Understanding biodiversity at the pondscape using environmental DNA: a focus on great crested newts

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, 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

***Corresponding author:**

Email: L.Harper@2015.hull.ac.uk

Word count: 9,563 words

eDNA metabarcoding represents a new tool for community biodiversity assessment in a broad range of aquatic and terrestrial habitats. However, much of the existing literature focuses on methodological development rather than testing of ecological hypotheses. Here, we use presence-absence data generated by eDNA metabarcoding of over 500 UK ponds to examine: 1) species associations between the great crested newt (*Triturus cristatus*) and other vertebrates, 2) determinants of great crested newt occurrence at the pondscape, and 3) determinants of vertebrate species richness at the pondscape. The great crested newt was significantly associated with nine vertebrate species. Occurrence in ponds was broadly reduced by more fish species, but enhanced by more waterfowl and other amphibian species. Abiotic determinants (including pond area, depth, and terrestrial habitat) were identified, which both corroborate and contradict existing literature on great crested newt ecology. Some of these abiotic factors (pond outflow) also determined species richness at the pondscape, but other factors were unique to great crested newt (pond area, depth, and ruderal habitat) or the wider biological community (pond density, macrophyte cover, terrestrial overhang, rough grass habitat, and overall terrestrial habitat quality) respectively. The great crested newt Habitat Suitability Index positively correlated with both eDNA-based great crested newt occupancy and vertebrate species richness. Our study is one of the first to use eDNA metabarcoding to test abiotic and biotic determinants of pond biodiversity. eDNA metabarcoding provided new insights at scales that were previously unattainable using established methods. This tool holds enormous potential for testing ecological hypotheses alongside biodiversity monitoring and pondscape management.

Freshwater ecosystems comprise <1% of the Earth's surface but provide vital

ecosystem services and are hotspots of biodiversity[1–3]. Nonetheless, freshwater organisms are experiencing a greater rate of decline than marine or terrestrial organisms[2,3]. Ponds especially represent critical habitat for biodiversity in a fragmented landscape[4] and support many rare and protected species[5], such as the great crested newt (*Triturus cristatus*) which is protected by UK and European legislation at all life stages[5,6]. Ponds contribute substantially to regional- and landscape-scale aquatic biodiversity[5,7–9] as well as non-aquatic biodiversity within pondscapes, i.e. a pond, its immediate catchment, and the terrestrial matrix of land between ponds[5]. Until recently, pondscapes were poorly understood[5] and neglected in research, scientific monitoring, and policy[4,7,8]. Effective management of pondscapes requires knowledge of abiotic and biotic factors that influence biodiversity, community structure and productivity. Moreover, the biodiversity that ponds support individually and in combination must be examined, but can only be maintained if stressors and threats to these systems are understood[4,5,7,8,10]. Exhaustive sampling of pond biodiversity is impeded by the complexity of these species-rich habitats, and numerous tools required for different taxa with associated bias[11] and cost[12]. However, large-scale community-level monitoring, encompassing alpha (site), beta (between-site) and gamma (landscape) diversity analyses, is necessary to understand biodiversity in changing environments[13].

Analysis of environmental DNA (eDNA, i.e. DNA released by organisms via skin cells, saliva, gametes, urine and faeces into the environment) is providing ecologists with exceptional power to detect single species or describe whole communities[14–18]. The great crested newt was the first and to date only UK protected species to be routinely monitored using eDNA analysed with targeted real-time quantitative PCR (qPCR)[19]. However, entire communities can be monitored using High-Throughput Sequencing, i.e. eDNA metabarcoding[16–18]. This approach has been

used to estimate species richness and assess diversity along environmental gradients[11,20–22], but studies have typically focused on species detection and methodological improvement. eDNA metabarcoding has unprecedented diagnostic power to test classic ecological hypotheses relating to the distribution of biodiversity and its response to environmental pressures. Ponds are ideal model systems for experimental validation and examination of biogeographical patterns as small, abundant ecosystems that span broad ecological gradients[8]; however, few eDNA metabarcoding studies to date have considered ponds[11,12,23–27].

Using ponds, we explore the potential of eDNA metabarcoding for hypothesis testing. We focus on the threatened great crested newt as its ecology is well-understood. Previous work established that both biotic (e.g. food availability, breeding substrate, and predators) and abiotic (e.g. pond depth, area, permanence, and temperature) variables strongly influence great crested newt breeding success[28]. These are encompassed in the Habitat Suitability Index (HSI) used in species surveys[29,30]. The HSI is comprised of 10 suitability indices (factors known to influence great crested newts) which are scored and combined to calculate a decimal score between 0 and 1 representing habitat suitability (where 1 = excellent habitat); although some research suggests HSI may not relate to great crested newt occupancy[31,32]. Fish species may negatively impact great crested newt populations[28,33–40] or effects may be negligible[41]. Larvae tend to swim in open water, increasing susceptibility to fish and waterfowl predation[34,36,38], and adults reportedly avoid ponds containing three-spined stickleback (*Gasterosteus aculeatus*)[42], ninespine stickleback (*Pungitius pungitius*)[40], crucian carp (*Carassius carassius*)[39,40], and common carp (*Carassius carpio*)[40]. Conversely, great crested newts and smooth newts (*Lissotriton vulgaris*) are positively associated due to shared habitat preferences[34,37,38,40]. Great crested newts are more likely in ponds with

better water quality (indicated by diverse macroinvertebrate communities)[29,40], higher nutrient content, and warmer temperature[43]. Water clarity is important for breeding displays, foraging success, and egg survival[34,38]. Higher density of ponds in a pondscape creates more opportunity for great crested newt occupation[33,36,37,40], but presence is negatively correlated with pond surface area[33]. Heavily shaded ponds[44], or those with high macrophyte cover[34,36,38], are less likely to support viable great crested newt populations. Great crested newts are also dependent on terrestrial habitat, preferring open, semi-rural pondscapes[37] containing pasture, extensively grazed and rough grassland, scrub, and coniferous and deciduous woodland[29,38,40,44,45].

The extensive literature on established determinants of the great crested newt provides an excellent opportunity to ground truth ecological patterns revealed by eDNA metabarcoding. We explore this tool's potential for biodiversity assessment at the pondscape using a dataset generated by eDNA metabarcoding of more than 500 ponds with comprehensive environmental metadata. We examined whether eDNA metabarcoding can test ecological hypotheses typically explored by established methods, and whether eDNA and established methods produce congruent results. Specifically, we sought to identify biotic determinants of great crested newt occurrence and species connections to the wider biological community. Using environmental metadata on pond properties and surrounding terrestrial habitat, we aimed to reaffirm abiotic determinants of great crested newts identified using established methods and revisit these important hypotheses at an unprecedented scale. We utilised eDNA metabarcoding for holistic biodiversity monitoring at the pondscape and uncovered abiotic determinants of vertebrate species richness - an impractical task by conventional means. Finally, we evaluated applicability of the great crested newt HSI[29,30] to eDNA-based great crested newt occupancy and vertebrate species richness of ponds.

Methods

Samples. 508 ponds, sampled as part of great crested newt surveys through Natural England's Great Crested Newt Evidence Enhancement Programme, were processed using eDNA metabarcoding alongside 24 ponds privately surveyed by ecological consultants. All water samples were collected using methodology outlined by Biggs *et al.* (2015)[19], detailed in Supplementary Methods. 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) for ethanol precipitation. Water subsamples from the same pond were pooled during DNA extraction to produce one eDNA sample per pond. Targeted qPCR detected great crested newt in 265 (49.81%) ponds. Egg searches performed at 506/508 ponds sampled for Natural England revealed great crested newt in 58 (11.46%) ponds[24].

Environmental metadata on pond characteristics and surrounding terrestrial habitat was collected for 504/508 ponds sampled for Natural England (Supplementary Fig. 1). Pond metadata included: maximum depth; circumference; width; length; area; density; terrestrial overhang; shading; macrophyte cover; HSI score[29]; HSI band (categorical classification of HSI score)[30]; 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 (see Supplementary Table 1 for details of environmental variables).

DNA reference database construction. A custom, phylogenetically curated reference

database of mitochondrial 12S rRNA sequences for UK fish species was previously created at University of Hull for an eDNA metabarcoding study of lake fish communities[46]. Similar reference databases for UK amphibians, reptiles, birds, and mammals were constructed using the ReproPhylo environment[47] in a Jupyter notebook (Jupyter Team 2016). Full details of reference database construction are provided in Harper *et al*. (in press)[24] and Supplementary Methods. Proportion of reference sequences available for species varied within each vertebrate group: amphibians 100.00% (N = 21), reptiles 90.00% (N = 20), mammals 83.93% (N = 112), and birds 55.88% (N = 621). Species without any representation in these databases (i.e. no records for that species or sister species within the same genus) are listed in Supplementary Table 2. The amphibian database was supplemented by Sanger sequences obtained from tissue of great crested newt, smooth newt, Alpine newt (*Mesotriton alpestris*) and common toad (*Bufo bufo*) supplied by DICE, University of Kent, under licence from Natural England, and common frog (*Rana temporaria*) supplied by University of Glasgow (see Supplementary Methods). Databases for each vertebrate group were combined and used for *in silico* validation of primers. 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.1193609.

