream barriers to impede passage, such as culverts narrowing the flight path or increasing predation rate by spiders (Blakely et al., 2006). It is likely that barriers in our study had taxon-specific reductive effects depending on barrier type, and dispersal mode. Opposing our primary hypothesis, but similar to a recent study on metacommunity structuring (Grönroos et al., 2013), we found no clear difference between overland and river network distance approaches to predict colonisation; however, these distances were highly correlated with each other.

Given the limited funds associated with restoration projects, it is critical to track and maximise their performance, rather than simply expecting that hydromorphological change will result in biotic improvement. While previous research has demonstrated the importance of spatial context for the biotic colonisation of restored river reaches (Sundermann et al., 2011b; Stoll et al., 2013), our results clearly quantify the importance of these large-scale factors such as dispersal and regional pools in conjunction with local-scale habitat restoration factors. Spatial landscape planning and selecting appropriate restoration sites could drastically improve the outcome of many restoration projects (Huxel & Hastings, 1999) and potentially assist with regional metapopulation maintenance. Without spatial planning of restoration projects within degraded landscapes, population recovery can lag behind habitat restoration (Tilman et al., 1997) and may lead to the requirement of labour-intensive approaches such as assisted migration to overcome fragmentation and dispersal limitation, despite uncertainty (McLachlan, Hellmann & Schwartz, 2007; Ricciardi & Simberloff, 2009; Vitt, Havens & Hoegh-Guldberg, 2009).

Our results specifically emphasise two key factors determining colonisation of restored sites: the taxon pool occurrence rate and proximity to this pool. As hypothesised, colonisation of restored sites declined with increasing distance to nearest source, particularly within the first km, but this influence was outweighed by the density of taxa within the surrounding pool. Presence and density of in-stream barriers had a minor but detectable effect on reducing overall colonisation rate. Knowledge of the regional species pool will thus assist with spatial prioritisation of restoration projects to enhance the outcome for communities inhabiting these river reaches. Where there is good regional richness, it is likely that there will be a measurable positive biotic response to habitat improvements at a restored site, especially on the short term. Given the clear differences in dispersal modes of many benthic invertebrate groups, distances, barriers and the dendritic nature of stream networks are likely to have highly variable effects on trait groups. Thus, this research opens the door for more in-depth analyses into trait-based responses to restoration within patchy environments, particularly between active and passive dispersers with and without flight stages.

Acknowledgments

References

Bednarek A.T. (2001) Undamming rivers: a review of the ecological impacts of dam removal. *Environmental Management*, **27**, 803–814.

Bernhardt E., Palmer M., Allan J., Alexander G., Barnas K., Brooks S. et al. (2005) Synthesizing U.S. river restoration efforts. *Science*, **308**, 636–637.

Bernhardt E.S. & Palmer M.A. (2011) River restoration: the fuzzy logic of repairing reaches to reverse catchment scale degradation. *Ecological applications*, **21**, 1926–1931.

Blakely T.J., Harding J.S., Mcintosh A.R. & Winterbourn M.J. (2006) Barriers to the recovery of aquatic insect communities in urban streams. *Freshwater Biology*, **51**, 1634–1645.

Bohn B.A. & Kershner J.L. (2002) Establishing aquatic restoration priorities using a watershed approach. *Journal of Environmental Management*, **64**, 355–363.

Bond N.R. & Lake P.S. (2003) Local habitat restoration in streams: constraints on the effectiveness of restoration for stream biota. *Ecological Management and Restoration*, **4**, 193–198.

Briers R.A., Gee J.H.R., Cariss H.M. & Geoghegan R. (2004) Inter-population dispersal by adult stoneflies detected by stable isotope enrichment. *Freshwater Biology*, **49**, 425–431.

Buston P.M. & Elith J. (2011) Determinants of reproductive success in dominant pairs of clownfish: a boosted regression tree analysis. *Journal of Animal Ecology*, **80**, 528–538.

Campbell Grant E.H., Lowe W.H. & Fagan W.F. (2007) Living in the branches: population dynamics and ecological processes in dendritic networks. *Ecology Letters*, **10**, 165–175.

Clements W.H., Vieira N.K.M. & Sonderegger D.L. (2010) Use of ecological thresholds to assess recovery in lotic ecosystems. *Journal of the North American Benthological Society*, **29**, 1017–1023.

Cornell H. & Lawton J. (1992) Species interactions, local and regional processes, and limits to the richness of ecological communities: a theoretical perspective. *Journal of Animal Ecology*, **61**, 1–12.

Cushing C. & Gaines W. (1989) Thoughts on recolonization of endorheic cold desert spring-streams. *Journal of the North American Benthological Society*, **8**, 277–287.

