Ground-truthing environmental DNA (eDNA) 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

E-mails: L.Harper@2015.hull.ac.uk, L.Lawson-Handley@hull.ac.uk, B.Haenfling@hull.ac.uk

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

E-mail: christoph.hahn@uni-graz.at

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

E-mails: Erin.Lewis@fera.co.uk, Ian.Adams@fera.co.uk

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

E-mail: neil.boonham@ncl.ac.uk

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

E-mail: Helen.Rees@adas.co.uk

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

E-mails: Peter.Brotherton@naturalengland.org.uk, susanna.phillips@naturalengland.org.uk

Statement of authorship: 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.

Data accessibility: Raw sequence reads have been archived on the NCBI Sequence Read Archive (Bioproject: PRJNA417951; SRA accessions: SRR6285413 - SRR6285678).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).

Running title: Hypothesis testing with eDNA metabarcoding

Key-words (up to 10): biodiversity assessment, community ecology, environmental DNA (eDNA), metabarcoding, ponds, pondscape, management, monitoring, survey, vertebrates

Article type: Letters

Abstract word count: 148 words

Main text word count: 5190 words

Number of references: 66

Number of figures, tables and boxes: 6

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

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

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

Tel: 01482 465804

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 great crested newts (*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 associations with smooth newt (*Lissotriton vulgaris*) and common toad (*Bufo bufo*), but negatively correlated to fish species richness, specifically stickleback presence. Both GCN occupancy and species richness correlated with the GCN Habitat Suitability Index score, supporting its application to GCN survey. We reaffirm reported associations (e.g. GCN preference for deeper ponds) but also provide novel insights, including a negative effect of pond outflow on GCN. Our findings demonstrate the prospects of eDNA metabarcoding for hypothesis testing at unprecedented scales.

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 (Rees *et al.* 2014; Lawson Handley 2015). 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. Hänfling *et al.* 2016; Kelly *et al.* 2016; Olds *et al.* 2016; Evans *et al.* 2017), its applications in community ecology are relatively unexplored.

Aquatic ecosystems are highly suited to eDNA studies (Muha *et al.* 2017) as eDNA exists in multiple states with rapid modes of transport and degradation, increasing detectability of contemporary biodiversity (Rees *et al.* 2014; Barnes & Turner 2015). 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 (Valentini *et al.* 2016; Bálint *et al.* 2017; Evans *et al.* 2017; Klymus *et al.* 2017; Ushio *et al.* 2017). These small and abundant water bodies span broad ecological gradients (De Meester *et al.* 2005) and comprise

96 pondscapes - a network of ponds and their surrounding terrestrial habitat (Hill *et al.*
97 2018). Pondscapes contribute substantially to aquatic and non-aquatic biodiversity
98 across spatial scales, with ponds supporting many rare and protected species in
99 fragmented landscapes (De Meester *et al.* 2005; Biggs *et al.* 2016; Hill *et al.* 2018).
100 Consequently, ponds are model systems for experimental validation and examination
101 of biogeographical patterns (De Meester *et al.* 2005). Habitat complexity and tools
102 required for different taxa with associated bias (Evans *et al.* 2017) and cost (Valentini
103 *et al.* 2016) once hindered exhaustive sampling of pond biodiversity (Hill *et al.* 2018),
104 but eDNA metabarcoding may overcome these barriers.

105 In the UK, the threatened great crested newt (*Triturus cristatus,* hereafter
106 GCN) is a flagship species for pond conservation. The extensive literature on the
107 GCN provides an excellent opportunity to ground-truth ecological patterns revealed
108 by eDNA metabarcoding. Both biotic (e.g. breeding substrate, prey, and predators)
109 and abiotic (e.g. pond area, depth, and temperature) factors are known to influence
110 GCN breeding success (Langton *et al.* 2001). The GCN Habitat Suitability Index
111 (HSI) (Oldham *et al.* 2000; ARG-UK 2010) accounts for these factors using 10
112 suitability indices that are scored and combined to calculate a decimal score between
113 0 and 1 (where 1 = excellent habitat). Larvae are susceptible to fish and waterfowl
114 predation (Rannap & Briggs 2006; Skei *et al.* 2006; Hartel *et al.* 2010), and adults
115 reportedly avoid ponds containing three-spined stickleback (*Gasterosteus aculeatus*)
116 (McLee & Scaife 1992), ninespine stickleback (*Pungitius pungitius*), crucian carp
117 (*Carassius carassius*), and common carp (*Carassius carpio*) (Rannap *et al.* 2009a, b).
118 Conversely, the GCN and smooth newt (*Lissotriton vulgaris*) prefer similar habitat
119 and often co-occur (Rannap & Briggs 2006; Skei *et al.* 2006; Rannap *et al.* 2009a;
120 Denoël *et al.* 2013). GCNs thrive in ponds with good water quality as indicated by

