Ground-truthing environmental DNA metabarcoding for ecological hypothesis testing at the pondscape

Lynsey R. Harper[1*], Lori Lawson Handley[1], Christoph Hahn[1,2], Neil Boonham[3,4], Helen C. Rees[5], Erin Lewis[3], Ian P. Adams[3], Peter Brotherton[6], Susanna Phillips[6] and Bernd Hänfling[1]

[1] School of Environmental Sciences, University of Hull, Hull, HU6 7RX, UK

[2] Institute of Zoology, University of Graz, Graz, Styria, Austria

[3] Fera Science Ltd (Fera), Sand Hutton, York, YO14 1LZ, UK

[4] Newcastle University, Newcastle upon Tyne, NE1 7RU, UK

[5] ADAS, School of Veterinary Medicine and Science, The University of Nottingham, Sutton Bonington Campus, Leicestershire, LE12 5RD, UK

[6] Natural England, Peterborough, PE1 1NG, UK

Running title: Hypothesis testing with eDNA metabarcoding

***Corresponding author: Lynsey R. Harper**

School of Environmental Sciences, University of Hull, Hull, HU6 7RX, UK

E-mail: L.Harper@2015.hull.ac.uk

Abstract

Environmental DNA (eDNA) metabarcoding is revolutionising biodiversity monitoring, but has unrealised potential for ecological hypothesis testing. Here, we ground-truth eDNA metabarcoding for describing vertebrate communities from 532 UK ponds, examine associations between the great crested newt (*Triturus cristatus*, GCN) and other vertebrates, and investigate factors determining GCN occurrence and vertebrate species richness at the pondscape. GCN occurrence was positively correlated with amphibian and waterfowl species richness, where GCN had strong positive associations with smooth newt (*Lissotriton vulgaris*), common coot (*Fulica atra*), and common moorhen (*Gallinula chloropus*), but a negative association with common toad (*Bufo bufo*). GCN occurrence was negatively correlated to fish species richness, specifically three-spined stickleback (*Gasterosteus aculeatus*) and ninespine stickleback (*Pungitius pungitius*) presence. We reaffirm reported abiotic associations (e.g. GCN preference for smaller, deeper ponds) but also provide novel insights, including a negative effect of pond outflow on GCN. Furthermore, we reveal novel determinants of vertebrate species richness at the pondscape, including pond density, macrophyte cover, and terrestrial habitat. Both GCN occupancy and species richness correlated with the GCN Habitat Suitability Index score, supporting its application to GCN survey. Our findings demonstrate the prospects of eDNA metabarcoding for hypothesis testing at unprecedented scales and enhancement of freshwater monitoring and research.

Keywords: biodiversity monitoring, environmental DNA (eDNA), hypothesis testing, metabarcoding, pondscape, *Triturus cristatus*

1. INTRODUCTION

Environmental DNA (eDNA) analysis offers ecologists exceptional power to detect organisms within and across ecosystems. DNA secreted or excreted by organisms into their environment can be sampled and analysed using different approaches to reveal the distribution of single or multiple species (Lawson Handley, 2015; Rees, Maddison, Middleditch, Patmore, & Gough, 2014). When combined with high-throughput sequencing (i.e. eDNA metabarcoding), eDNA can yield efficient, comprehensive assessments of entire communities (Deiner et al., 2017), providing a step change in biodiversity monitoring (Hering et al., 2018). eDNA metabarcoding has untapped potential to test ecological hypotheses by enabling biodiversity monitoring at phenomenal scales with minimal impact to communities under investigation. Although this tool has been used to estimate species richness and assess diversity along environmental gradients (e.g. Evans et al., 2017; Hänfling et al., 2016; Kelly et al., 2016; Olds et al., 2016), its applications in community ecology are relatively unexplored.

Aquatic ecosystems are highly suited to eDNA studies (Muha, Rodríguez-Rey, Rolla, & Tricarico, 2017) as eDNA exists in multiple states with rapid modes of transport and degradation, increasing detectability of contemporary biodiversity (Barnes & Turner, 2015; Rees et al., 2014). Lentic systems provide further opportunities for eDNA research, being discrete water bodies with variable physicochemical properties that do not experience flow dynamics (Muha et al., 2017). Ponds in particular have enormous biodiversity and experimental virtue that has not been maximised in previous eDNA metabarcoding assessments of this habitat (Bálint et al., 2017; Evans et al., 2017; Klymus, Richter, Thompson, & Hinck, 2017; Ushio et

al., 2017; Valentini et al., 2016). These small and abundant water bodies span broad ecological gradients (De Meester et al., 2005) and comprise pondscapes - a network of ponds and their surrounding terrestrial habitat (Hill et al., 2018). Pondscapes contribute substantially to aquatic and non-aquatic biodiversity across spatial scales, with ponds supporting many rare and protected species in fragmented landscapes (Biggs, von Fumetti, & Kelly-Quinn, 2016; De Meester et al., 2005; Hill et al., 2018). Consequently, ponds are model systems for experimental validation and examination of biogeographical patterns (De Meester et al., 2005). Habitat complexity and tools required for different taxa with associated bias (Evans et al., 2017) and cost (Valentini et al., 2016) once hindered exhaustive sampling of pond biodiversity, but eDNA metabarcoding may overcome these barriers (Biggs et al., 2016; Hill et al., 2018).

In the UK, the threatened great crested newt (*Triturus cristatus*, hereafter GCN) is a flagship species for pond conservation (Hill et al., 2018). The extensive literature on the GCN provides an excellent opportunity to ground truth ecological patterns revealed by eDNA metabarcoding. Both biotic (e.g. breeding substrate, prey, and predators) and abiotic (e.g. pond area, depth, and temperature) factors are known to influence GCN breeding success (Langton, Beckett, & Foster, 2001). The GCN Habitat Suitability Index (HSI) (ARG-UK, 2010; Oldham, Keeble, Swan, & Jeffcote, 2000) accounts for these factors using 10 suitability indices that are scored and combined to calculate a decimal score between 0 and 1 (where 1 = excellent habitat). Larvae are susceptible to fish and waterfowl predation (Hartel, Nemes, & Oellerer, 2010; R. Rannap & Briggs, 2006; Skei, Dolmen, Rønning, & Ringsby, 2006), and adults reportedly avoid ponds containing three-spined stickleback (*Gasterosteus aculeatus*) (McLee & Scaife, 1992), ninespine stickleback (*Pungitius pungitius*), crucian carp (*Carassius carassius*), and common carp (*Carassius carpio*) (Rannap,

Lõhmus, & Briggs, 2009a, b). Conversely, the GCN and smooth newt (*Lissotriton vulgaris*) prefer similar habitat and often co-occur (Denoël, Perez, Cornet, & Ficetola, 2013; Rannap & Briggs, 2006; Rannap et al., 2009a; Skei et al., 2006). GCNs thrive in ponds with good water quality as indicated by diverse macroinvertebrate communities (Oldham et al., 2000; Rannap et al., 2009a), and water clarity is important for breeding displays, foraging success, and egg survival (Rannap & Briggs, 2006; Skei et al., 2006). Pond networks encourage GCN occupancy (Denoël et al., 2013; Hartel et al., 2010; Joly, Miaud, Lehmann, & Grolet, 2001; Rannap et al., 2009a), but larger area discourages presence (Joly et al., 2001). Ponds with heavy shading (Vuorio, Heikkinen, & Tikkanen, 2013) or dense macrophyte cover (Hartel et al., 2010; Rannap & Briggs, 2006; Skei et al., 2006) are unlikely to support viable populations. GCNs also depend on terrestrial habitat, preferring open, semi-rural pondscapes (Denoël et al., 2013) containing pasture, extensively grazed and rough grassland, scrub, and coniferous and deciduous woodland (Gustafson, Malmgren, & Mikusiński, 2011; Oldham et al., 2000; Rannap & Briggs, 2006; Rannap et al., 2009a; Vuorio et al., 2013).

We assessed vertebrate communities at the pondscape using a dataset generated by eDNA metabarcoding for over 500 ponds with comprehensive environmental metadata. We evaluated eDNA metabarcoding as a tool for ecological hypothesis testing, and compared its outputs to previous results generated by established methods. Specifically, we aimed to identify biotic (community presence-absence data) and abiotic determinants (environmental metadata on ponds and surrounding terrestrial habitat) of the GCN at an unparalleled scale, and determinants of vertebrate species richness at the pondscape - an impractical task by conventional means. Finally, we investigated applicability of the HSI to predict eDNA-based GCN

occupancy and vertebrate species richness of ponds.

2. MATERIALS AND METHODS

2.1 Samples

Samples from 508 ponds included in Natural England's Great Crested Newt Evidence Enhancement Programme were processed using eDNA metabarcoding alongside 24 privately surveyed ponds. Water samples were collected using established methodology (Biggs et al., 2015), detailed in Supporting Information. In brief, 20 x 30 mL water samples were collected from each pond and pooled. Six 15 mL subsamples were taken from the pooled sample and each added to 33.5 mL absolute ethanol and 1.5 mL sodium acetate 3 M (pH 5.2). Subsamples were pooled during DNA extraction to produce one eDNA sample per pond. Targeted qPCR detected GCN in 265 (49.81%) ponds (Harper et al., 2018).