David B.O., Tonkin J.D., Taipeti K.W.T. & Hokianga H.T. (2014) Learning the ropes: mussel spat ropes improve fish and shrimp passage through culverts. *Journal of Applied Ecology*, **51**, 214–223.

DeGraff J.V. & Evans J.E. eds. (2013) *The Challenges of Dam Removal and River Restoration*. Geological Society of America, Boulder, CO.

Didham R., Blakely T., Ewers R.M., Hitchings T., Ward J. & Winterbourn M.J. (2012) Horizontal and vertical structuring in the dispersal of adult aquatic insects in a fragmented landscape. *Fundamental and Applied Limnology*, **180**, 27–40.

Dodds W.K., Clements W.H., Gido K., Hilderbrand R.H. & King R.S. (2010) Thresholds, breakpoints, and nonlinearity in freshwaters as related to management. *Journal of the North American Benthological Society*, **29**, 988–997.

Doyle M.W., Stanley E.H., Orr C.H., Selle A.R., Sethi S.A. & Harbor J.M. (2005) Stream ecosystem response to small dam removal: lessons from the Heartland. *Geomorphology*, **71**, 227–244.

Dudgeon D., Arthington A.H., Gessner M.O., Kawabata Z.-I., Knowler D.J., Lévêque C. et al. (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological reviews of the Cambridge Philosophical Society*, **81**, 163–182.

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

Fagan W. (2002) Connectivity, fragmentation, and extinction risk in dendritic metapopulations. *Ecology*, **83**, 3243–3249.

Fullerton A.H., Burnett K.M., Steel E.A., Flitcroft R.L., Pess G.R., Feist B.E. et al. (2010) Hydrological connectivity for riverine fish: measurement challenges and research opportunities. *Freshwater Biology*, **55**, 2215–2237.

Gore J.A. & Milner A.M. (1990) Island biogeographical theory: can it be used to predict lotic recovery rates? *Environmental Management*, **14**, 737–753.

Grönroos M., Heino J., Siqueira T., Landeiro V.L., Kotanen J. & Bini L.M. (2013) Metacommunity structuring in stream networks: roles of dispersal mode, distance type, and regional environmental context. *Ecology and Evolution*, **3**, 4473–4487.

Haase P., Hering D., Jähnig S.C., Lorenz A.W. & Sundermann A. (2013) The impact of hydromorphological restoration on river ecological status: a comparison of fish, benthic invertebrates, and macrophytes. *Hydrobiologia*, **704**, 475–488.

Haase P., Lohse S., Pauls S., Schindeh K., Sundermann A., Rolauffs P. et al. (2004a) Assessing streams in Germany with benthic invertebrates: development of a practical standardised protocol for macroinvertebrate sampling and sorting. *Limnologica*, **34**, 349–365.

Haase P., Pauls S., Sundermann A. & Zenker A. (2004b) Testing different sorting techniques in macroinvertebrate samples from running waters. *Limnologica*, **34**, 366–378.

Hanski I. & Gilpin M.E. (1997) *Metapopulation Biology: Ecology, Genetics, and Evolution*. Academic press, San Diego.

Hilderbrand R.H., Watts A.C. & Randle A.M. (2005) The myths of restoration ecology. *Ecology and Society* **10**, art19.

Holl K.D., Crone E.E. & Schultz C.B. (2003) Landscape restoration: moving from generalities to methodologies. *BioScience*, **53**, 491.

Huxel G.R. & Hastings A. (1999) Habitat loss, fragmentation, and restoration. *Restoration Ecology*, **7**, 309–315.

Jähnig S.C., Lorenz A.W., Hering D., Antons C., Sundermann A., Jedicke E. et al. (2011) River restoration success: a question of perception. *Ecological Applications*, **21**, 2007–2015.

Lake P.S., Bond N. & Reich P. (2007) Linking ecological theory with stream restoration. *Freshwater Biology*, **52**, 597–615.

Langford T.E.L., Shaw P.J., Ferguson A.J.D. & Howard S.R. (2009) Long-term recovery of macroinvertebrate biota in grossly polluted streams: re-colonisation as a constraint to ecological quality. *Ecological Indicators* **9**, 1064–1077.

Liermann C.R., Nilsson C., Robertson J. & Ng R.Y. (2012) Implications of dam obstruction for global freshwater fish diversity. *BioScience*, **62**, 539–548.

Louhi P., Mykrä H., Paavola R., Huusko A., Vehanen T., Mäki-Petäys A. *et al.* (2011) Twenty years of stream restoration in Finland: little response by benthic macroinvertebrate communities. *Ecological Applications*, **21**, 1950–1961.

Macneale K.H., Peckarsky B.L. & Likens G.E. (2005) Stable isotopes identify dispersal patterns of stonefly populations living along stream corridors. *Freshwater Biology*, **50**, 1117–1130.