121 diverse macroinvertebrate communities (Oldham *et al.* 2000; Rannap *et al.* 2009a),
122 and water clarity is important for breeding displays, foraging success, and egg
123 survival (Rannap & Briggs 2006; Skei *et al.* 2006). Pond networks encourage GCN
124 occupancy (Joly *et al.* 2001; Rannap *et al.* 2009a; Hartel *et al.* 2010; Denoël *et al.*
125 2013), but larger area discourages presence (Joly *et al.* 2001). Ponds with heavy
126 shading (Vuorio *et al.* 2013) or dense macrophyte cover (Rannap & Briggs 2006; Skei
127 *et al.* 2006; Hartel *et al.* 2010) are unlikely to support viable populations. GCNs also
128 depend on terrestrial habitat, preferring open, semi-rural pondscapes (Denoël *et al.*
129 2013) containing pasture, extensively grazed and rough grassland, scrub, and
130 coniferous and deciduous woodland (Oldham *et al.* 2000; Rannap & Briggs 2006;
131 Rannap *et al.* 2009a; Gustafson *et al.* 2011; Vuorio *et al.* 2013).

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

142

143

144

145 **MATERIALS AND METHODS**

146

147 **Samples**

148

149 Samples from 508 ponds included in Natural England's Great Crested Newt Evidence
150 Enhancement Programme were processed using eDNA metabarcoding alongside 24
151 privately surveyed ponds. Water samples were collected using established
152 methodology (Biggs *et al.* 2015), detailed in Supporting Information. In brief, 20 x 30
153 mL water samples were collected from each pond and pooled. Six 15 mL subsamples
154 were taken from the pooled sample and each added to 33.5 mL absolute ethanol and
155 1.5 mL sodium acetate 3 M (pH 5.2). Subsamples were pooled during DNA
156 extraction to produce one eDNA sample per pond. Targeted qPCR detected GCN in
157 265 (49.81%) ponds (Harper *et al.* 2018).

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

166

167

168 **DNA reference database construction**

169

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

184

185

186 **Primer validation**

187

188 Reference databases were combined for *in silico* validation of published 12S rRNA
189 primers 12S-V5-F (5'-ACTGGGATTAGATACCCC-3') and 12S-V5-R (5'-
190 TAGAACAGGCTCCTCTAG-3') (Riaz *et al.* 2011) using ecoPCR software (Ficetola
191 *et al.* 2010). Set parameters allowed a 50-250 bp fragment and three mismatches
192 between each primer and reference sequence. Primers were validated *in vitro* for UK

193 fish by Hänfling *et al*. (2016) and here for six UK amphibian species (Fig. S2).

194

195

196 **eDNA metabarcoding**

197

198 Our two-step PCR workflow is fully described in Harper *et al*. (2018) and Supporting
199 Information. eDNA was first amplified with the aforementioned primers, where PCR
200 positive controls (six per PCR plate; N = 114) were cichlid (*Rhamphochromis esox*)
201 DNA and PCR negative controls (six per PCR plate; N = 114) were sterile molecular
202 grade water (Fisher Scientific UK Ltd, UK). PCR products were individually purified
203 using E.Z.N.A.® Cycle Pure V-Spin Clean-Up Kits (Omega Bio-tek, GA, USA)
204 following the manufacturer's protocol. The second PCR bound Multiplex
205 Identification (MID) tags to the purified products. PCR products were individually
206 purified using magnetic bead clean-up and quantified with a Quant-IT™ PicoGreen™
207 dsDNA Assay (Invitrogen, ThermoFisher Scientific, UK). Samples were normalised,
208 pooled, and libraries quantified using a Qubit™ dsDNA HS Assay Kit (Invitrogen,
209 ThermoFisher Scientific, UK). Libraries were sequenced on an Illumina MiSeq using
210 2 x 300 bp V3 chemistry (Illumina Inc., San Diego, CA, USA) and raw sequence
211 reads processed using metaBEAT (metaBarcoding and Environmental Analysis Tool)
212 v0.8 (https://github.com/HullUni-bioinformatics/metaBEAT). After quality trimming,
213 merging, chimera detection, and clustering, non-redundant query sequences were
214 compared against our reference database using BLAST (Zhang *et al.* 2000). Putative
215 taxonomic identity was assigned using a lowest common ancestor (LCA) approach
216 based on the top 10% BLAST matches for any query matching with at least 98%
217 identity to a reference sequence across more than 80% of its length. Unassigned