Environmental metadata (Table S1) were collected for 504 of 532 ponds (Figure S1), including: maximum depth; circumference; width; length; area; density; terrestrial overhang; shading; macrophyte cover; HSI score (Oldham et al., 2000); HSI band (categorical classification of HSI score by ARG-UK, 2010); permanence; water quality; pond substrate; presence of inflow or outflow; presence of pollution; presence of other amphibians, fish and waterfowl; woodland; rough grass; scrub/hedge; ruderals; other good terrestrial habitat (i.e. good terrestrial habitat that did not conform to aforementioned habitat types); and overall terrestrial habitat

quality.

2.2 DNA reference database construction

A custom, phylogenetically curated reference database of mitochondrial 12S ribosomal RNA (rRNA) sequences for UK fish species was previously constructed for eDNA metabarcoding of lake fish communities (Hänfling et al., 2016). In this study, additional reference databases for UK amphibians, reptiles, birds, and mammals were constructed (see Harper et al., 2018 and Supporting Information). Reference sequences available for species varied across vertebrate groups: amphibians 100.00% (N = 21), reptiles 90.00% (N = 20), mammals 83.93% (N = 112), and birds 55.88% (N = 621). Table S2 lists species without database representation, i.e. no records for any species in a genus. Sanger sequences were obtained from tissue of GCN, smooth newt, Alpine newt (*Mesotriton alpestris*), common toad (*Bufo bufo*), and common frog (*Rana temporaria*) to supplement the amphibian database (see Supporting Information). The complete reference databases compiled in GenBank format have been deposited in a dedicated GitHub repository for this study, permanently archived at: https://doi.org/10.5281/zenodo.1304107.

2.3 Primer validation

Reference databases were combined for *in silico* validation of published 12S rRNA primers 12S-V5-F (5'-ACTGGGATTAGATACCCC-3') and 12S-V5-R (5'-

TAGAACAGGCTCCTCTAG-3') (Riaz et al., 2011) using ecoPCR software (Ficetola *et al.* 2010). Set parameters allowed a 50-250 bp fragment and three mismatches between each primer and reference sequence. Primers were validated *in vitro* for UK fish by Hänfling et al. (2016) and here for six UK amphibian species (Figure S2).

2.4 eDNA metabarcoding

Our two-step PCR workflow is fully described in Harper et al. (2018) and Supporting Information. eDNA was first amplified with the aforementioned primers, where PCR positive controls (six per PCR plate; N = 114) were cichlid (*Rhamphochromis esox*) DNA and PCR negative controls (six per PCR plate; N = 114) were sterile molecular grade water (Fisher Scientific UK Ltd, UK). PCR products were individually purified using E.Z.N.A.® Cycle Pure V-Spin Clean-Up Kits (Omega Bio-tek, GA, USA) following the manufacturer's protocol. The second PCR bound Multiplex Identification (MID) tags to the purified products. PCR products were individually purified using magnetic bead clean-up and quantified with a Quant-IT™ PicoGreen™ dsDNA Assay (Invitrogen, ThermoFisher Scientific, UK). Samples were normalised, pooled, and libraries quantified using a Qubit™ dsDNA HS Assay Kit (Invitrogen, ThermoFisher Scientific, UK). Libraries were sequenced on an Illumina MiSeq using 2 x 300 bp V3 chemistry (Illumina Inc., San Diego, CA, USA) and raw sequence reads processed using metaBEAT (metaBarcoding and Environmental Analysis Tool) v0.8 (https://github.com/HullUni-bioinformatics/metaBEAT). After quality trimming, merging, chimera detection, and clustering, non-redundant query sequences were

compared against our reference database using BLAST (Zhang, Schwartz, Wagner, & Miller, 2000). Putative taxonomic identity was assigned using a lowest common ancestor (LCA) approach based on the top 10% BLAST matches for any query matching with at least 98% identity to a reference sequence across more than 80% of its length. Unassigned sequences were subjected to a separate BLAST against the complete NCBI nucleotide (nt) database at 98% identity to determine the source via LCA as described above. The bioinformatic analysis has been deposited in the GitHub repository for reproducibility.

2.5 Data analysis

Analyses were performed in the statistical programming environment R v.3.4.2 (R Core Team, 2017). Data and R scripts have been deposited in the GitHub repository. Assignments from different databases were merged, and spurious assignments (i.e. non-UK species, invertebrates and bacteria) removed from the dataset. The genus *Strix* was reassigned to tawny owl (*Strix aluco*) as it is the only UK representative of this genus. Where family and genera assignments containing a single UK representative had reads assigned to species, reads from all assignment levels were merged and manually assigned to that species. Consequently, all taxonomic assignments in the final dataset were of species resolution. The only misassignment was Scottish wildcat (*Felis silvestris*), which does not occur at the sampling localities (Kent, Lincolnshire and Cheshire) and was therefore reassigned to domestic cat (*Felis catus*).

To minimise risk of false positives, species were only classed as present at

sites if their sequence frequency exceeded species-specific thresholds. Thresholds were defined using the maximum sequence frequency of each species in PCR positive controls (*N* = 114; Table S3). For example, the GCN threshold was 0.028% to omit false positives in PCR positive controls. After applying thresholds, the read count data were converted to a species presence-absence matrix. Analyses were based on species-specific thresholds, but also performed for different blanket sequence thresholds (0.05 - 30%, Tables S4-9). We tested biotic and abiotic determinants of GCN occupancy and vertebrate species richness, and appropriateness of the HSI. Hypotheses are summarised in Table 1.

All Generalised Linear Mixed Models (GLMMs) were executed using the R package 'lme4' v1.1-12 (Bates, Mächler, Bolker, & Walker, 2015). GCN occurrence relating to number of other vertebrate species was investigated with a binomial GLMM, and species associations identified using the R package 'cooccur' v1.3 (Griffith, Veech, & Marsh, 2016) (*N* = 532). Identified associations informed candidate biotic variables for inclusion with abiotic variables (Table S1) in a binomial GLMM of GCN occurrence (*N* = 504). Candidate explanatory variables were assessed for collinearity, relative importance, and non-linearity (see Supporting Information). HSI score and band were collinear, thus HSI score was analysed in a separate binomial GLMM. Using the R package 'ncf' v1.1-7 (Bjørnstad, 2009), spline correlograms of the Pearson residuals from a binomial Generalised Linear Model (GLM) and GLMM were compared to assess spatial autocorrelation potential in our dataset. Sample was treated as a random effect in the GLMM to account for dependencies within ponds (Zuur, Ieno, Walker, Saveliev, & Smith, 2009). The mixed model successfully accounted for spatial autocorrelation within sites. The same steps were performed to identify explanatory variables and a modeling framework for

vertebrate species richness. A Poisson distribution was specified for all species richness models as the response variable was integer count data.

Binomial and Poisson models considered respectively were nested thus best models were chosen using stepwise backward deletion of terms based on Likelihood Ratio Tests (LRTs). Akaike's Information Criteria (AIC) was employed to evaluate model fit (Akaike, 1973). Final models were tested for overdispersion using the R package 'RVAideMemoire' v 0.9-69 (Hervé, 2018) and a custom function testing overdispersion of the Pearson residuals. Model fit was assessed using the Hosmer and Lemeshow Goodness of Fit Test (Hosmer & Lemeshow, 2000) within the R package 'ResourceSelection' v0.3-2 (Lele, Keim, & Solymos, 2016), quantile-quantile plots and partial residual plots (Zuur et al., 2009). Model predictions were obtained using the predictSE() function in the 'AICcmodavg' package v2.1-1 (Mazerolle, 2016) and upper and lower 95% CIs were calculated from the standard error of the predictions. Results were plotted using the R package 'ggplot2' v 2.1.0 (Wickham, 2009).

3. RESULTS

3.1 eDNA metabarcoding

A total of 532 eDNA samples and 228 PCR controls were processed across two sequencing runs. The runs generated raw sequence read counts of 36,236,862 and 32,900,914 respectively. After trimming and merging of paired-end reads, 26,294,906 and 26,451,564 sequences remained. Following removal of chimeras and redundancy

via clustering, the libraries contained 14,141,237 and 14,081,939 sequences (average read counts of 36,826 and 36,671 per sample respectively), of which 13,126,148 and 13,113,143 sequences were taxonomically assigned. The final dataset (thresholds applied and assignments corrected) contained 60 vertebrate species (Table S10), including six amphibians, 14 fish, 18 birds, and 22 mammals (Figure S3).