Primer validation. Published 12S ribosomal RNA (rRNA) primers 12S-V5-F (5'-ACTGGGATTAGATACCCC-3') and 12S-V5-R (5'-TAGAACAGGCTCCTCTAG-3')[48] were validated *in silico* using ecoPCR software[49] against our custom reference database for UK vertebrates. Parameters set allowed a 50-250 bp fragment and maximum of three mismatches between the primer pair and each sequence in the

reference database. Primers were previously validated *in vitro* for UK fish communities by Hänfling *et al.* (2016)[46] and in the present study for six UK amphibian species (Supplementary Fig. 2).

eDNA metabarcoding. Full details of the eDNA metabarcoding workflow performed are provided in Harper *et al*. (in press)[24] and Supplementary Methods. eDNA was amplified with a two-step PCR protocol, using the aforementioned 12S rRNA primers in the first PCR. DNA from a cichlid (*Rhamphochromis esox*) was used for PCR positive controls (six per PCR plate; N = 114), whilst sterile molecular grade water (Fisher Scientific) substituted template DNA for negative controls (six per PCR plate; N = 114). All PCR products were individually purified using E.Z.N.A. Cycle Pure V-Spin Clean-Up Kits (VWR International) following manufacturer's protocol. A second PCR was then used to bind Multiplex Identification (MID) tags to the amplified product. PCR products were individually purified using a magnetic bead clean-up prior to quantification with a Quant-IT™ PicoGreen™ dsDNA Assay. Using concentration values, samples were normalised and pooled to create 4 nM pooled libraries, which were quantified using a Qubit™ dsDNA HS Assay. Sequencing was performed on an Illumina MiSeq using 2 x 300 bp V3 chemistry. Raw sequence reads were taxonomically assigned against our UK vertebrate reference database using a custom pipeline for reproducible analysis of metabarcoding data: metaBEAT (metaBarcoding and Environmental Analysis Tool) v0.8 (https://github.com/HullUni-bioinformatics/metaBEAT). After quality trimming, merging, chimera detection, and clustering, non-redundant sets of query sequences were compared against our custom reference database using BLAST[50]. 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. Sequences that could not be assigned were subjected to a separate BLAST search against the complete NCBI nucleotide (nt) database at 98% identity to determine the source via LCA as described above. To ensure reproducibility, the bioinformatic analysis has been deposited in the GitHub repository.

Data analysis. All downstream 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.

Non-target sequence assignments and original assignments at 98% identity were merged. Any spurious assignments (i.e. non-UK species, invertebrates and bacteria) were removed from the dataset. Assignments to genera or families which contained only a single UK representative were manually assigned to that species. In our dataset, only genus *Strix* was reassigned to tawny owl *Strix aluco*. 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 included in the final database were of species resolution. Misassignments in our dataset were then corrected; again, only one instance was identified. Scottish wildcat *Felis silvestris* was reassigned to domestic cat *Felis catus* on the basis that Scottish wildcat does not occur where ponds were sampled (Kent, Lincolnshire and Cheshire).

To reduce the potential for false positives, we applied species-specific thresholds: a species was only classed as present at a given site if its sequence frequency exceeded a species-specific threshold. Thresholds for each species were defined by analysing sequence data from PCR positive controls (N = 114) and identifying the

maximum sequence frequency for a given species across all PCR positive controls (Supplementary Table 3). For example, the great crested newt species-specific false positive sequence threshold was 0.028% to omit all false detections in the PCR positive controls. After thresholds were applied, the read count data for detected species were converted to a presence-absence matrix for downstream analyses. In the main text, we focus on the results inferred using the species-specific thresholds but all downstream analyses were also performed across a variety of blanket sequence thresholds (0.05 - 30%, see Supplementary Tables 4-9). We tested the influence of fish and waterfowl presence, pond characteristics and surrounding terrestrial habitat on great crested newt occurrence as inferred by eDNA metabarcoding. We were particularly interested in the appropriateness of HSI for eDNA-based great crested newt occupancy. Hypotheses are summarised in Table 1.

All Generalised Linear Mixed Models (GLMMs) were executed using the R package 'lme4' v1.1-12[51]. First, correlations between great crested newt occurrence and number of other vertebrate species were investigated using a binomial GLMM (N = 532). Individual species associations were then investigated using the method of Veech (2013)[52] in the R package 'cooccur' v1.3[53] (N = 532). Identified associations informed candidate biotic variables to be included with abiotic variables (Table S1) in a binomial GLMM of great crested newt occurrence (N = 504). Collinearity and spatial autocorrelation within the dataset were investigated before the most appropriate regression model was determined. Collinearity between explanatory variables was assessed using a Spearman's rank pairwise correlation matrix. After collinear variables were removed, variance inflation factors (VIFs) of remaining variables were calculated using the R package 'car' v2.1-6[54] to identify remnant multicollinearity. Variables corresponding to HSI (HSI score, HSI band) were multicollinear and subsequently

removed prior to model selection (see Supplementary Methods), and HSI score analysed separately in a binomial GLMM.

A large number of explanatory variables remained: max. depth; area; density, overhang; macrophyte cover; permanence; water quality; pond substrate; inflow; outflow; pollution; presence of amphibians, waterfowl and fish; woodland; rough grass; scrub/hedge; ruderals; terrestrial other; and overall terrestrial habitat quality. The relative importance of these for determining great crested newt occurrence was inferred using a classification tree within the R package 'rpart' v4.1-13[55]. A pruning diagram was applied to the data to cross-validate the classification tree and remove unimportant explanatory variables (see Supplementary Methods). Many variables occurred more than once in the classification tree, indicative of weak non-linear relationships with the response variable. Generalised Additive Models (GAMs) were performed to deal with non-linearity but several explanatory variables were in fact linear (estimated one degree of freedom for smoother). A parametric, binomial Generalised Linear Model (GLM) was applied and the potential for spatial autocorrelation assessed using spline correlograms of the data using R package 'ncf' v1.1-7[56]. A binomial GLMM was employed to account for dependencies within sites, handled with the introduction of random effects[57–59]. Each eDNA sample represented a different pond and thus sample was treated as a random effect. The mixed model successfully accounted for spatial autocorrelation within sites when a spline correlogram of the Pearson residuals was examined.

After identification of a suitable set of explanatory variables and modelling framework, the variables which are the most important determinants of great crested newt occurrence and a suitable, parsimonious approximating model to make predictions were determined. An information-theoretic approach using Akaike's Information

Criteria (AIC) to evaluate model fit was employed[60]. A binomial distribution was specified as the response variable was presence-absence data. After following a similar workflow for identification of a suitable set of explanatory variables and modelling framework (see Supplementary Methods), a set of variables that best explain vertebrate species richness were constructed. A Poisson distribution was specified for all species richness models as the response variable was integer count data. Model fit was evaluated as above using AIC.

All binomial and Poisson models considered were nested and so the best models of great crested newt occurrence and vertebrate species richness respectively were chosen using stepwise backward deletion of terms based on Likelihood Ratio Tests (LRTs). The final binomial and Poisson models were tested for overdispersion using the R package 'RVAideMemoire' v 0.9-69[61] and custom functions to test overdispersion of the Pearson residuals. Model fit was assessed using the Hosmer and Lemeshow Goodness of Fit Test[62] within the R package 'ResourceSelection' v0.3-2[63], quantile-quantile plots and partial residual plots[59,64]. Model predictions were obtained using the predictSE() function in the 'AICcmodavg' package v2.1-1[65] and upper and lower 95% CIs were calculated from the standard error of the predictions. All values were bound in a new data frame and model results plotted for evaluation using the R package 'ggplot2' v 2.2.1[66].

Data availability. 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.1193609).

Results and discussion

Across two sequencing runs, 532 eDNA samples and 228 PCR controls were processed. 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. In the final dataset (thresholds applied and assignments corrected), a total of 60 vertebrate species were detected by eDNA metabarcoding across the 532 ponds surveyed (Supplementary Table 10). These consisted of six amphibian species, 14 fish species, 18 bird species, and 22 mammal species (Supplementary Fig. 3). Amphibian species detection ranged from 1 - 152 ponds (median 81 ponds) whilst fish species detection ranged from 1 - 72 ponds (median 15 ponds). Bird species detection ranged between 1 and 215 ponds (median 3 ponds) whereas mammal species detection ranged between 1 and 179 ponds (median 9 ponds). The most common species detected across all vertebrate groups were common moorhen (*Gallinula chloropus*, N = 215), cow (*Bos taurus*, N = 179), smooth newt (N = 152), great crested newt (N = 149), pig (*Sus scrofa*, N = 140), and common frog (N = 120). All detected species and their frequency of detection are listed in Supplementary Table 10.

We discuss great crested newt occupancy in the context of broad trends found across vertebrate groups (GLMM: overdispersion χ^2_{525} = 517.636, P = 0.582; fit χ^2_8 = 22.524, P = 0.004, R^2 = 9.43%) and individual species associations (Table 1).