218 sequences were subjected to a separate BLAST against the complete NCBI nucleotide
219 (nt) database at 98% identity to determine the source via LCA as described above.
220 The bioinformatic analysis has been deposited in the GitHub repository for
221 reproducibility.
222
223
224 **Data analysis**
225
226 Analyses were performed in the statistical programming environment R v.3.4.2 (R
227 Core Team 2017). Data and R scripts have been deposited in the GitHub repository.
228 Assignments from different databases were merged, and spurious assignments (i.e.
229 non-UK species, invertebrates and bacteria) removed from the dataset. The genus
230 *Strix* was reassigned to tawny owl (*Strix aluco*) as it is the only UK representative of
231 this genus. Where family and genera assignments containing a single UK
232 representative had reads assigned to species, reads from all assignment levels were
233 merged and manually assigned to that species. Consequently, all taxonomic
234 assignments in the final dataset were of species resolution. The only misassignment
235 was Scottish wildcat (*Felis silvestris*), which does not occur at the sampling localities
236 (Kent, Lincolnshire and Cheshire) and was therefore reassigned to domestic cat (*Felis*
237 *catus*).
238 To minimise risk of false positives, species were only classed as present at
239 sites if their sequence frequency exceeded species-specific thresholds. Thresholds
240 were defined using the maximum sequence frequency of each species in PCR positive
241 controls (N = 114; Table S3). For example, the GCN threshold was 0.028% to omit
242 false positives in PCR positive controls. After applying thresholds, the read count data

243 were converted to a species presence-absence matrix. Analyses were based on
244 species-specific thresholds, but also performed for different blanket sequence
245 thresholds (0.05 - 30%, Tables S4-9). We tested biotic and abiotic determinants of
246 GCN occupancy and vertebrate species richness, and appropriateness of the HSI.
247 Hypotheses are summarised in Table 1.

248 All Generalised Linear Mixed Models (GLMMs) were executed using the R
249 package 'lme4' v1.1-12 (Bates *et al.* 2015). GCN occurrence relating to number of
250 other vertebrate species was investigated with a binomial GLMM, and species
251 associations identified using the R package 'cooccur' v1.3 (Griffith *et al.* 2016) (N =
252 532). Identified associations informed candidate biotic variables for inclusion with
253 abiotic variables (Table S1) in a binomial GLMM of GCN occurrence (N = 504).
254 Candidate explanatory variables were assessed for collinearity, relative importance,
255 and non-linearity (see Supporting Information). HSI score and band were collinear,
256 thus HSI score was analysed in a separate binomial GLMM. Using the R package
257 'ncf' v1.1-7 (Bjornstad 2009), spline correlograms of the Pearson residuals from a
258 binomial Generalised Linear Model (GLM) and GLMM were compared to assess
259 spatial autocorrelation potential in our dataset. Sample was treated as a random effect
260 in the GLMM to account for dependencies within ponds (Zuur *et al.* 2009). The
261 mixed model successfully accounted for spatial autocorrelation within sites. The same
262 steps were performed to identify explanatory variables and a modeling framework for
263 vertebrate species richness. A Poisson distribution was specified for all species
264 richness models as the response variable was integer count data.

265 Binomial and Poisson models considered respectively were nested thus best
266 models were chosen using stepwise backward deletion of terms based on Likelihood
267 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 *et al.* 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).

RESULTS

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 (Fig. S3).

Pondscape biodiversity

All native amphibians were found as well as 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), European bullhead (*Cottus gobio*, *N* = 14), 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), but also introduced species, including common carp (*N* = 41), ruffe (*Gymnocephalus cernua*, *N* = 1), and rainbow trout (*Oncorhynchus mykiss*, *N* = 3). Regular and irregular waterfowl were identified, ranging from common moorhen (*Gallinula chloropus*, *N* = 215) to grey heron (*Ardea cinerea*, *N* = 1) and Eurasian oystercatcher (*Haematopus ostralegus*, *N* = 1). Terrestrial fauna were often detected in fewer than five ponds (Fig. S3c, d). Buzzard (*Buteo buteo*, *N* = 4), Eurasian jay (*Garrulus glandarius*, *N* = 7), dunnock (*Prunella modularis*, *N* = 4), and starling (*Sturnus vulgaris*, *N* = 4) were the most frequently detected terrestrial birds. 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. Nonetheless, we detected several mammals with Biodiversity Actions Plans and/or of conservation concern (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* =