3.2 Pondscape biodiversity

All native amphibians were found alongside the non-native marsh frog (*Pelophylax ridibundus*). The GCN (*N* = 149), smooth newt (*N* = 152) and common frog (*N* = 120) were widespread, but the common toad (*N* = 42), palmate newt (*Lissotriton helveticus*, *N* = 5) and marsh frog were uncommon (*N* = 1). The threatened European eel (*Anguilla anguilla*, *N* = 15) and crucian carp (*Carassius carassius*, *N* = 2) were detected alongside native fishes, such as pike (*Esox lucius*, *N* = 17) and roach (*Rutilus rutilus*, *N* = 72), and introduced species, including common carp (*N* = 41) and ruffe (*Gymnocephalus cernua*, *N* = 1). Regular and irregular waterfowl were identified, ranging from common moorhen (*Gallinula chloropus*, *N* = 215) to grey heron (*Ardea cinerea*, *N* = 1). Terrestrial fauna were often detected in fewer than five ponds (Figures S3c, d). Eurasian jay (*Garrulus glandarius*, *N* = 7), buzzard (*Buteo buteo*, *N* = 4), dunnock (*Prunella modularis*, *N* = 4), and starling (*Sturnus vulgaris*, *N* = 4) were the most frequently detected terrestrial birds (Figure S3c). Domesticated, including cow (*Bos taurus*, *N* = 179) and pig (*Sus scrofa*, *N* = 140), or introduced mammals (Mathews *et al.* 2018), such as grey squirrel (*Sciurus carolinensis*, *N* = 57) and Reeve's muntjac (*Muntiacus reevesi*, *N* = 3), outweighed native mammals (Figure

S3d). Nonetheless, we detected several mammals of conservation concern with Biodiversity Actions Plans (Mathews et al., 2018), including otter (*Lutra lutra*, N = 1), water vole (*Arvicola amphibius*, N = 16), European polecat (*Mustela putorius*, N = 1), brown hare (*Lepus europaeus*, N = 1) and water shrew (*Neomys fodiens*, N = 9). Notably, the invasive American mink (*Neovison vison*) was absent despite widespread UK distribution (Mathews et al., 2018). All species and their detection frequencies are listed in Table S10.

3.3 Biotic determinants of GCN occurrence

GCN occupancy was negatively correlated with fish species richness, but positively influenced by amphibian and waterfowl species richness (Figure 1, GLMM: overdispersion χ^2_{525} = 517.636, P = 0.582; fit χ^2_8 = 22.524, P = 0.004, R^2 = 9.43%). The GCN had significant positive associations with four species (Figure 2), including the smooth newt, common coot (*Fulica atra*), common moorhen, and pig. However, the GCN had significant negative associations with five species (Figure 2), including the common toad, ninespine stickleback, three-spined stickleback, grey squirrel, and common pheasant (*Phasianus colchicus*). Non-significant negative associations were also observed between the GCN and common carp, green-winged teal (*Anas carolinensis*), and badger (*Meles meles*). Only presence-absence of the smooth newt, common toad, three-spined stickleback, and grey squirrel were retained by model selection as explanatory variables for the GLMM (Figures 3a-d) with abiotic determinants. Results of analyses are summarised and compared to previously reported determinants in Table 1.

GCNs were more likely to occupy ponds with more amphibian species (Figure 1a). GCNs were detected in 51.97% of ponds (N = 152) containing smooth newts, but in only 11.91% of ponds (N = 42) with common toads (Figure S3a). GCN occurrence probability was lower in ponds with more fish species, and GCNs were absent from ponds with more than four fish species (Figure 1b). The GCN was only found in 14.29% (N = 56) and 6.67% (N = 15) of ponds inhabited by the three-spined stickleback and ninespine stickleback respectively (Figure S3b). In contrast, GCNs were more likely to occur in ponds with more waterfowl species (Figure 1c). GCNs occupied 41.67% (N = 48) and 35.81% (N = 215) of ponds with common coot and common moorhen respectively (Figure S3c). No significant relationships between terrestrial bird or mammal species richness and GCN occupancy were found (Figures 1d, e). Yet, species associations between the GCN and pig, common pheasant, and grey squirrel were identified. The GCN was detected in 37.14% of ponds (N = 140) where pig was present (Figure S3d) as opposed to 12.00% (N = 25) and 15.79% (N = 57) of ponds with common pheasant (Figure S3c) and grey squirrel (Figure S3d) records respectively. GCN occupancy positively correlated with overall vertebrate species richness (Figure 3h), irrespective of individual species associations.

3.4 Abiotic determinants of GCN occurrence

Five abiotic determinants were retained in the GLMM explaining GCN occupancy (GLMM: overdispersion χ^2_{490} = 413.394, P = 0.995; fit χ^2_8 = 11.794, P = 0.161, R^2 = 38.58%). The probability of GCN occurrence increased with greater pond depth but decreased in ponds with larger area, outflow, some other good terrestrial habitat, and

where ruderal habitat was absent (Table 1, Figures. 3e-g, i-j).

3.5 Abiotic determinants of vertebrate species richness

Our analysis (Table 2, GLMM: overdispersion χ^2_{494} = 431.959, P = 0.979; fit χ^2_8 = -42.708, P = 1.000, R^2 = 8.94%) revealed species richness was greater in ponds with outflow (Figure 4a), but reduced in those with some rough grass habitat compared to ponds with no or important rough grass habitat (Figure 4b). Overall quality of terrestrial habitat was also influential (Figure 4c), where ponds had higher species richness in areas considered to be poor or moderate habitat for GCNs. Species richness was reduced as percentages of terrestrial overhang (Figure 4d) and macrophyte cover increased (Figure 4e), but improved with pond density (Figure 4f).

3.6 Applicability of the HSI to GCN and community eDNA surveys

HSI score positively correlated with both GCN occurrence (GLMM: overdispersion χ^2_{501} = 506.763, P = 0.4198; fit χ^2_8 = 8.118, P = 0.422, R^2 = 4.99%) and vertebrate species richness (GLMM: overdispersion χ^2_{501} = 389.744, P = 0.999; fit χ^2_8 = -145.12, P = 1.000, R^2 = 1.10%). GCN occupancy probability (Table 1, Figure 3k) and vertebrate species richness (Table 2, Figure 4g) were improved in ponds with higher HSI score.

4. DISCUSSION

We have ground-truthed eDNA metabarcoding as a tool for ecological hypothesis testing using the community data generated by this tool in combination with environmental metadata for ponds. We identified biotic and abiotic determinants of GCN occupancy and wider vertebrate biodiversity, and evaluated the appropriateness of the HSI for eDNA survey. Our findings demonstrate the power of eDNA metabarcoding to reboot freshwater monitoring and research by providing biodiversity data en masse at low cost.

4.1 Pondscape biodiversity

eDNA metabarcoding detected six amphibian, 14 fish, 18 bird, and 22 mammal species across 532 UK ponds. This diverse species inventory emphasises the importance of ponds as habitat for aquatic taxa, but also as stepping stones for semi-aquatic and terrestrial taxa (De Meester et al., 2005; Hill et al., 2018) through provision of drinking, foraging, dispersal, and reproductive opportunities (Biggs et al., 2016; Klymus et al., 2017). Some species detections may be the result of eDNA transport from water bodies in the surrounding area (Hänfling et al., 2016) to ponds via inflow; however, this signifies the capacity of ponds to provide natural samples of freshwater biodiversity in the wider catchment.

4.2 Biotic determinants of GCN occurrence

GCN occurrence was positively associated with vertebrate species richness, which may support its status as an indicator or umbrella species for pond biodiversity (Gustafson, Pettersson, & Malmgren, 2006). GCNs were more likely to occur in ponds with higher amphibian species richness - particularly ponds containing smooth newts and absent of common toads. GCNs and smooth newts have similar habitat requirements and tend to breed in the same ponds (Denoël et al., 2013; Rannap et al., 2009a; Skei et al., 2006), with >60% overlap reported (Rannap & Briggs, 2006). However, smooth newts can inhabit a broader range of habitat (Rannap & Briggs, 2006; Skei et al., 2006) than GCNs, which depend on larger, deeper ponds with abundant macrophytes and no fish located in open, semi-rural landscapes (Denoël et al., 2013). Common toads inhabit fish-containing ponds (Manenti & Pennati, 2016) which may explain their negative association with GCNs, but newts may also predate toad eggs and larvae (Langton et al., 2001).

GCNs were negatively associated with higher fish species richness, and specifically the presence of three-spined and ninespine sticklebacks - fish that are common in and typical of ponds. All GCN life stages may be predated by fishes (Langton et al., 2001) and negative effects of fish presence on GCN occupancy, distribution, and abundance are repeatedly reported (Denoël & Ficetola, 2008; Denoël et al., 2013; Hartel et al., 2010; Joly et al., 2001; Rannap & Briggs, 2006; Rannap et al., 2009a, b; Skei et al., 2006). Three-spined sticklebacks predate GCN eggs and larvae (Jarvis, 2010; McLee & Scaife, 1992), and have non-consumptive effects on GCN embryos (Jarvis, 2010). GCN larvae were also found to alter their behaviour when exposed to the predatory three-spined stickleback, but not the non-predatory

crucian carp (Jarvis, 2012), another fish characteristic of ponds.