Significant species associations with the great crested newt revealed by the co-occurrence analysis were carried forward as candidate variables for analysis of biotic and abiotic determinants (GLMM: overdispersion $\chi^2_{490} = 413.394$, $P = 0.995$; fit $\chi^2_8 = 11.794$, $P = 0.161$, $R^2 = 38.58\%$). Associations with support from both analyses are summarised in Table 1. Great crested newt occupancy was best explained by smooth newt occurrence (+), common toad occurrence (-), three-spined stickleback occurrence (-), grey squirrel occurrence (-), vertebrate species richness (+), pond outflow (-), ruderal habitat (-), other good terrestrial habitat (-), pond area (-), and max. depth (+).

Biotic determinants of great crested newt occurrence. We found a positive correlation between great crested newt occurrence and increasing number of other amphibian species (Table 1, Fig. 1a). Smooth newts were commonly detected in ponds with great crested newts but palmate newts (*Lissotriton helveticus*) were not. Similarly, common toad and common frog records were less frequent in ponds containing great crested newts, and there was only one record of marsh frogs (*Pelophylax ridibundus*) in a great crested newt pond (Supplementary Fig. 3a). Of these observations, a positive association between the great crested newt and smooth newt (Table 1, Figs. 2, 3a) and a negative association between the great crested newt and common toad were significant (Table 1, Figs. 2, 3b). Great crested newts and smooth newts share similar terrestrial and aquatic habitat requirements resulting in selection of the same ponds for breeding[34,37,40], with more than 60% overlap in ponds reported[38]. Notably, research suggests smooth newts are more versatile and capable of inhabiting a broader range of habitat[34,38] whereas great crested newts may be associated with larger, deeper ponds with an abundance of macrophytes and absence of fish located in open, semi-rural landscapes[37]. Conversely, the negative association observed between the great crested

newt and common toad may be attributable to toads inhabiting fish-containing ponds[67] or great crested newt predation on toad eggs and larvae[28].

Great crested newt occurrence was reduced in ponds containing a greater number of fish species (Table 1), and newts were absent from ponds containing more than four fish species (Fig. 1b). Nonetheless, all detected fish species were recorded in great crested newt ponds to some extent, except Eurasian ruffe (*Gymnocephalus cernua*) (Fig. S3b). It is important to note that some fish species detections may result from eDNA transport into ponds via inflows from larger stream or river catchments, when these species do not actually inhabit ponds. We discuss only associations with fish species that are established inhabitants of ponds. The great crested newt had significant negative associations with ninespine stickleback (Table 1, Fig. 2) and three-spined stickleback (Table 1, Figs. 2, 3c), the latter of which was shared by the smooth newt (Fig. 2). A non-significant negative co-occurrence was also observed between great crested newts and common carp (Table 1). Common carp are ecosystem engineers: their benthic foraging activity increases water turbidity and reduces invertebrate density and macrophyte cover, affecting species that depend on these groups[68,69]. Introduced fish species exerted a negative effect on site occupation of both newt species in Belgium[37] and both species only colonised a site in England once three-spined stickleback were removed[42]. Smooth newts are known to avoid fish occupied sites, including ponds and wetlands[70,71], and negative effects of fish species on great crested newt populations have been frequently reported[28,33–36,38,39]. Conversely, other research suggests no or minimal negative interaction between fish and great crested newts[41,72]. Fish species characteristic of ponds, such as crucian carp, are unlikely to be damaging predators to amphibian populations[72,73]. Indeed, great crested newt detection was equal in ponds containing or absent of crucian carp. However, consumption of

macroinvertebrates by fish can alter habitat suitability for great crested newts[35] as many fish species share the same trophic status as newt species[70]. Fish also tend to be associated with algal ponds where macrophyte diversity is impaired. Reduced macrophyte availability imposes restrictions on egg-laying in great crested newts and restricts the ecological niches that invertebrate prey may inhabit[38].

Unexpectedly, great crested newt occurrence was positively associated with increasing number of waterfowl species (Table 1, Fig. 1c), despite absence of great crested newts in ponds with certain waterfowl species (Supplementary Fig. 3c). Furthermore, the great crested newt had significant positive associations with the common coot (*Fulica atra*) and common moorhen (Table 1, Fig. 2), and a non-significant negative association with the green-winged teal (Table 1). Great crested newts are typically found in ponds with high macrophyte diversity as macrophyte species dictate reproductive success and invertebrate prey availability[72,74]. Common moorhen and common coot share macrophytes and macroinvertebrates as resources but feed on both directly[75–77] thus competition between great crested newts and omnivorous waterfowl may be reduced or indirect. Great crested newt breeding in April to June[28] may be impacted by coots pulling up submerged vegetation, damaging vegetation banks[78]. However, coot diet tends to be macrophyte-dominated in late summer and autumn[77]. Both coot and moorhen also crop emergent macrophytes in their search for invertebrate prey[75,76], but in doing so they may expose prey items to great crested newts and confer indirect benefits.

The most common terrestrial detections in this study were domesticated or introduced pest species, such as grey squirrel (*Sciurus carolinensis*) and muntjac deer (*Muntiacus reevesi*)[69,79] (Supplementary Table 10). Nonetheless, we identified wild species which emphasise the importance of ponds as stepping stones for both semi-

aquatic and terrestrial taxa[7,8], through provision of drinking, foraging, dispersal, and reproductive opportunities[10,25,80]. The most frequent terrestrial bird detections included buzzard (*Buteo buteo*), Eurasian jay (*Garrulus glandarius*), dunnock (*Prunella modularis*), and starling (*Sturnus vulgaris*) (Supplementary Table 10), which utilise different habitats. We detected several mammal species with Biodiversity Actions Plans and/or of conservation concern, including otter (*Lutra lutra*), water vole (*Arvicola amphibius*), European polecat (*Mustela putorius*), brown hare (*Lepus europaeus*) and water shrew (*Neomys fodiens*)[79]. Notably, some mammals were only identified in one pond (Supplementary Table 10) and American mink (*Neovison vison*) was absent despite widespread UK distribution[79].

Records of great crested newts in relation to terrestrial species, and any significant associations identified below, are unlikely to reflect direct species interactions. Rather, these records and associations are a probable outcome of land-use and indirect interaction. No significant relationships between the numbers of terrestrial bird or mammal species and great crested newt occurrence were found (Table 1, Figs. 1d, e), but great crested newts were entirely absent from ponds where certain terrestrial species were present e.g. great spotted woodpecker *(Dendrocopos major*), tawny owl, badger (*Meles meles*), and red deer (*Cervus elaphus*) (Supplementary Figs. 3c, d).The great crested newt had a significant positive association with pig (Table 1, Fig. 2), but significant negative associations with the grey squirrel (*Sciurus carolinensis*) (Table 1, Figs. 2, 3d) and common pheasant (*Phasianus colchicus*) (Table 1, Fig. 2). A non-significant negative co-occurrence between the great crested newt and badger was also identified (Table 1). Excluding breeding, adult great crested newts live outside ponds in terrestrial habitat for foraging, shelter, and hibernation[35,43]. Juveniles also spend two to three years on land after emerging from ponds[70]. During time spent outside of ponds,

great crested newts may suffer dessication, or predation[28,45,70] from terrestrial species. There have been anecdotal records of pheasant predation on herpetofauna, including great crested newt[81], but existing literature on amphibian and terrestrial species interactions is sparse with even less study of great crested newt interactions.

Critically, data points in all analyses of biotic determinants were not evenly distributed across different species, or number of species within each vertebrate group. Ponds containing a higher number of vertebrate species were much fewer than ponds containing a lower number of vertebrate species (Fig. 1). Similarly, some species were detected more frequently in ponds than others (Supplementary Table 10 and Fig. 3). This uneven distribution is likely a natural outcome of species accumulation, but may reduce capability of models to make accurate predictions.

Abiotic determinants of great crested newt occurrence. The probability of great crested newt occurrence increased with greater pond depth but decreased in ponds with larger area, outflow, without ruderal habitat, and with some other good terrestrial habitat (Table 1, Figs. 3e-g, i-j). Previous work has shown great crested newts utilise small and large ponds[34,38], although very small ponds (less than 124 m^2) were incapable of supporting all life stages and larger ponds had greater occurrence of fish[38]. Large ponds may also be more susceptible to eutrophication due to agricultural or polluted run-off[38]. Yet, some studies found no effect of pond area on great crested newt occurrence[35,41,45], and pond area has been deemed a poor predictor of great crested newt reproductive success[44]. In contrast, past research showed a positive influence of pond depth on great crested newt occupancy[35]. Conditions in shallow ponds may be too unpredictable for great crested newt occupation, as they are susceptible to drying out or freezing and may contain less prey. However, pond depth and surrounding terrestrial

habitat may be linked as detrimental effects of shallow water are more typically observed in open farmland[33,35,71]. Temporary water bodies can be occupied provided fish are absent[71] and drying may reduce predators such as fish and dragonfly larvae[28]. Unlike pond area and depth, there is little research on outflow to support an effect on great crested newt occurrence. Pond inflow is known to affect biodiversity due to polluted agricultural run-off and connections to streams and rivers containing large, predatory fish. We suggest outflow (facilitated by drains, pipes or streams) may stabilise maximum water level[82] and minimise fluctuations in pond depth, affecting subsequent colonisation and structure of biological communities.