McLachlan J.S., Hellmann J.J. & Schwartz M.W. (2007) A framework for debate of assisted migration in an era of climate change. *Conservation Biology*, **21**, 297–302.

Milner A.M., Knudsen E.E., Soiseth C., Robertson A.L., Schell D., Phillips I.T. *et al.* (2000) Colonization and development of stream communities across a 200-year gradient in Glacier Bay National Park, Alaska, U.S.A. *Canadian Journal of Fisheries and Aquatic Sciences*, **57**, 2319–2335.

Mouillot D., Bellwood D.R., Baraloto C., Chave J., Galzin R., Harmelin-Vivien M. *et al.* (2013) Rare species support vulnerable functions in high-diversity ecosystems. *Plos Biology*, **11**, e1001569.

Nilsson C., Reidy C.A., Dynesius M. & Revenga C. (2005) Fragmentation and flow regulation of the world's large river systems. *Science*, **308**, 405–408.

Palmer M.A. (2009) Reforming watershed restoration: science in need of application and applications in need of science. *Estuaries and Coasts*, **32**, 1–17.

Palmer M.A., Bernhardt E.S., Allan J.D., Lake P.S., Alexander G., Brooks S. *et al.* (2005) Standards for ecologically successful river restoration. *Journal of Applied Ecology*, **42**, 208–217.

Palmer M.A., Menninger H.L. & Bernhardt E. (2010) River restoration, habitat heterogeneity and biodiversity: a failure of theory or practice? *Freshwater Biology*, **55**, 205–222.

Parkyn S.M. & Smith B.J. (2011) Dispersal constraints for stream invertebrates: setting realistic timescales for biodiversity restoration. *Environmental management*, **48**, 602–614.

Petersen I., Masters Z., Hildrew A.G. & Ormerod S.J. (2004) Dispersal of adult aquatic insects in catchments of differing land use. *Journal of Applied Ecology*, **41**, 934–950.

Pretty J.L., Harrison S.S.C., Shepherd D.J., Smith C., Hildrew A.G. & Hey R.D. (2003) River rehabilitation and fish populations: assessing the benefit of instream structures. *Journal of Applied Ecology*, **40**, 251–265.

R Core Team (2013) *R: A language and environment for statistical computing*. R Foundation of Statistical Computing, Vienna, Austria.

Resh V.H. (2005) Stream crossings and the conservation of diadromous invertebrates in South Pacific island streams. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **15**, 313–317.

Ricciardi A. & Simberloff D. (2009) Assisted colonization is not a viable conservation strategy. *Trends in Ecology & Evolution*, **24**, 248–253.

Riley S.C. & Fausch K.D. (1995) Trout population response to habitat enhancement in six northern Colorado streams. *Canadian Journal of Fisheries and Aquatic Sciences*, **52**, 34–53.

Saura S., Bodin Ö. & Fortin M.-J. (2014) Stepping stones are crucial for species' long-distance dispersal and range expansion through habitat networks. *Journal of Applied Ecology*, **51**, 171–182.

Stoll S., Kail J., Lorenz A.W., Sundermann A. & Haase P. (2014) The importance of the regional species pool, ecological species traits and local habitat conditions for the colonization of restored river reaches by fish. *PLoS ONE*, **9**, e84741.

Stoll S., Sundermann A., Lorenz A.W., Kail J. & Haase P. (2013) Small and impoverished regional species pools constrain colonisation of restored river reaches by fishes. *Freshwater Biology*, **58**, 664–674.

Sundermann A., Antons C., Cron N., Lorenz A.W., Hering D. & Haase P. (2011a) Hydromorphological restoration of running waters: effects on benthic invertebrate assemblages. *Freshwater Biology*, **56**, 1689–1702.

Sundermann A., Stoll S. & Haase P. (2011b) River restoration success depends on the species pool of the immediate surroundings. *Ecological Applications*, **21**, 1962–1971.

Tilman D., Lehman C.L. & Kareiva P. (1997) Population dynamics in spatial habitats. In: *Spatial Ecology* (Eds D. Tilman & P. Kareiva), pp. 3–20. Princeton University Press, Princeton, NJ.

Vitt P., Havens K. & Hoegh-Guldberg O. (2009) Assisted migration: part of an integrated conservation strategy. *Trends in Ecology & Evolution*, **24**, 473–474.

Winterbourn M.J., Chadderton W.L., Entrekin S.A., Tank J.L. & Harding J.S. (2007) Distribution and dispersal of adult stream insects in a heterogeneous montane environment. *Fundamental and Applied Limnology*, **168**, 127–135.

Supporting Information

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

Table S1. Summary table of the nine individual logistic regression models predicting taxa presence at a restored site as a function of distance and total barrier score.

(Manuscript accepted 8 April 2014)