In our study, we detected GCNs in 50% of ponds inhabited by crucian carp, but <20% of ponds containing large and/or predatory fishes, e.g. pike and three-spined stickleback. Although fewer ponds contained crucian carp than pike or stickleback, previous research also indicates large and/or predatory fish are more detrimental to GCN occurrence (Chan, 2011; Hartel et al., 2010; Skei et al., 2006). Crucian carp do not hinder GCN oviposition, larval behaviour, or recruitment success (Chan, 2011; Jarvis, 2012), or pond invertebrate and macrophyte diversity (Stefanoudis et al., 2017). In contrast, common carp foraging reduces invertebrate density and macrophyte cover (Maceda-Veiga, López, & Green, 2017), which lowers GCN reproductive and foraging success and heightens predator exposure (Chan, 2011; Gustafson et al., 2006; Rannap & Briggs, 2006). The crucian carp and common carp are both included among fish species assumed to negatively impact GCNs and whose presence-absence is assessed for the GCN HSI (ARG-UK, 2010). However, it is evident that crucian carp do not directly predate GCNs or indirectly alter their behaviour, reproductive success, or habitat. Therefore, we advocate a systematic re-evaluation of problematic fish species for GCN conservation.

Unexpectedly, GCNs were positively associated with waterfowl species richness, namely presence of common coot and common moorhen. These waterfowl species share macrophytes and macroinvertebrates as resources with amphibians, feeding on both directly (Paillisson & Marion, 2001; Perrow et al., 1997; Wallau et al., 2010). Coots and moorhens crop emergent macrophytes to search for invertebrate prey (Paillisson & Marion, 2001; Wallau et al., 2010), which may indirectly benefit GCN foraging. Although coots can also pull up submerged vegetation and damage vegetation banks (Lauridsen, Jeppesen, & Andersen, 1993), diet is macrophyte-

dominated in late summer and autumn (Perrow et al., 1997) and unlikely to impact GCN breeding in spring (Langton et al., 2001). The positive association identified here between GCNs and these waterfowl most likely reflects a shared preference for macrophyte-rich ponds.

The GCN had negative associations with the common pheasant and grey squirrel, but a positive association with pig. GCNs are at risk of predation during terrestrial as well as aquatic phases of their life history (Gustafson et al., 2011; Langton et al., 2001). There have been anecdotal records of pheasant predation on herpetofauna, including the GCN (Rice, 2016), which our results would support. However, the terrestrial associations identified may instead reflect land-use and indirect interaction. GCNs prefer ponds surrounded by deciduous forest and pasture (Gustafson et al., 2011), where pheasants and domestic animals are commonplace, over those in urban areas (Denoël & Ficetola, 2008; Hartel et al., 2010), which support dense populations of grey squirrels.

4.3 Abiotic determinants of GCN occurrence

GCNs were more likely to occupy deeper ponds, but less likely to inhabit larger ponds, with outflow present, no ruderal habitat, and some other good terrestrial habitat. Consistent with our results, pond depth was previously found to positively influence GCN occupancy (Denoël & Ficetola, 2008). Shallow ponds can be inhospitable due to drying or freezing and may contain less prey, but detrimental effects are often observed in open farmland (Denoël & Ficetola, 2008). Although our results indicate GCNs prefer larger ponds, pond area does not always influence

occurrence (Denoël & Ficetola, 2008; Gustafson et al., 2011; Maletzky, Kyek, & Goldschmid, 2007) and was deemed a poor predictor of reproductive success (Vuorio et al., 2013). GCNs have been found to utilise small and large ponds (Rannap & Briggs, 2006; Skei et al., 2006); however, very small ponds (<124 m^2) may be unable to support all life stages, and larger ponds may contain fish and experience eutrophication from agricultural or polluted run-off (Rannap & Briggs, 2006). Effects of pond outflow (facilitated by drains, pipes or streams) are understudied, whereas inflow impacts biodiversity via polluted agricultural run-off and connections to streams and rivers containing large, predatory fish. Based on our findings that show outflow presence deters GCN occupation, we suggest outflow minimises fluctuations in pond depth (Freshwater Habitats Trust, 2015) and alters patterns of colonisation and community structure to those of ponds which fluctuate with precipitation.

Our results support the importance of good terrestrial habitat to GCNs for shelter, foraging and dispersal (Langton et al., 2001). We observed higher GCN occurrence in ponds surrounded by ruderal or other good quality terrestrial habitat. However, the majority of terrestrial habitat variables were not retained by our model selection. Hartel et al. (2010) also found landscape variables, excluding urbanisation, were inadequate predictors of GCN distribution, although their study area was mostly rural and thus optimal for amphibians. Conversely, research on more diverse landscapes found GCN occupancy was lower in coniferous forest, yet higher in deciduous or herb-rich forest and pasture (Gustafson et al., 2011; Vuorio et al., 2013). In our study, the metadata available were qualitative, preventing detailed analyses on terrestrial habitat quality in relation to GCN occupancy. Better understanding of occupancy and interactions with terrestrial species could be achieved with quantitative data on terrestrial habitat type, density, distance to ponds, and species

utilisation, but this is a phenomenal task for large numbers of ponds across a vast landscape (Denoël & Ficetola, 2008).

4.4 Abiotic determinants of vertebrate species richness

Vertebrate species richness was higher in ponds where outflow was present, with poor or moderate overall terrestrial habitat for GCNs, and more ponds nearby. Conversely, species richness was lower in ponds with higher percentages of macrophyte cover and terrestrial overhang, and absent of or surrounded by some rough grass habitat. We compare our results to previous studies of aquatic species richness, but these largely focus on species assemblages or guilds, primarily macrophytes, macroinvertebrates, and amphibians. Outflow and inflow have been understudied in relation to pond biodiversity, although outflow may release harmful pollutants and pathogens (Beutel & Larson, 2015) that would accumulate in a closed pond system. Species richness increased as pond density increased, which echoes the positive relationship between pond density and macrophyte and macroinvertebrate richness observed by Gledhill, James, & Davies (2008). These findings combined again highlight the importance of pond networks for aquatic and non-aquatic taxa (Hill et al., 2018). Shade was identified as a principal, negative driver of macroinvertebrate and macrophyte diversity in ponds (Sayer et al., 2012), but canopy and macrophyte cover positively influenced amphibian species richness (Piha, Luoto, & Merilä, 2007). In our study, vertebrate diversity decreased as shading increased, which corroborates findings of Sayer et al. (2012). Plentiful rough grass habitat may create more ecological niches and foraging opportunity for different vertebrates, but quantitative data on type and

abundance of terrestrial habitat are needed to understand species preferences.

4.5 Applicability of the HSI to GCN and community eDNA surveys

We found the HSI can predict eDNA-based GCN occupancy at the pondscape. This contradicts previous conventional studies which deemed the index inappropriate for predicting GCN occupancy or survival probabilities (Unglaub, Steinfartz, Drechsler, & Schmidt, 2015). Our results also suggest some indices comprising the GCN HSI, for example, outflow and terrestrial habitat, represent suitability criteria for other biodiversity. An adapted HSI, designed to predict species richness, could help select areas for management and enhancement of aquatic and non-aquatic biodiversity. Nonetheless, the GCN HSI also confers protection to pond biodiversity by identifying optimal habitat for pond creation and restoration to encourage populations of this threatened amphibian. The HSI is not without issue due to qualitative data used for score calculation and subjective estimation of indices (Oldham et al., 2000). For future application of this index in GCN eDNA survey, we recommend metabarcoding to quantify some qualitatively assessed indices (e.g. water quality via macroinvertebrate diversity, fish and waterfowl presence) alongside GCN detection. Provided rigorous spatial and temporal sampling are undertaken, eDNA metabarcoding can also generate site occupancy data to estimate relative species abundance (Hänfling et al., 2016; Valentini et al., 2016).

4.6 Prospects of eDNA metabarcoding for conservation, management, and research

We have demonstrated the effectiveness of eDNA metabarcoding for landscape-scale biodiversity monitoring and ecological hypothesis testing. We investigated associations between aquatic and non-aquatic vertebrates, and combined metabarcoding with environmental metadata to revisit important ecological hypotheses at an unprecedented scale. Our findings indicate preferred habitat of a threatened amphibian. GCNs were more likely to occupy ponds where smooth newts were present, but common toads, three-spined sticklebacks, and grey squirrels were absent. GCNs prefer small but deep ponds that are absent of outflow, and surrounded by ruderal and good quality terrestrial habitat. These findings will guide management in the face of increasing land-use and habitat fragmentation - a poignant issue as protective legislation for the GCN in the UK is changing. Whilst conservation of threatened species and their habitat should be a priority, the bigger picture should not be ignored. eDNA metabarcoding could enhance our understanding of freshwater networks - particularly pondscapes - to enable more effective monitoring, protection, and management of aquatic and non-aquatic biodiversity. We are only now beginning to realise and explore these opportunities.

ACKNOWLEDGEMENTS

This work was funded by the University of Hull. We would like to thank Jennifer Hodgetts (Fera Science Ltd) for assisting with sample collection, and Jianlong Li

(University of Hull) for primer design and advice on laboratory protocols. Tissue samples for primer validation and Sanger sequencing were provided by Andrew Buxton and Richard Griffiths (DICE, University of Kent) under licence from Natural England, and Barbara Mabel and Elizabeth Kilbride (University of Glasgow).

DATA ACCESSIBILITY

Raw sequence reads have been archived on the NCBI Sequence Read Archive (Bioproject: PRJNA417951; SRA accessions: SRR6285413 - SRR6285678). Jupyter notebooks, R scripts and corresponding data are deposited in a dedicated GitHub repository (https://github.com/HullUni-bioinformatics/Harper_et_al_2018) which has been permanently archived (https://doi.org/10.5281/zenodo.1304107).