Our results support previous work demonstrating that good terrestrial habitat is key to great crested newt success and serves multiple purposes, including daytime and long-term shelter from extreme conditions in refugia, as well as foraging and dispersal opportunities[28]. Previous research determined great crested newt occupancy and breeding success was sub-optimal in coniferous forest yet enhanced in deciduous or herb-rich forest and pasture[44,45]. Similarly, extensively grazed grassland and deadwood positively influenced great crested newt presence whilst intensively grazed grasslands were unoccupied[38]. Lower great crested newt abundance has been observed in cultivated habitats[33], and modern forestry and increasing land use were deemed the biggest great crested newt decline factors using a spatially explicit population model[44]. Conversely, others have found minimal effect of landscape context (excluding urban areas) on great crested newts[36], suggesting terrestrial habitat may not restrict species distribution; although, habitat degradation may increase isolation of ponds. Our results indicate terrestrial habitat does influence great crested newt occupancy but without quantitative data, these discrete effects cannot be teased apart. Data on type, density, and great crested newt utilisation of terrestrial habitat, as well as distance of ponds to

terrestrial habitat, are necessary to fully understand great crested newt occupancy and interactions with terrestrial species. However, this is a phenomenal task for large numbers of ponds across a vast landscape[35].

HSI in relation to eDNA-based great crested newt occupancy. In a separate analysis (GLMM: overdispersion $\chi^2_{501} = 506.763$, $P = 0.4198$; fit $\chi^2_8 = 8.118$, $P = 0.422$, $R^2 = 4.99\%$), eDNA-based great crested newt occurrence positively correlated with HSI score (Table 1), where probability of great crested newt occupancy was greater in ponds with higher HSI score (Fig. 4). It has been suggested that HSI may be inappropriate for predicting great crested newt occupancy or survival probabilities[32], but our finding indicates HSI can be used to predict great crested newt occupancy at the pondscape. HSI may help establish protection of ponds and the biodiversity they host by identifying those which may be occupied by great crested newt. Optimal habitat can also be identified for creation of new ponds or restoration of old ponds to encourage new populations of this threatened amphibian species.

Nevertheless, issues remain with the HSI. Great crested newt occurrence may indicate good quality habitat but may not reflect successful breeding and population viability, albeit one study found ponds with higher HSI did have higher reproduction probability[32]. Other issues include the use of qualitative data for score calculation, and subjective, scorer-dependent estimation of indices[29]. For future application of HSI in great crested newt eDNA survey, we recommend metabarcoding for quantification of some indices which are qualitatively assessed (e.g. water quality via macroinvertebrate diversity, fish and waterfowl presence) alongside detection of great crested newts. Provided rigorous spatial and temporal sampling are undertaken, eDNA metabarcoding can also generate site occupancy data to estimate relative abundance of species[12,46].

However, only conventional surveys can provide data on true great crested newt abundance to enable effective mitigation (e.g. translocation), understand population dynamics, and generate survival and reproduction probabilities.

Abiotic and biotic determinants of vertebrate species richness. In another analysis (GLMM: overdispersion χ^2_{494} = 431.959, P = 0.979; fit χ^2_8 = -42.708, P = 1.000, R^2 = 8.94%), species richness was greater in ponds with outflow (0.214 ± 0.063, χ^2_1 = 11.220, P = 0.0008, Fig. 5a), but reduced in those with some rough grass habitat (-0.297 ± 0.074) compared to ponds with no (-0.1402 ± 0.0795) or important rough grass habitat (χ^2_2 = 16.715, P = 0.0002, Fig. 5b). Overall quality of terrestrial habitat was also influential (χ^2_2 = 8.244, P = 0.016, Fig. 5c) where species richness was higher in ponds that were in areas considered to be poor (0.115 ± 0.089) or moderate (0.216 ± 0.078) habitat for great crested newts. Species richness was reduced as percentage of terrestrial overhang (-0.0026 ± 0.0008, χ^2_1 = 9.575, P = 0.002, Fig. 5d) and percentage of macrophyte cover increased (-0.002 ± 0.001, χ^2_1 = 4.117, P = 0.043, Fig. 5e) but improved with pond density (0.006 ± 0.003, χ^2_1 = 4.564, P = 0.033, Fig. 5f). Many studies have focused on species richness of aquatic invertebrates as a range of invertebrate groups can be surveyed simultaneously using conventional tools. Until application of eDNA metabarcoding, this was not possible for aquatic and non-aquatic vertebrates. Instead indicator groups, such as amphibians, were chosen as representatives of pond biodiversity - although amphibians may in fact be poor surrogates of macroinvertebrate and macrophyte diversity[83,84]. Consequently, the literature on vertebrate species richness in aquatic ecosystems is sparse and we may only compare our results to studies which have investigated species richness of different vertebrate assemblages or species guilds, primarily amphibians and waterfowl.

Plentiful rough grass habitat can create more ecological niches and foraging opportunity for a variety of vertebrates, but quantitative data on type and abundance of terrestrial habitat surrounding ponds would be required to understand which species prefer open or covered habitat. Pond outflow and inflow have received little investigation in studies of freshwater biodiversity. Outflow may release harmful pollutants and pathogens[85] that would otherwise accumulate and be retained in a closed pond system. Outflow may benefit vertebrate biodiversity at risk of human health in urban areas[85], but pollution was not identified as a candidate for model selection. Shade has been identified as a principal driver of macroinvertebrate and macrophyte diversity in freshwater ponds, negatively correlating with macrophyte cover, and can create anoxic conditions in water bodies thereby decreasing productivity[9]. This can have knock-on effects for consumers at higher trophic levels. For example, amphibians have been observed to avoid ponds that are densely vegetated[86]. Yet, canopy and macrophyte cover were also identified as positive drivers of amphibian species richness[86,87] and abundance[88]. Our own results indicate highly shaded ponds are inconducive to high vertebrate species richness but high densities of ponds support higher species richness, providing further evidence of the importance of ponds for aquatic and non-aquatic taxa[7,8].

In a separate analysis (GLMM: overdispersion $\chi^2_{501} = 389.744$, $P = 0.999$; fit $\chi^2_8 = -145.12$, $P = 1.000$, $R^2 = 1.10\%$), vertebrate species richness positively correlated with HSI score (0.459 ± 0.002, $\chi^2_1 = 5.034$, $P = 0.025$, $R^2 = 1.10\%$), where species richness was improved in ponds with higher HSI score (Fig. 5g). HSI score also positively correlated with probability of great crested newt occurrence (Fig. 4), and a positive association was identified between vertebrate species richness and great crested newt occurrence ($P < 0.0001$, Fig. 3b). Our results suggest some indices which

comprise the great crested newt HSI also represent key habitat criteria for broader biodiversity, for example, outflow and terrestrial habitat. However, several indices which affect great crested newts were not identified as determinants of vertebrate species richness. Nonetheless, it may be possible to adapt HSI to more accurately represent and predict vertebrate species richness in order to identify areas for pond creation and management to enhance aquatic and non-aquatic biodiversity.

Implications for biodiversity assessment at the pondscape. Many species associations were identified using eDNA metabarcoding (Fig. 2). However, there is no literature available to confirm the nature, or even existence, of the majority of these relationships. Lack of appropriate survey methods has caused freshwater research to focus on single species or guilds and assemblages when studying predictors of species diversity, richness, and abundance, or investigating impact of environmental change and gradients. New methods are required for holistic biodiversity assessment in response to ecosystem drivers and stressors. We have demonstrated how eDNA metabarcoding can be used for landscape-scale biodiversity monitoring and ecological study. Our results provide new insights and unparalleled biological understanding of aquatic and non-aquatic biodiversity at the UK pondscape. Continued use of eDNA metabarcoding could enhance our understanding of freshwater networks to enable more effective protection and management for both aquatic and non-aquatic biodiversity. Huge quantities of data can be generated to reduce the noise typically observed in ecological datasets and at comparable cost to single-species eDNA monitoring[24]. We investigated associations between aquatic and non-aquatic vertebrates and combined metabarcoding with environmental metadata to revisit important ecological hypotheses at an unprecedented scale, identifying determinants of great crested newts and broader

biodiversity. Our findings indicate preferred habitat of a threatened amphibian and will guide management in the face of increasing land-use and habitat fragmentation - a poignant issue as protective legislation for the great crested newt in the UK is under review. Whilst conservation of threatened biodiversity and their habitat should be a priority, the bigger picture should not be ignored. eDNA metabarcoding can create both fine and broad-scale species inventories and allow researchers to examine the response of entire communities' to environmental change, thereby allowing prioritisation of regional- and landscape-scale conservation effort. eDNA metabarcoding holds great promise for improved biodiversity monitoring and we are only beginning to realise and explore these opportunities.