AUTHOR CONTRIBUTIONS

B.H., L.R.H., L.L.H and N.B. conceived and designed the study. H.C.R. and N.B. contributed samples for processing. L.R.H. performed laboratory work and analysed the data. I.P.A. and E.L. offered advice on and supervised sequencing. C.H. assisted with bioinformatics analysis. P.B. and S.P. contributed datasets for analysis. L.R.H. wrote the manuscript, which all authors revised.

REFERENCES

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, *60*, 255–265. https://doi.org/10.2307/2334537

ARG-UK. (2010). *ARG UK Advice Note 5: Great Crested Newt Habitat Suitability Index*. Amphibian and Reptile Groups of the United Kingdom.

Bálint, M., Nowak, C., Márton, O., Pauls, S., Wittwer, C., Aramayo, J. L., … Jansen, M. (2017). Twenty-five species of frogs in a liter of water: eDNA survey for exploring tropical frog diversity. *bioRxiv*, 176065. https://doi.org/10.1101/176065

Barnes, M. A., & Turner, C. R. (2015). The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, *17*, 1–17. https://doi.org/10.1007/s10592-015-775-4

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. https:/doi.org/10.18637/jss.v067.i01

Beutel, M. W., & Larson, L. (2015). Pathogen removal from urban pond outflow using rock biofilters. *Ecological Engineering*, *78*, 72–78. https://doi.org/10.1016/j.ecoleng.2014.05.001

Biggs, J., Ewald, N., Valentini, A., Gaboriaud, C., Dejean, T., Griffiths, R. A., … Dunn, F. (2015). Using eDNA to develop a national citizen science-based monitoring programme for the great crested newt (*Triturus cristatus*). *Biological Conservation*, *183*, 19–28. https://doi.org/10.1016/j.biocon.2014.11.029

Biggs, J., von Fumetti, S., & Kelly-Quinn, M. (2016). The importance of small waterbodies for biodiversity and ecosystem services: implications for policy makers. *Hydrobiologia*, *793*, 3–39. https://doi.org/10.1007/s10750-016-3007-0

Bjørnstad, O. N. (2009). ncf: spatial nonparametric covariance functions. R package version 1.1-7.

Chan, K. (2011). Can great crested newts (*Triturus cristatus*) coexist with fish?. Masters Thesis. University College London (UCL).

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., … Bernatchez, L. (2017). Environmental DNA metabarcoding: transforming how we survey animal and plant communities. *Molecular Ecology*, *26*, 5872–5895. https://doi.org/10.1111/mec.14350

De Meester, L., Declerck, S., Stoks, R., Louette, G., Van De Meutter, F., De Bie, T., … Brendonck, L. (2005). Ponds and pools as model systems in conservation biology, ecology and evolutionary biology. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *15*, 715–725. https://doi.org/10.1002/aqc.748

Denoël, M., & Ficetola, G. F. (2008). Conservation of newt guilds in an agricultural landscape of Belgium: the importance of aquatic and terrestrial habitats. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *18*, 714–728. https://doi.org/10.1002/aqc.853

Denoël, M., Perez, A., Cornet, Y., & Ficetola, G. F. (2013). Similar local and landscape processes affect both a common and a rare newt species. *PloS ONE*, *8*, e62727. https://doi.org/10.1371/journal.pone.0062727

Evans, N. T., Li, Y., Renshaw, M. A., Olds, B. P., Deiner, K., Turner, C. R., … Pfrender, M. E. (2017). Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. *Canadian Journal of Fisheries and Aquatic Sciences*, *74*, 1362-1374. https://doi.org/10.1139/cjfas-2016-0306

Ficetola, G. F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., … Pompanon, F. (2010). An *In silico* approach for the evaluation of DNA barcodes. *BMC Genomics*, *11*, 434. https://doi.org/10.1186/1471-2164-11-434

Freshwater Habitats Trust (2015). Pond Habitat Survey. Survey Manual. Freshwater Habitats Trust, Oxford.

Gledhill, D. G., James, P., & Davies, D. H. (2008). Pond density as a determinant of aquatic species richness in an urban landscape. *Landscape Ecology*, *23*, 1219–1230. https://doi.org/10.1007/s10980-008-9292-x

Griffith, D., Veech, J., & Marsh, C. (2016). cooccur: Probabilistic Species Co-Occurrence Analysis in R. *Journal of Statistical Software*, *69*, 1–17. https://doi.org/10.18637/jss.v069.c02

Gustafson, D. H., Malmgren, J. C., & Mikusiński, G. (2011). Terrestrial Habitat Predicts use of Aquatic Habitat for Breeding Purposes — A Study on the Great Crested Newt (*Triturus cristatus*). *Annales Zoologici Fennici*, *48*, 295–307. https://doi.org/10.5735/086.048.0504

Gustafson, D. H., Pettersson, C. J., & Malmgren, J. C. (2006). Great crested newts (*Triturus cristatus*) as indicators of aquatic plant diversity. *The Herpetological Journal*, *16*, 347–352.

Hänfling, B., Lawson Handley, L., Read, D. S., Hahn, C., Li, J., Nichols, P., … Winfield, I. J. (2016). Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. *Molecular Ecology*, *25*, 3101–3119. https://doi.org/10.1111/mec.13660

Harper, L. R., Lawson Handley, L., Hahn, C., Boonham, N., Rees, H. C., Gough, K. C., … Hänfling, B. (2018). Needle in a haystack? A comparison of eDNA metabarcoding and targeted qPCR for detection of the great crested newt (*Triturus cristatus*). *Ecology and Evolution*, *8*, 6330–6341. https://doi.org/10.1002/ece3.4013

Hartel, T., Nemes, S., & Oellerer, K. (2010). Using connectivity metrics and niche modelling to explore the occurrence of the northern crested newt *Triturus cristatus* (Amphibia, Caudata) in a traditionally managed landscape. *The Environmentalist*, *37*, 195–200. https://doi.org/10.1017/S037689291000055X

Hering, D., Borja, A., Jones, J. I., Pont, D., Boets, P., Bouchez, A., … Kelly, M. (2018). Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive. *Water Research*, *138*, 192–205. https://doi.org/10.1016/j.watres.2018.03.003

Hervé, M. (2015). RVAideMemoire: Diverse basic statistical and graphical functions. R package version 0.9-36. 2014.

Hill, M. J., Hassall, C., Oertli, B., Fahrig, L., Robson, B. J., Biggs, J., … Wood, P. J. (2018). New policy directions for global pond conservation. *Conservation Letters*, *e12447*. https://doi.org/10.1111/conl.12447

Hosmer, D. W., & Lemeshow, S. (2000). Multiple Logistic Regression. In *Applied Logistic Regression* (pp. 31–46). John Wiley & Sons, Inc.

Jarvis, L. E. (2010). Non-consumptive effects of predatory three-spined sticklebacks (*Gasterosteus aculeatus*) on great crested newt (*Triturus cristatus*) embryos. *The Herpetological Journal*, *20*, 271–275.

Jarvis, L.E. (2012). Microhabitat preferences of the great crested newt *Triturus cristatus* in a woodland area. PhD Thesis. Open University.

Joly, P., Miaud, C., Lehmann, A., & Grolet, O. (2001). Habitat Matrix Effects on Pond Occupancy in Newts. *Conservation Biology*, *15*, 239–248. https://doi.org/10.1111/j.1523-1739.2001.99200.x

Kelly, R. P., O'Donnell, J. L., Lowell, N. C., Shelton, A. O., Samhouri, J. F., Hennessey, S. M., … Williams, G. D. (2016). Genetic signatures of ecological diversity along an urbanization gradient. *PeerJ*, *4*, e2444. https://doi.org/10.7717/peerj.2444

Klymus, K. E., Richter, C. A., Thompson, N., & Hinck, J. E. (2017). Metabarcoding of Environmental DNA Samples to Explore the Use of Uranium Mine Containment Ponds as a Water Source for Wildlife. *Diversity*, *9*, 1-18. https://doi.org/10.3390/d9040054

Langton, T., Beckett, C. & Foster, J. (2001). Great crested newt conservation handbook. Froglife, Halesworth.

Lauridsen, T. L., Jeppesen, E., & Andersen, F. Ø. (1993). Colonization of submerged macrophytes in shallow fish manipulated Lake Væng: impact of sediment composition and waterfowl grazing. *Aquatic Botany*, *46*, 1–15. https://doi.org/10.1016/0304-3770(93)90061-Z

Lawson Handley, L. (2015). How will the "molecular revolution" contribute to biological recording? *Biological Journal of the Linnean Society*, *115*, 750–766. https://doi.org/10.1111/bij.12516

Lele, S. R., Keim, J. L., & Solymos, P. (2016). ResourceSelection: Resource Selection (Probability) Functions for Use-Availability Data.

Maceda-Veiga, A., López, R., & Green, A. J. (2017). Dramatic impact of alien carp *Cyprinus carpio* on globally threatened diving ducks and other waterbirds in Mediterranean shallow lakes. *Biological Conservation*, *212*, 74–85. https://doi.org/10.1016/j.biocon.2017.06.002

Maletzky, A., Kyek, M., & Goldschmid, A. (2007). Monitoring status, habitat features and amphibian species richness of crested newt (*Triturus cristatus* superspecies) ponds at the edge of the species range (Salzburg, Austria). *Annales de Limnologie - International Journal of Limnology*, *43*, 107–115. https://doi.org/10.1051/limn/2007017

Manenti, R., & Pennati, R. (2016). Environmental factors associated with amphibian breeding in streams and springs: effects of habitat and fish occurrence. *Amphibia-Reptilia*, *37*, 237–242. https://doi.org/10.1163/15685381-00003040

Mathews, F., Kubasiewicz, L.M., Gurnell, J., Harrower, C.A., McDonald, R.A. & Shore, R.F. (2018). A Review of the Population and Conservation Status of British Mammals: Technical Summary. A report by the Mammal Society under contract to Natural England, Natural Resources Wales and Scottish Natural Heritage. Natural England, Peterborough.

Mazerolle, M. J. (2016). AICcmodavg: Model Selection and Multimodel Inference Based on (Q)AIC(c).

McLee, A. G., & Scaife, R. W. (1992). The colonisation by great crested newts (*Triturus cristatus*) of a water body following treatment with a piscicide to remove a large population of sticklebacks (*Gasterosteus aculeatus*). *Herpetological Bulletin*, *42*, 6–9.

Muha, T. P., Rodríguez-Rey, M., Rolla, M., & Tricarico, E. (2017). Using Environmental DNA to Improve Species Distribution Models for Freshwater Invaders. *Frontiers in Ecology and Evolution*, *5*, 158. https://doi.org/10.3389/fevo.2017.00158

Oldham, R. S., Keeble, J., Swan, M. J. S., & and Jeffcote, M. (2000). Evaluating the suitability of habitat for the great crested newt (*Triturus cristatus*). *The Herpetological Journal*, *10*, 143–155.

Olds, B. P., Jerde, C. L., Renshaw, M. A., Li, Y., Evans, N. T., Turner, C. R., … Lamberti, G. A. (2016). Estimating species richness using environmental DNA. *Ecology and Evolution*, *6*, 4214–4226. https://doi.org/10.1002/ece3.2186

Paillisson, J.-M., & Marion, L. (2001). Interaction between coot (*Fulica atra*) and waterlily (*Nymphaea alba*) in a lake: the indirect impact of foraging. *Aquatic Botany*, *71*, 209–216. https://doi.org/10.1016/S0304-3770(01)00182-6

Perrow, M. R., Schutten, H. J., Howes, J. R., Holzer, T., Madgwick, F. J., & Jowitt, A. (1997). Interactions between coot (*Fulica atra*) and submerged macrophytes: the role of birds in the restoration process. *Hydrobiologia*, *342*, 241–255. https://doi.org/10.1023/A:1017007911190

Piha, H., Luoto, M., & Merilä, J. (2007). Amphibian occurrence is influenced by current and historic landscape characteristics. *Ecological Applications*, *17*, 2298–2309. https://doi.org/10.1890/06-1509.1

Rannap, R. & Briggs, L. (2006). THE CHARACTERISTICS OF GREAT CRESTED NEWT TRITURUS CRISTATUS'BREEDING PONDS. Protection of the Great Crested Newt. Project Report. The experiences of LIFE-Nature project "Protection of *Triturus cristatus* in the Eastern Baltic Region" LIFE04NAT/EE/000070. Ministry of the Environment of the Republic of Estonia, Tallinn.

Rannap, R., Lõhmus, A., & Briggs, L. (2009a). Niche position, but not niche breadth, differs in two coexisting amphibians having contrasting trends in Europe. *Diversity and Distributions*, *15*, 692–700. https://doi.org/10.1111/j.1472-4642.2009.00575.x

Rannap, R., Lõhmus, A., & Briggs, L. (2009b). Restoring ponds for amphibians: a success story. *Hydrobiologia*, *634*, 87–95. https://doi.org/10.1007/s10750-009-9884-8

Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R. M., & and Gough, K. C. (2014). REVIEW: The detection of aquatic animal species using environmental DNA–a review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*, *51*, 1450–1459. https://doi.org/10.1111/1365-2664.12306

Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P., & Coissac, E. (2011). ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, *39*, e145. https://doi.org/10.1093/nar/gkr732

Rice, C.N. (2016). Abundance, impacts and resident perceptions of non-native common pheasants (*Phasianus colchicus*) in Jersey, UK Channel Islands. Masters Thesis. University of Kent.

Sayer, C., Andrews, K., Shilland, E., Edmonds, N., Edmonds-Brown, R., Patmore, I., … Axmacher, J. (2012). The role of pond management for biodiversity conservation in an agricultural landscape. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *22*, 626–638. https://doi.org/10.1002/aqc.2254

Skei, J. K., Dolmen, D., Rønning, L., & Ringsby, T. H. (2006). Habitat use during the aquatic phase of the newts *Triturus vulgaris* (L.) and *T. cristatus* (Laurenti) in central Norway: proposition for a conservation and monitoring area. *Amphibia-Reptilia*, *27*, 309–324. https://doi.org/10.1163/156853806778189972

Stefanoudis, P. V., Sayer, C. D., Greaves, H. M., Davidson, T. A., Robson, H., Almeida, D., & Smith, E. (2017). Consequences of fish for cladoceran, water beetle and macrophyte communities in a farmland pond landscape: implications for conservation. *Fundamental and Applied Limnology*, *190*, 141–156. https://doi.org/10.1127/fal/2017/1004

Unglaub, B., Steinfartz, S., Drechsler, A., & Schmidt, B. R. (2015). Linking habitat suitability to demography in a pond-breeding amphibian. *Frontiers in Zoology*, *12*. https://doi.org/10.1186/s12983-015-0103-3

Ushio, M., Fukuda, H., Inoue, T., Makoto, K., Kishida, O., Sato, K., … Miya, M. (2017). Environmental DNA enables detection of terrestrial mammals from forest pond water. *Molecular Ecology Resources*, *17*, e63-e65. https://doi.org/10.1111/1755-0998.12690

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., … Dejean, T. (2016). Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, *25*, 929–942. https://doi.org/10.1111/mec.13428

Vuorio, V., Heikkinen, R. K., & Tikkanen, O.-P. (2013). Breeding Success of the Threatened Great Crested Newt in Boreal Forest Ponds. *Annales Zoologici Fennici*, *50*, 158–169. https://doi.org/10.5735/086.050.0303

Wallau, G. L., Della-Flora, F., Bueno, A. S., Corso, J., Ortiz, M. F., & Cáceres, N. C. (2010). Behaviour of the Common Moorhen in Rio Grande do Sul, Brazil. *Acta Ethologica*, *13*, 127–135. https://doi.org/10.1007/s10211-010-0082-5

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer New York, USA.

Zhang, Z., Schwartz, S., Wagner, L., & Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, *7*, 203–214. https://doi.org/10.1089/10665270050081478

Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. Springer New York, USA.

Table 1 Summary of established and novel determinants of GCN occupancy. Reported effects on GCN occupancy in the literature and hypothesised effects on eDNA-based GCN occupancy are given for each determinant. Any determinants not reported in the literature are listed as UNK. Direction of observed effects on eDNA-based GCN occupancy determined by each analysis (GLMM assessing species richness in each vertebrate group, N = 532; co-occur analysis, N = 532; GLMM combining abiotic and biotic factors N = 504; and GLMM assessing HSI score, N = 504) are given. No, negative and positive effects are listed as 0, - and + respectively. For categorical variables with more than one level, effect size and standard error (SE) are only given for levels reported in the model summary. Test statistic is for LRT used and significant P-values (<0.05) are in bold. Variables included for model selection but not retained in the final model are listed as NR. Co-occur analysis was not applicable (NA) to abiotic factors.