References

1. Strayer, D. L. & Dudgeon, D. Freshwater biodiversity conservation: recent progress and future challenges. *J. North Am. Benthol. Soc.* **29,** 344–358 (2010).
2. Collen, B. *et al.* Global patterns of freshwater species diversity, threat and endemism. *Glob. Ecol. Biogeogr.* **23,** 40–51 (2014).
3. Dudgeon, D. *et al.* Freshwater biodiversity: importance, threats, status and conservation challenges. *Biol. Rev. Camb. Philos. Soc.* **81,** 163–182 (2006).
4. Céréghino, R., Biggs, J., Oertli, B. & Declerck, S. The ecology of European ponds: defining the characteristics of a neglected freshwater habitat. *Hydrobiologia* **597,** 1–6 (2008).
5. Wood, P. J., Greenwood, M. T. & Agnew, M. D. Pond biodiversity and habitat loss in the UK. *Area* **35,** 206–216 (2003).
6. Rees, H. C. *et al.* The application of eDNA for monitoring of the Great Crested Newt in the UK. *Ecol. Evol.* **4,** 4023–4032 (2014).
7. Oertli, B., Céréghino, R., Hull, A. & Miracle, R. Pond conservation: from science to practice. *Hydrobiologia* **634,** 1–9 (2009).
8. De Meester, L. *et al.* Ponds and pools as model systems in conservation biology, ecology and evolutionary biology. *Aquat. Conserv.* **15,** 715–725 (2005).
9. Sayer, C. *et al.* The role of pond management for biodiversity conservation in an agricultural landscape. *Aquat. Conserv.* **22,** 626–638 (2012).
10. Biggs, J., von Fumetti, S. & Kelly-Quinn, M. The importance of small waterbodies for biodiversity and ecosystem services: implications for policy makers. *Hydrobiologia* **793,** 3–39 (2016).

11. Evans, D. N. T. *et al.* Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. *Can. J. Fish. Aquat. Sci.* **74,** 1362-1374 (2017).
12. Valentini, A. *et al.* Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Mol. Ecol.* **25,** 929–942 (2016).
13. Hajibabaei, M., Baird, D. J., Fahner, N. A., Beiko, R. & Golding, G. B. A new way to contemplate Darwin's tangled bank: how DNA barcodes are reconnecting biodiversity science and biomonitoring. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **371,** (2016).
14. Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R. M. & and Gough, K. C. REVIEW: The detection of aquatic animal species using environmental DNA – a review of eDNA as a survey tool in ecology. *J. Appl. Ecol.* **52,** 827-831 (2014).
15. Thomsen, P. F. & Willerslev, E. Environmental DNA – an emerging tool in conservation for monitoring past and present biodiversity. *Biol. Conserv.* **183,** 4-18 (2015).
16. Lawson Handley, L. How will the 'molecular revolution' contribute to biological recording? *Biol. J. Linnaean. Soc.* **115,** 750–766 (2015).
17. Deiner, K. *et al.* Environmental DNA metabarcoding: transforming how we survey animal and plant communities. *Mol. Ecol.* **26,** 5872-5895 (2017).
18. Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* **21,** 2045–2050 (2012).
19. Biggs, J. *et al.* Using eDNA to develop a national citizen science-based monitoring programme for the great crested newt (*Triturus cristatus*). *Biol. Conserv.* **183,** 19–28 (2015).
20. Kelly, R. P. *et al.* Genetic signatures of ecological diversity along an urbanization gradient. *PeerJ* **4,** e2444 (2016).
21. Olds, B. P. *et al.* Estimating species richness using environmental DNA. *Ecol. Evol.* **6,** 4214–4226 (2016).
22. Andruszkiewicz, E. A. *et al.* Biomonitoring of marine vertebrates in Monterey Bay using eDNA metabarcoding. *PLoS One* **12,** e0176343 (2017).
23. Thomsen, P. F. *et al.* Monitoring endangered freshwater biodiversity using environmental DNA. *Mol. Ecol.* **21,** 2565–2573 (2012).
24. Harper, L. R. *et al.* Needle in a haystack? A comparison of eDNA metabarcoding and targeted qPCR for detection of the great crested newt (*Triturus cristatus*). *Ecol. Evol.* (in press). Pre-print at https://www.biorxiv.org/content/early/2018/03/05/215897.
25. Klymus, K. E., Richter, C. A., Thompson, N. & Hinck, J. E. Metabarcoding of Environmental DNA Samples to Explore the Use of Uranium Mine Containment Ponds as a Water Source for Wildlife. *Diversity* **9,** 54 (2017).
26. Bálint, M. *et al.* Twenty-five species of frogs in a liter of water: eDNA survey for exploring tropical frog diversity. Pre-print at https://www.biorxiv.org/content/early/2017/08/15/176065 (2017).

27. Ushio, M. *et al.* Environmental DNA enables detection of terrestrial mammals from forest pond water. *Mol. Ecol. Resour.* **17,** e63-e65 (2017).
28. Langton, T., Beckett, C. & Foster, J. *Great crested newt conservation handbook*. (Froglife, Halesworth, 2001).
29. Oldham, R. S., Keeble, J., Swan, M. J. S. & and Jeffcote, M. Evaluating the suitability of habitat for the great crested newt (*Triturus cristatus*). *Herpetol. J.* **10,** 143-155 (2000).
30. ARG-UK. *ARG UK Advice Note 5: Great Crested Newt Habitat Suitability Index*. (Amphibian and Reptile Groups of the United Kingdom, 2010).
31. Sewell, D., Beebee, T. J. C. & Griffiths, R. A. Optimising biodiversity assessments by volunteers: The application of occupancy modelling to large-scale amphibian surveys. *Biol. Conserv.* **143,** 2102–2110 (2010).
32. Unglaub, B., Steinfartz, S., Drechsler, A. & Schmidt, B. R. Linking habitat suitability to demography in a pond-breeding amphibian. *Front. Zool.* **12,** 9 (2015).
33. Joly, P., Miaud, C., Lehmann, A. & Grolet, O. Habitat Matrix Effects on Pond Occupancy in Newts. *Conserv. Biol.* **15,** 239–248 (2001).
34. Skei, J. K., Dolmen, D., Rønning, L. & Ringsby, T. H. 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. *Amphib-Reptil.* **27,** 309–324 (2006).
35. Denoël, M. & Ficetola, G. F. Conservation of newt guilds in an agricultural landscape of Belgium: the importance of aquatic and terrestrial habitats. *Aquat. Conserv.* **18,** 714–728 (2008).
36. Hartel, T., Nemes, S. & Oellerer, K. Using connectivity metrics and niche modelling to explore the occurrence of the northern crested newt *Triturus cristatus* (Amphibia, Caudata) in a traditionally managed landscape. *Environ. Conserv.* **37,** 195-200 (2010).
37. Denoël, M., Perez, A., Cornet, Y. & Ficetola, G. F. Similar local and landscape processes affect both a common and a rare newt species. *PLoS One* **8,** e62727 (2013).
38. Rannap, R. & Briggs, L. THE CHARACTERISTICS OF GREAT CRESTED NEWT TRITURUS CRISTATUS'BREEDING PONDS. (Project Report: "Protection of the Great Crested Newt in the Eastern Baltic region", LIFE2004NAT/EE/000070, Action A1, 2006).
39. Rannap, R., Lõhmus, A. & Briggs, L. Restoring ponds for amphibians: a success story. *Hydrobiologia* **634,** 87–95 (2009).
40. Rannap, R., Lõhmus, A. & Briggs, L. Niche position, but not niche breadth, differs in two coexisting amphibians having contrasting trends in Europe. *Divers. Distributions* **15,** 692–700 (2009).
41. Maletzky, A., Kyek, M. & Goldschmid, A. Monitoring status, habitat features and amphibian species richness of crested newt (*Triturus cristatus* superspecies) ponds at the edge of the species range (Salzburg, Austria). *Ann. Limnol. – Int. J. Lim.* **43,** 107-115 (2007).

42. McLee, A. G. & Scaife, R. W. 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*). *Herpetol. Bull.* **42,** 6-9 (1992).
43. Gustafson, D. H., Andersen, A. S. L., Mikusiński, G. & Malmgren, J. C. Pond Quality Determinants of Occurrence Patterns of Great Crested Newts (*Triturus cristatus*). *J. Herpetol.* **43,** 300–310 (2009).
44. Vuorio, V., Heikkinen, R. K. & Tikkanen, O.-P. Breeding Success of the Threatened Great Crested Newt in Boreal Forest Ponds. *Ann. Zool. Fennici* **50,** 158–169 (2013).
45. Gustafson, D. H., Malmgren, J. C. & Mikusiński, G. Terrestrial Habitat Predicts use of Aquatic Habitat for Breeding Purposes — A Study on the Great Crested Newt (*Triturus cristatus*). *Ann. Zool. Fennici* **48,** 295–307 (2011).
46. Hänfling, B. *et al.* Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. *Mol. Ecol.* **25**, 3101-3119 (2016).
47. Szitenberg, A., John, M., Blaxter, M. L. & Lunt, D. H. ReproPhylo: An Environment for Reproducible Phylogenomics. *PLoS Comput. Biol.* **11,** e1004447 (2015).
48. Riaz, T. *et al.* ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Res.* **39,** e145 (2011).
49. Ficetola, G. F. *et al.* An *In silico* approach for the evaluation of DNA barcodes. *BMC Genomics* **11,** 434 (2010).
50. Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **7,** 203–214 (2000).
51. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *J. of Stat. Softw.* **67,** 1–48 (2015).
52. Veech, J. A. A probabilistic model for analysing species co-occurrence. *Glob. Ecol. Biogeogr.* **22,** 252-260 (2013).
53. Griffith, D., Veech, J. & Marsh, C. cooccur: Probabilistic Species Co-Occurrence Analysis in R. *J. Stat. Softw.* **69,** 1–17 (2016).
54. Fox, J. & Weisberg, S. *An R companion to applied regression, Second Edition*. (Sage Publications, Thousand Oaks, CA, USA, 2011).
55. Therneau, T., Atkinson, B. & Ripley, B. rpart: Recursive Partitioning and Regression Trees. R package version 4.1-13 (2018).
56. Bjørnstad, O. N. ncf: spatial nonparametric covariance functions. R package version 1.1-7 (2016).
57. Pinheiro, J. C. & Bates, D. M. *Mixed-effects in S and S-Plus*. (Springer, New York, USA, 2000).
58. McCulloch, C. E. & Searle, S. R. *Generalized, Linear, and Mixed Models* (John Wiley & Sons, Inc., Hoboken, NJ, USA, 2001).
59. Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A. & Smith, G. M. *Mixed effects models and extensions in ecology with R* (Springer, New York, USA, 2009).