Determinant	**Effect reported**	**Hypothesised effect**	**Analysis**					
			cooccur		**GLMM**			
			Effect	***P***	**DF**	**Effect size (SE)**	**χ^2**	***P***
Fish	-/0	-			1	-0.239 (0.124)	4.065	**0.044**
Three-spined stickleback	-	-	-	**0.009**	1	-1.432 (0.561)	9.453	**0.002**
Ninespine stickleback	-	-	-	**0.047**				
Common carp	-	-	-	0.070				
Crucian carp	-	-	0	NA				
Waterfowl	-	-			1	0.617 (0.181)	13.050	**<0.001**
Coot	UNK		+	**0.023**				
Moorhen	UNK		+	**<0.001**				
Green-winged teal	UNK		-	0.099				
Amphibians	UNK				1	0.558 (0.149)	16.641	**<0.001**
Smooth newt	+	+	+	**< 0.001**	1	1.081 (0.303)	7.434	**<0.001**
Common toad	UNK		-	**0.009**	1	-1.635 (0.696)	8.228	**0.004**
Terrestrial birds	UNK				1	-0.335 (0.291)	1.444	0.230
Common pheasant	UNK		-	**0.048**				
Terrestrial mammals	UNK				1	0.028 (0.091)	0.095	0.758
Grey squirrel	UNK		-	**0.018**	1	-1.591 (0.534)	12.432	**<0.001**
Badger	UNK		-	0.099				
Pig	UNK		+	**0.004**				
Cow	UNK		+	0.097				
Pond area	-/+	-	NA	NA	1	-0.0004 (0.0002)	6.453	**0.011**
Pond density	+	+	NA	NA		NR		
Pond depth	+	+	NA	NA	1	0.282 (0.139)	4.266	**0.039**
Water quality	+	+	NA	NA		NR		
Outflow	UNK		NA	NA	1	-0.713 (0.359)	4.467	**0.035**
Macrophyte cover	-/+	-	NA	NA		NR		
Shading	-/+	-	NA	NA		NR		

Woodland	+	+	NA	NA		NR		
Grassland	+	+	NA	NA		NR		
HSI	0/+	+	NA	NA	1	3.020 (0.791)	15.709	**<0.001**
Ruderal	UNK		NA	NA	2		6.507	**0.039**
None						-0.617 (0.527)		
Some						0.032 (0.528)		
Other good terrestrial habitat	UNK		NA	NA	2		7.918	**0.019**
None						0.428 (0.429)		
Some						-0.316 (0.424)		
Species richness	UNK		NA	NA	1	0.527 (0.105)	60.267	**<0.001**

834

Table 2 Summary of analyses testing for variation in vertebrate species richness of ponds (*N* = 504), attributable to aquatic and terrestrial habitat. Notably, HSI score was analysed in a GLMM independent of other habitat variables. For categorical variables with more than one level, effect size and standard error (SE) are only given for levels reported in the model summary. Test statistic is for LRT used and significant P-values (<0.05) are in bold.

Model variables	**DF**	**Effect size (SE)**	χ^2	***P***
Outflow	1		11.220	**<0.001**
Present		0.214 (0.063)		
Rough grass	2		16.715	**<0.001**
Some		-0.297 (0.074)		
None		-0.1402 (0.0795)		
Overall terrestrial habitat	2		8.244	**0.016**
Moderate		0.216 (0.078)		
Poor		0.115 (0.089)		
Pond density	1	0.006 (0.003)	4.564	**0.033**
Macrophyte cover	1	-0.002 (0.001)	4.117	**0.043**
Terrestrial overhang	1	-0.0026 (0.0008)	9.575	**0.002**
HSI score	1	0.459 (0.002)	5.034	**0.025**

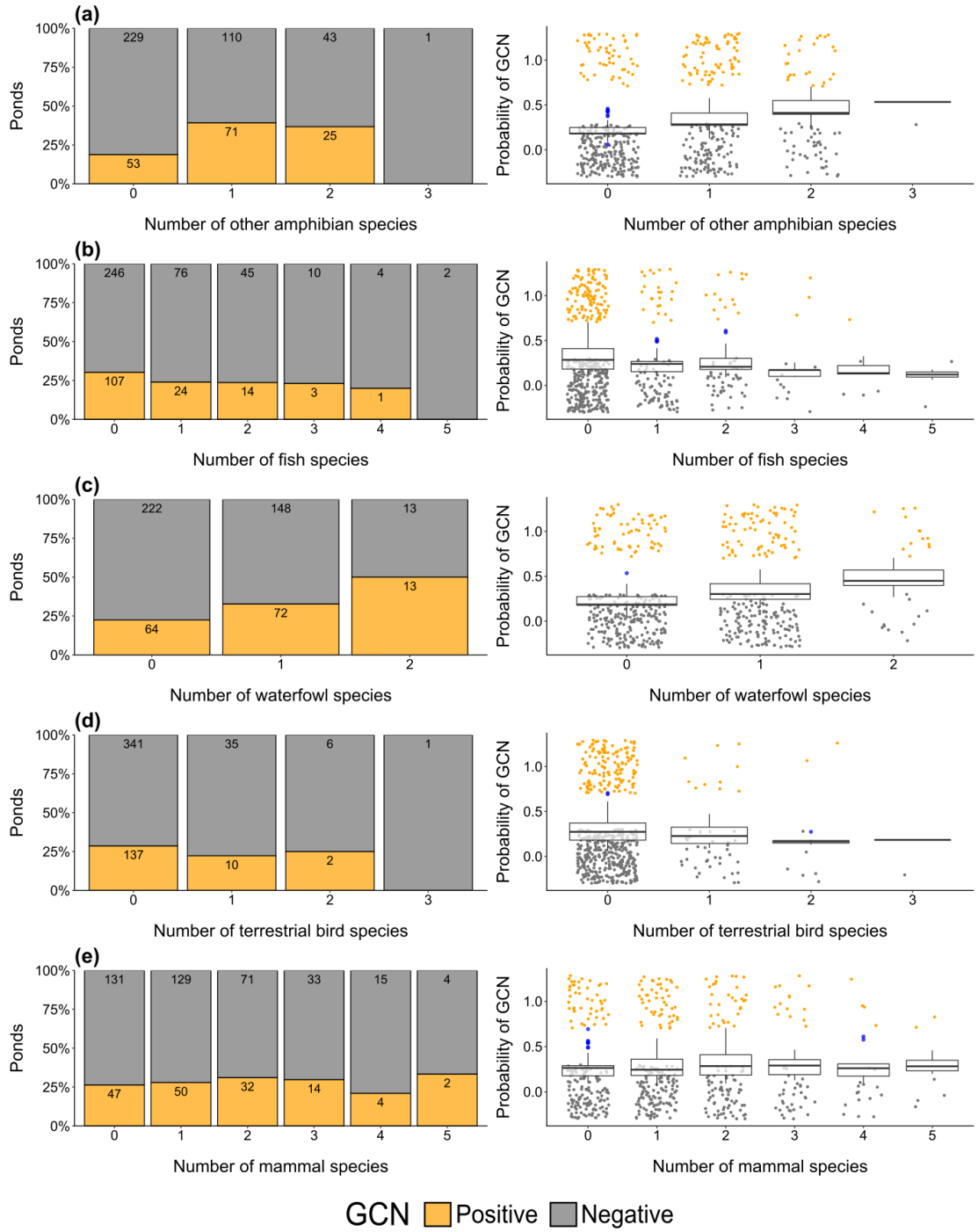

Figure 1 GCN presence (orange) and absence (grey) in relation to number of species from different vertebrate groups detected by eDNA (N = 532 ponds): (a) other amphibians, (b) fish, (c) waterfowl, (d) terrestrial birds, and (e) mammals. Observed proportion of ponds with and without GCN (left) is plotted alongside predicted probability of GCN occurrence in ponds as determined by the binomial GLMM (right). Numbers on barplots of observed occupancy are the number of ponds for each category. In plots showing predicted GCN occupancy, the observed data is shown as points which have been jittered around 0 and 1 to clarify variation in point density. Blue points are outliers and boxes are the model predictions.

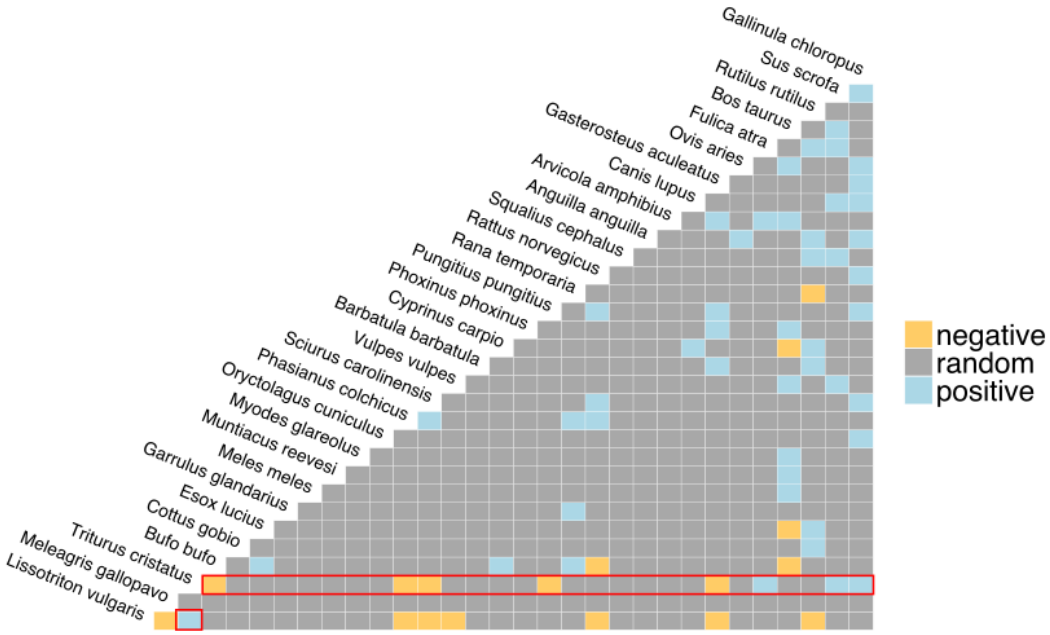

Figure 2 Heat map showing significant (P < 0.05) positive and negative species associations determined by the probabilistic co-occurrence model for the eDNA metabarcoding presence-absence data (N = 532 ponds). Species names are positioned to indicate the columns and rows that represent their pairwise relationships with other species. Species are ordered by those with the most negative interactions to those with the most positive interactions (left to right). Associations relevant to GCN are highlighted in red.