60. Akaike, H. Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika* **60,** 255-265 (1973).
61. Hervé, M. RVAideMemoire: Diverse basic statistical and graphical functions. R package version 0.9-69. (2018).
62. Hosmer, D. W. & Lemeshow, S. *Applied Logistic Regression* (John Wiley & Sons, Inc., Hoboken, NJ, USA, 2000).
63. Lele, S. R., Keim, J. L. & Solymos, P. ResourceSelection: Resource Selection (Probability) Functions for Use-Availability Data. R package version 0.3-2 (2017).
64. Landwehr, J. M., Pregibon, D. & Shoemaker, A. C. Graphical Methods for Assessing Logistic Regression Models. *J. Am. Stat. Assoc.* **79,** 61–71 (1984).
65. Mazerolle, M. J. AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c). R package version 2.1-1 (2017).
66. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer, New York, USA, 2009).
67. Manenti, R. & Pennati, R. Environmental factors associated with amphibian breeding in streams and springs: effects of habitat and fish occurrence. *Amphib-Reptil.* **37,** 237–242 (2016).
68. Maceda-Veiga, A., López, R. & Green, A. J. Dramatic impact of alien carp *Cyprinus carpio* on globally threatened diving ducks and other waterbirds in Mediterranean shallow lakes. *Biol. Conserv.* **212,** 74–85 (2017).
69. Manchester, S. J. & Bullock, J. M. The impacts of non-native species on UK biodiversity and the effectiveness of control. *J. Appl. Ecol.* **37,** 845–864 (2000).
70. Aronsson, S. & Stenson, J. A. E. Newt-fish interactions in a small forest lake. *Amphib-Reptil.* **16,** 177–184 (1995).
71. Ficetola, G. F. & De Bernardi, F. Amphibians in a human-dominated landscape: the community structure is related to habitat features and isolation. *Biol. Conserv.* **119,** 219–230 (2004).
72. Chan, K. Can great crested newts (*Triturus cristatus*) coexist with fish? (University College London, 2011).
73. Hecnar, S. & M'Closkey, R. Amphibian species richness and distribution in relation to pond water chemistry in south-western Ontario, Canada. *Freshw. Biol.* (1996).
74. Gustafson, D. H., Pettersson, C. J. & Malmgren, J. C. Great crested newts (*Triturus cristatus*) as indicators of aquatic plant diversity. *Herpetol. J.* **16,** 347–352 (2006).
75. Paillisson, J.-M. & Marion, L. Interaction between coot (*Fulica atra*) and waterlily (*Nymphaea alba*) in a lake: the indirect impact of foraging. *Aquat. Bot.* **71,** 209–216 (2001).
76. Wallau, G. L. *et al.* Behaviour of the Common Moorhen in Rio Grande do Sul, Brazil. *Acta Ethol.* **13,** 127–135 (2010).
77. Perrow, M. R. *et al.* Interactions between coot (*Fulica atra*) and submerged macrophytes: the role of birds in the restoration process. *Hydrobiologia* **342/343,** 241–255 (1997).

78. Lauridsen, T. L., Jeppesen, E. & Andersen, F. Ø. Colonization of submerged macrophytes in shallow fish manipulated Lake Væng: impact of sediment composition and waterfowl grazing. *Aquat. Bot.* **46,** 1–15 (1993).
79. Battersby, J. & Partnership, T. M. *UK Mammals: Species Status and Population Trends. First Report by the Tracking Mammals Partnership* (JNCC/Tracking Mammals Partnership, Peterborough, 2005).
80. Almeida, D., Rodolfo, N., Sayer, C. D. & Copp, G. H. Seasonal use of ponds as foraging habitat by Eurasian otter with description of an alternative handling technique for common toad predation. *Folia Zool. Brno.* **62,** 214–221 (2013).
81. Rice, C. N. Abundance, impacts and resident perceptions of non-native common pheasants (*Phasianus colchicus*) in Jersey, UK Channel Islands. (University of Kent, 2016).
82. Freshwater Habitats Trust. *Pond Habitat Survey: Survey Manual* (Freshwater Habitats Trust, Oxford, 2015).
83. Ilg, C. & Oertli, B. Effectiveness of amphibians as biodiversity surrogates in pond conservation. *Conserv. Biol.* **31,** 437-445 (2016).
84. Sewell, D. & Griffiths, R. Can a Single Amphibian Species Be a Good Biodiversity Indicator? *Diversity* **1,** 102–117 (2009).
85. Beutel, M. W. & Larson, L. Pathogen removal from urban pond outflow using rock biofilters. *Ecol. Eng.* **78,** 72–78 (2015).
86. Piha, H., Luoto, M. & Merilä, J. Amphibian occurrence is influenced by current and historic landscape characteristics. *Ecol. Appl.* **17,** 2298–2309 (2007).
87. Semlitsch, R. D., Peterman, W. E., Anderson, T. L., Drake, D. L. & Ousterhout, B. H. Intermediate pond sizes contain the highest density, richness, and diversity of pond-breeding amphibians. *PLoS One* **10,** e0123055 (2015).
88. Landi, M., Piazzini, S. & Saveri, C. The response of amphibian communities to fish and habitat features in Mediterranean permanent ponds. *Biologia* **69,** (2014).

Acknowledgements

This work was funded by University of Hull. We would like to thank Jennifer Hodgetts (Fera) for assisting with sample collection, and Jianlong Li (University of Hull) for primer design and advice on laboratory protocols. Furthermore, Barbara Mabel (University of Glasgow), Andrew Buxton and Richard Griffiths (DICE, University of Kent) provided tissue samples for primer validation and Sanger sequencing to supplement the reference database.

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.

Competing interests

The authors declare no competing financial interests.

Materials and correspondence

All requests should be addressed to L.R.H., B.H. or L.L.H.

Table 1 | Summary of established and newly identified abiotic and biotic determinants of great crested newt occupancy. Reported effects on great crested newt occupancy in the literature and hypothesised effects on eDNA-based crested newt occurrence are given for each determinant. Any determinants not reported in the literature are listed as NR. Direction of observed effects on eDNA-based great crested newt occupancy determined by each analysis (GLMM assessing number of species in each vertebrate group, *N* = 532; co-occur analysis, *N* = 532; GLMM combining abiotic and biotic factors *N* = 504; and GLMM assessing HSI, *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. Significant P-values (<0.05) are in bold.

Determinant	**Effect reported**	**Hypothesised effect**	**Analysis**					
			Co-occur		**GLMM**			
			Effect	***P***	**DF**	**Effect size (SE)**	χ^2	***P***
Fish	-/0	-			1	-0.239 (0.124)	4.065	**0.044**
Three-spined stickleback	-	-	-	**0.0091**	1	-1.432 (0.561)	9.453	**0.0021**
Ninespine stickleback	-	-	-	**0.0472**				
Common carp	-	-	-	0.0704				
Crucian carp	-	-						
Waterfowl	-	-			1	0.617 (0.181)	13.050	**0.0003**
Coot	NR		+	**0.0232**				
Moorhen	NR		+	**0.0007**				
Green-winged teal	NR		-	0.0987				
Amphibians	NR				1	0.558 (0.149)	16.640	**4.158×10^{-5}**
Smooth newt	+	+	+	**< 0.0001**	1	1.081 (0.303)	17.434	**2.975×10^{-5}**
Common toad	NR		-	**0.0088**	1	-1.635 (0.696)	8.228	**0.0041**
Terrestrial birds	NR				1	-0.335 (0.291)	1.444	0.2295
Common pheasant	NR		-	**0.0479**				
Mammals	NR				1	0.028 (0.091)	0.095	0.7583
Grey squirrel	NR		-	**0.0183**	1	-1.591 (0.534)	12.432	**0.0004**
Badger	NR		-	0.0987				
Pig	NR		+	**0.00395**				
Cow	NR		+	0.0971				

Pond area	-/+	-	1	-0.0004 (0.0002)	6.453	**0.0111**
Pond density	+	+				
Pond depth	+	+	1	0.282 (0.139)	4.266	**0.0389**
Water quality	+	+				
Outflow	NR		1	-0.713 (0.359)	4.467	**0.0346**
Macrophyte cover	-/+	-				
Shading	-/+	-				
Woodland	+	+				
Grassland	+	+				
HSI	0/+	+	1	3.0198 (0.7912)	15.709	**7.388×10^{-5}**
Ruderal	NR		2		6.507	**0.0387**
None				-0.617 (0.527)		
Some				0.032 (0.528)		
Other good terrestrial habitat	NR		2		7.918	**0.0191**
None				0.428 (0.429)		
Some				-0.316 (0.424)		
Species richness	NR		1	0.527 (0.105)	60.267	**8.281×10^{-15}**

841

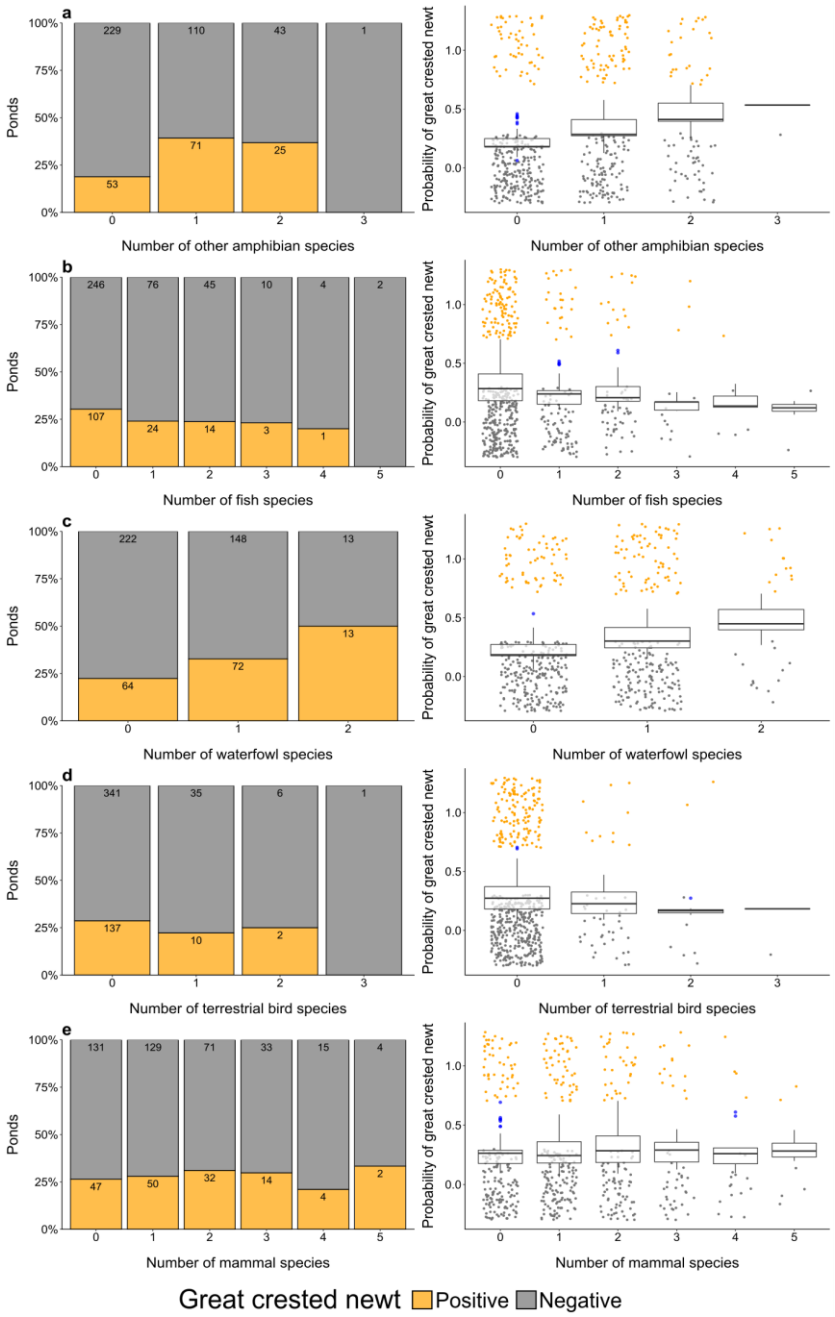

Figure 1 | Great crested newt presence (orange) and absence (grey) in relation to number of species from different vertebrate groups detected by eDNA (*N* = 532 ponds). a, other amphibians; **b,** fish; **c,** waterfowl; **d,** terrestrial birds; **e,** mammals. Observed proportion of ponds with and without great crested newt (left) is plotted alongside predicted probability of great crested newt 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 crested newt 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.

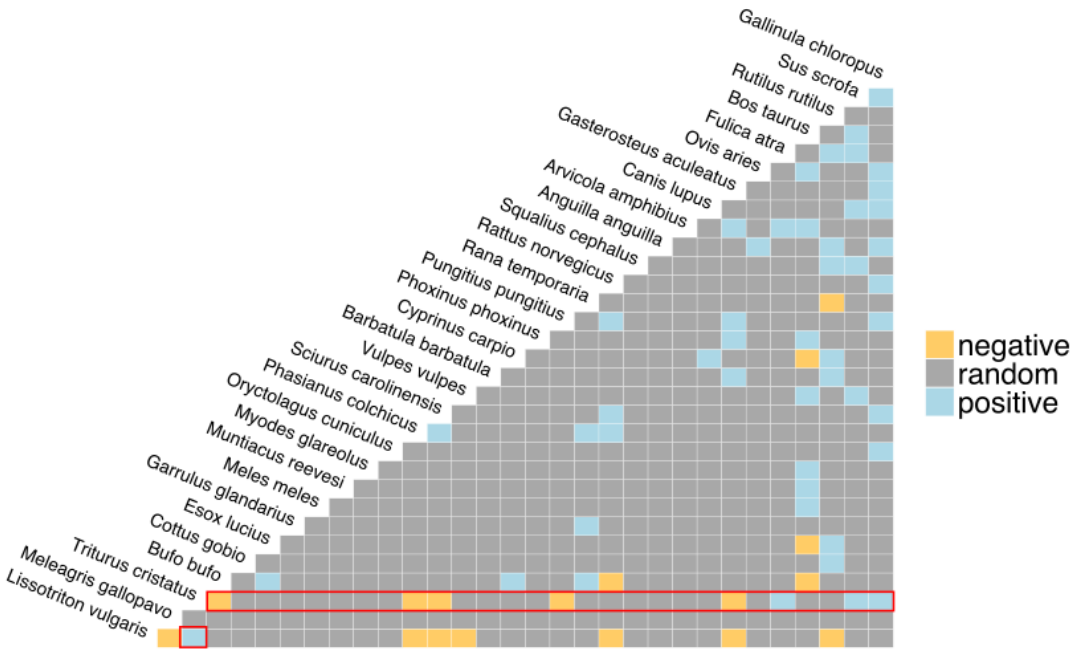

Figure 2 | Heat map showing significant (*P* < 0.05) positive and negative species associations determined by the probabilistic co-occurrence model for the eDNA metabarcoding presence-absence data (*N* = 532 ponds). Species names are positioned to indicate the columns and rows that represent their pairwise 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 great crested newt are highlighted in red.