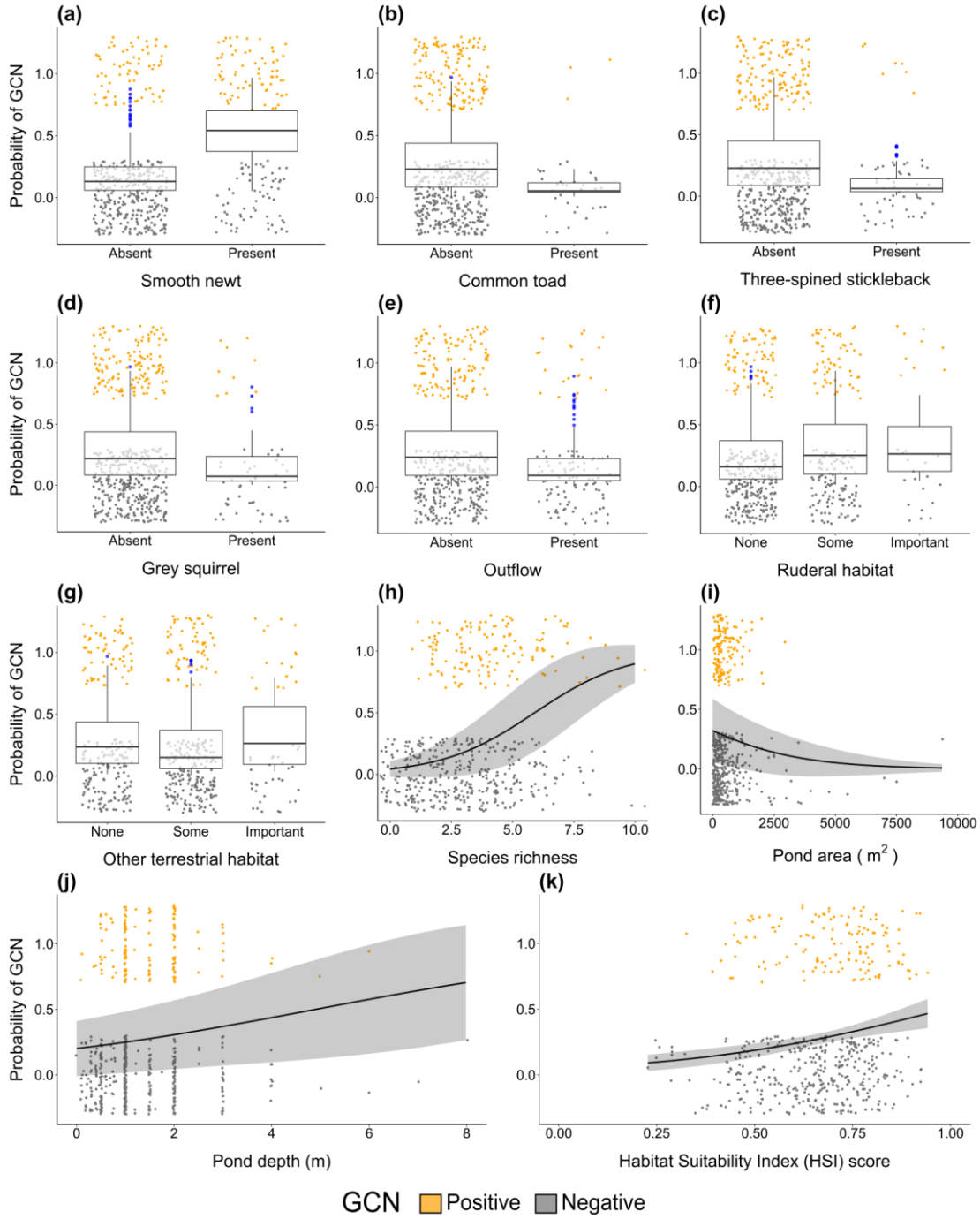

Figure 3 Biotic and abiotic determinants of GCN occurrence, as predicted by the binomial GLMMs (N = 504 ponds): (a) smooth newt occurrence, (b) common toad occurrence, (c) three-spined stickleback occurrence, (d) grey squirrel occurrence, (e) pond outflow, (f) ruderal habitat, (g) other good quality terrestrial habitat, (h) species richness, (i) pond area, (j) pond depth, and (k) HSI score. The 95% CIs, as calculated using the predicted GCN probability values and standard error for these predictions, are given for each relationship. The observed GCN presence (orange) and absence (grey) data are also displayed as points, which have been jittered around 0 and 1 to clarify variation in point density, against the predicted relationships (boxes/lines). Outliers are indicated by blue points.

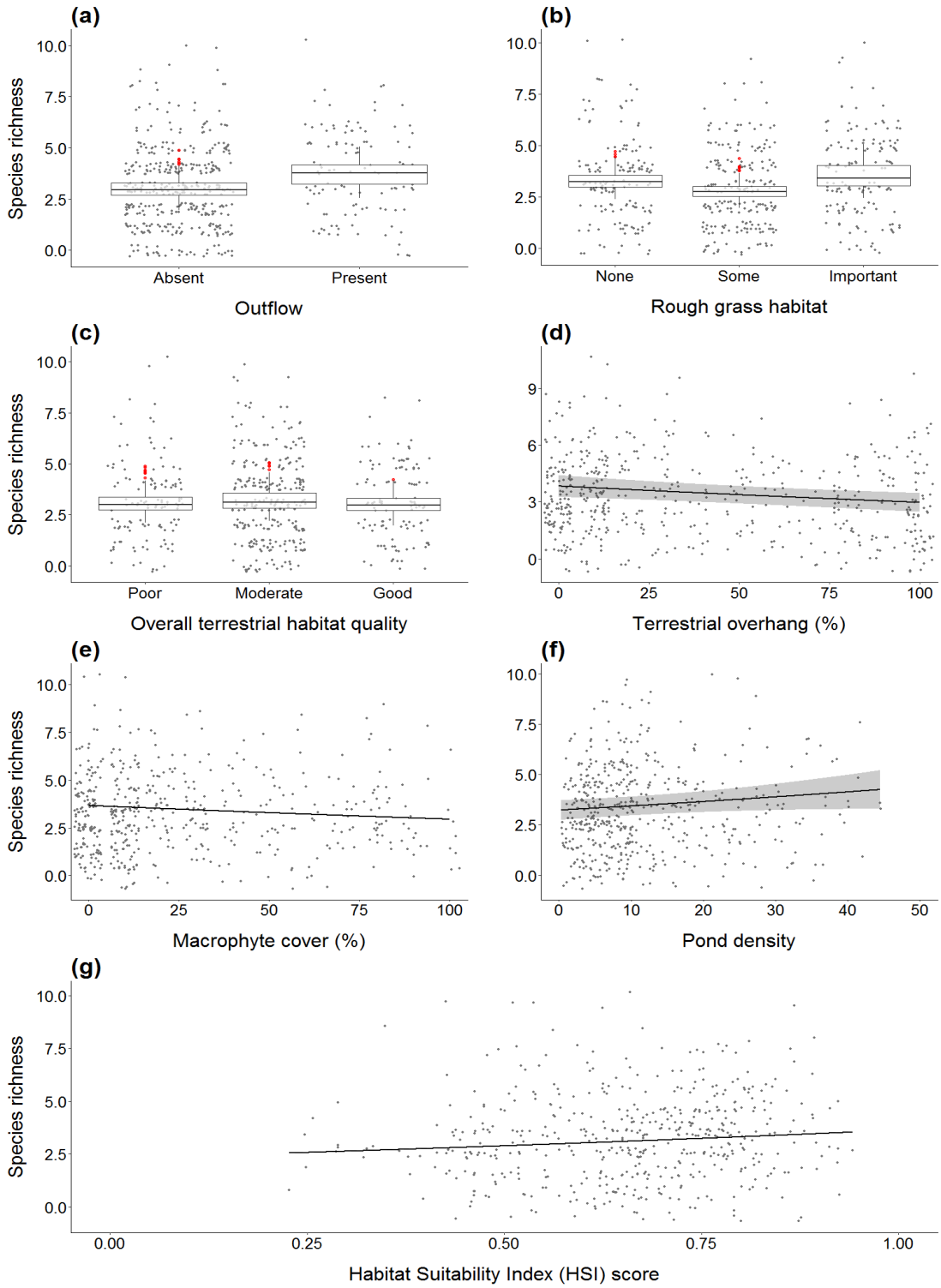

Figure 4 Abiotic and biotic determinants of vertebrate species richness, as predicted by the Poisson GLMMs (N = 504 ponds): (a) outflow, (b) rough grass habitat, (c) overall quality of terrestrial habitat, (d) percentage of terrestrial overhang, (e) percentage of macrophyte cover, (f) pond density, and (g) HSI score. The 95% CIs, as calculated using the predicted species richness values and standard error for these predictions, are given for each relationship. The observed data are also displayed as points, which have been jittered around 0 and 10 to clarify variation in point density, against the predicted relationships (boxes/lines). Outliers are indicated by red points.