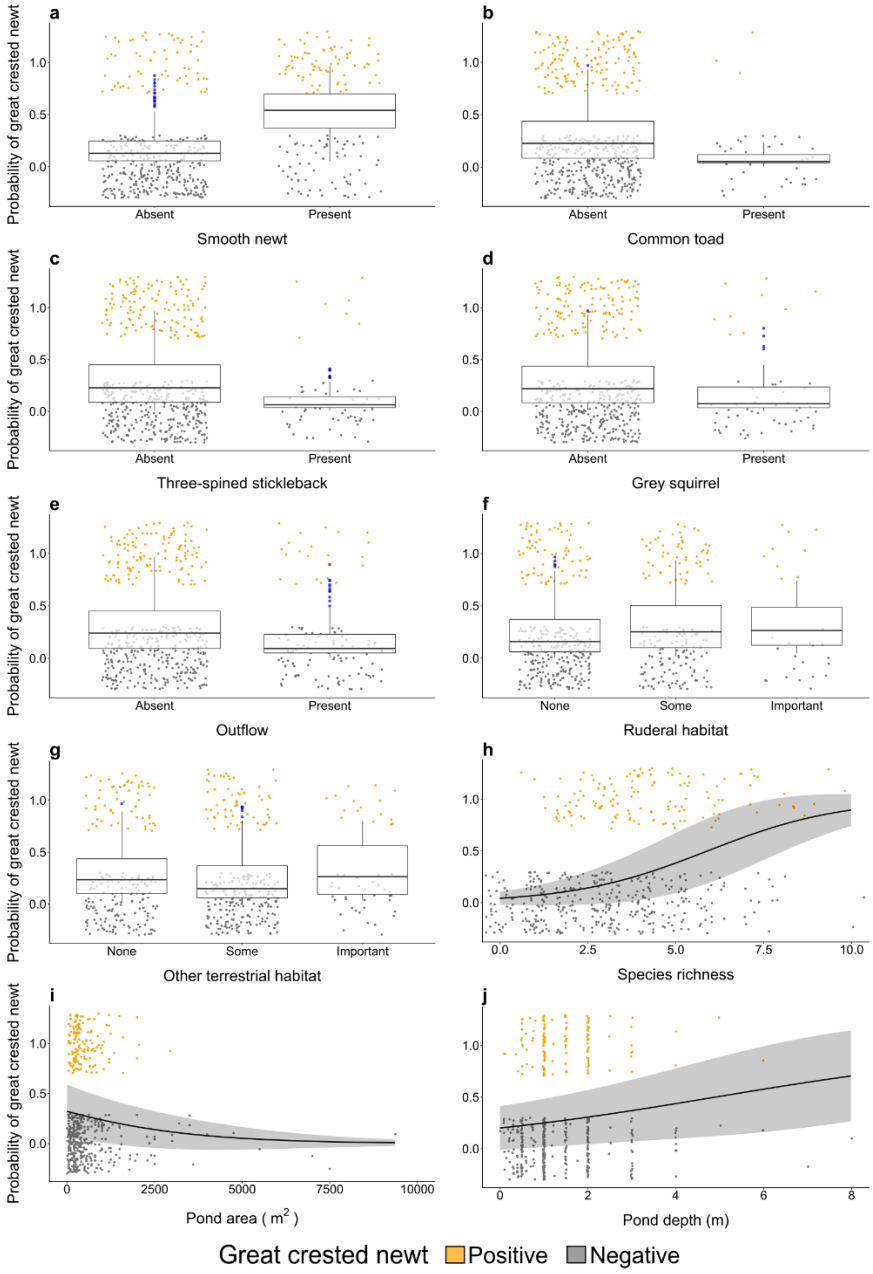

Figure 3 | Biotic and abiotic determinants of great crested newt occurrence, as predicted by the binomial GLMM (*N* = 504 ponds). a, smooth newt occurrence, **b,** common toad occurrence, **c,** three-spined stickleback occurrence, **d,** grey squirrel occurrence, **e,** pond outflow, **f,** ruderal habitat and **g,** other good quality terrestrial habitat, **h,** species richness, **i,** pond area, **j,** pond depth. The 95% CIs, as calculated using the predicted great crested newt probability values and standard error for these predictions, are given for each relationship. The observed great crested newt 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.

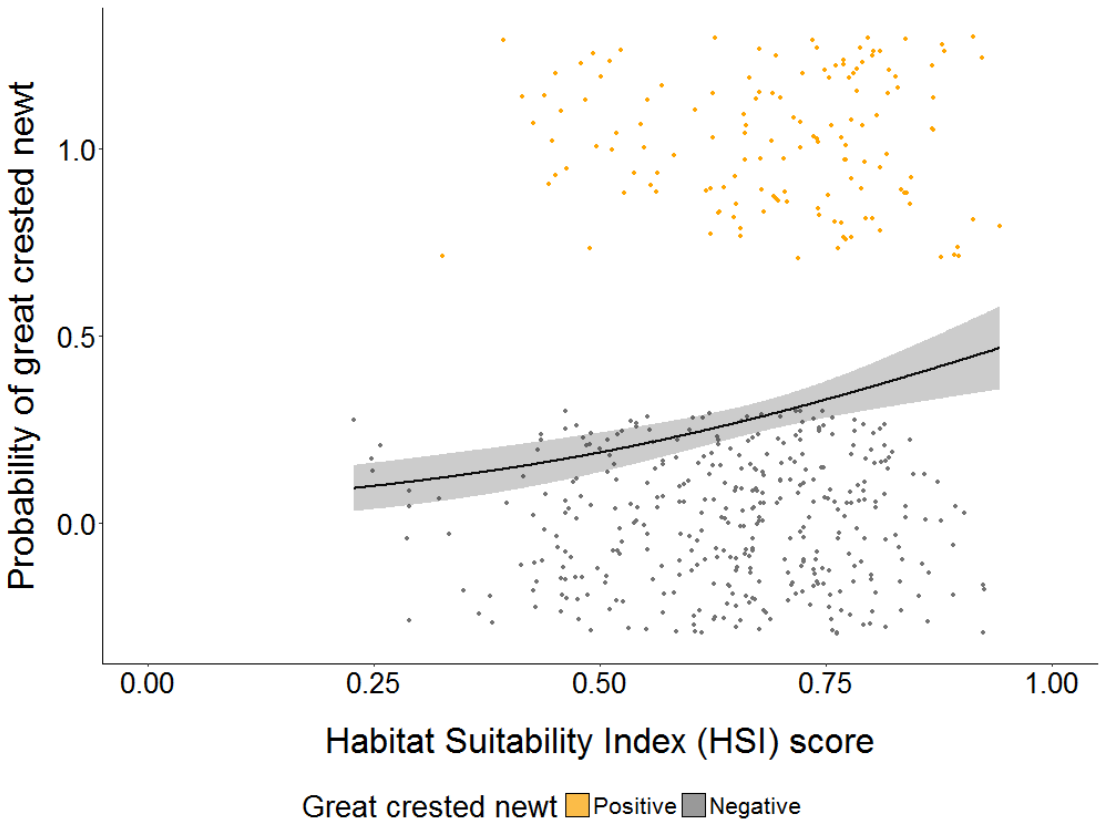

Figure 4 | Relationship between great crested newt occupancy and HSI score, as predicted by the binomial GLMM (*N* = 504 ponds). The 95% CIs, as calculated using the predicted great crested newt probability values and standard error for these predictions, are given. The observed great crested newt presence (orange) and absence (grey) data are shown as points, which have been jittered around 0 and 1 to clarify variation in point density, against the predicted relationship (line).

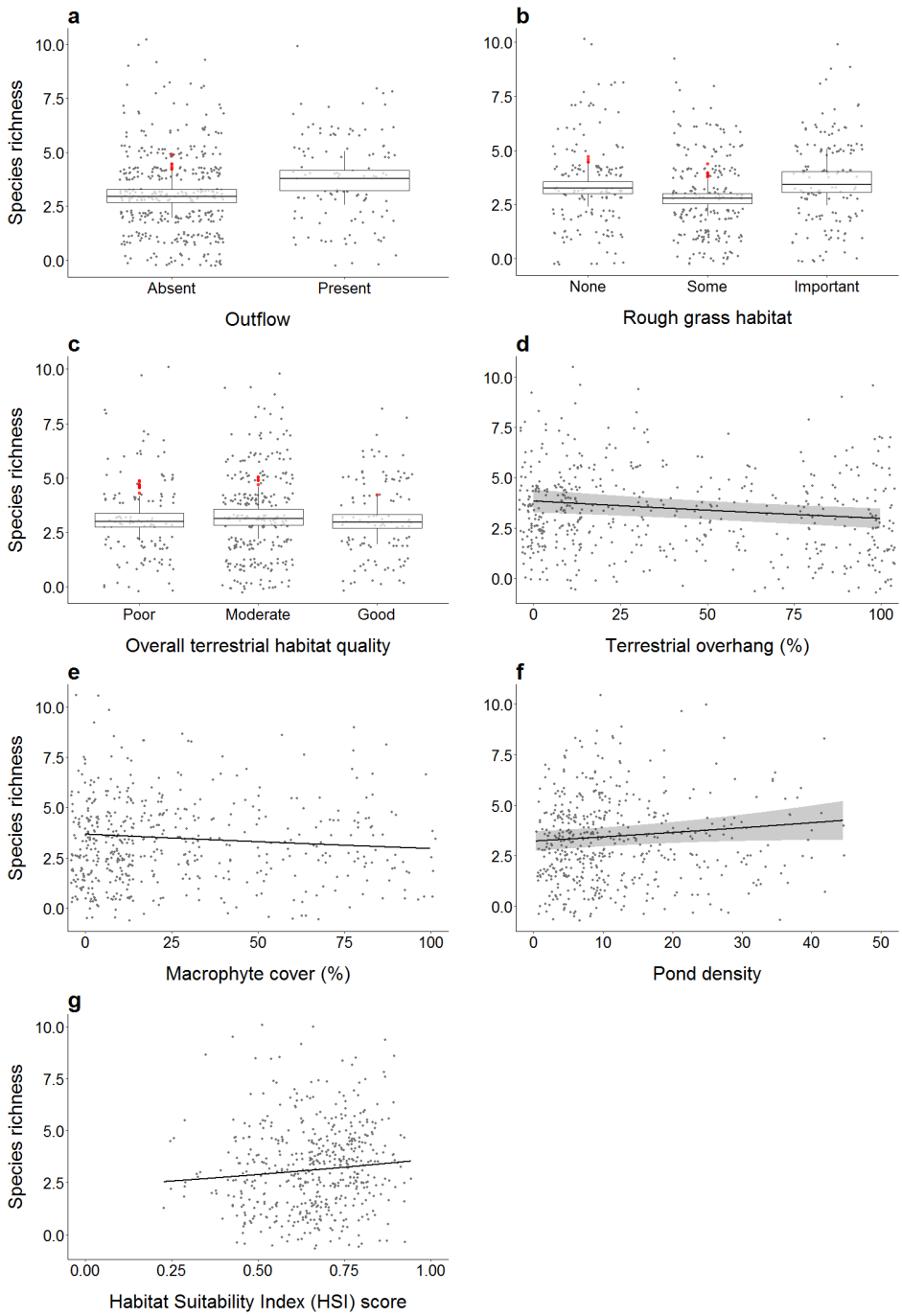

Figure 5 | Abiotic and biotic determinants of vertebrate species richness, as predicted by the Poisson GLMM (*N* = 504 ponds). a, outflow, **b,** rough grass habitat, **c,** overall quality of terrestrial habitat, **d,** percentage of terrestrial overhang, **e,** percentage of macrophyte cover, **f,** 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.