318 1) and water shrew (*Neomys fodiens*, N = 9). Notably, the invasive American mink
319 (*Neovison vison*) was absent despite widespread UK distribution (Mathews *et al.*
320 2018). All species and their detection frequencies are listed in Table S10.
321
322
323 **Biotic determinants of GCN occurrence**
324
325 GCN occupancy was negatively correlated with fish species richness, but positively
326 influenced by amphibian and waterfowl species richness (Fig. 1, GLMM:
327 overdispersion χ^2_{525} = 517.636, P = 0.582; fit χ^2_8 = 22.524, P = 0.004, R^2 = 9.43%).
328 The GCN had significant positive associations with four species (Fig. 2), including
329 the smooth newt, common coot (*Fulica atra*), common moorhen, and pig. However,
330 the GCN had significant negative associations with five species (Fig. 2), including the
331 common toad, ninespine stickleback, three-spined stickleback, grey squirrel, and
332 common pheasant (*Phasianus colchicus*). Non-significant negative associations were
333 also observed between the GCN and common carp, green-winged teal (*Anas*
334 *carolinensis*), and badger (*Meles meles*). Only presence-absence of the smooth newt,
335 common toad, three-spined stickleback, and grey squirrel were retained by model
336 selection as explanatory variables for the GLMM (Figs. 3a-d) with abiotic
337 determinants. Results of analyses are summarised and compared to previously
338 reported determinants in Table 1.
339 GCNs were more likely to occupy ponds with more amphibian species (Fig.
340 1a). GCNs were detected in 51.97% of ponds (N = 152) containing smooth newts, but
341 in only 11.91% of ponds (N = 42) with common toads (Fig. S3a). GCN occurrence
342 probability was lower in ponds with more fish species, and GCNs were absent from

343 ponds with more than four fish species (Fig. 1b). The GCN was only found in 14.29%
344 (N = 56) and 6.67% (N = 15) of ponds inhabited by the three-spined stickleback and
345 ninespine stickleback respectively (Fig. S3b). In contrast, GCNs were more likely to
346 occur in ponds with more waterfowl species (Fig. 1c). GCNs occupied 41.67% (N =
347 48) and 35.81% (N = 215) of ponds with common coot and common moorhen
348 respectively (Fig. S3c). No significant relationships between terrestrial bird or
349 mammal species richness and GCN occupancy were found (Figs. 1d, e). Yet, species
350 associations between the GCN and pig, common pheasant, and grey squirrel were
351 identified. The GCN was detected in 37.14% of ponds (N = 140) where pig was
352 present (Fig. S3d) as opposed to 12.00% (N = 25) and 15.79% (N = 57) of ponds with
353 common pheasant (Fig. S3c) and grey squirrel (Fig. S3d) records respectively. GCN
354 occupancy positively correlated with overall vertebrate species richness (Fig. 3h),
355 irrespective of individual species associations.
356
357
358 **Abiotic determinants of GCN occurrence**
359
360 Five abiotic determinants were retained in the GLMM explaining GCN occupancy
361 (GLMM: overdispersion χ^2_{490} = 413.394, P = 0.995; fit χ^2_8 = 11.794, P = 0.161, R^2 =
362 38.58%). The probability of GCN occurrence increased with greater pond depth but
363 decreased in ponds with larger area, outflow, some other good terrestrial habitat, and
364 where ruderal habitat was absent (Table 1, Figs. 3e-g, i-j).
365
366

367 **Abiotic determinants of vertebrate species richness**

368

369 Our analysis (Table 2, GLMM: overdispersion χ^2_{494} = 431.959, P = 0.979; fit χ^2_8 = -
370 42.708, P = 1.000, R^2 = 8.94%) revealed species richness was greater in ponds with
371 outflow (Fig. 4a), but reduced in those with some rough grass habitat compared to
372 ponds with no or important rough grass habitat (Fig. 4b). Overall quality of terrestrial
373 habitat was also influential (Fig. 4c), where ponds had higher species richness in areas
374 considered to be poor or moderate habitat for GCNs. Species richness was reduced as
375 percentages of terrestrial overhang (Fig. 4d) and macrophyte cover increased (Fig.
376 4e), but improved with pond density (Fig. 4f).

377

378

379 **Applicability of the HSI to GCN and community eDNA surveys**

380

381 HSI score positively correlated with both GCN occurrence (GLMM: overdispersion
382 χ^2_{501} = 506.763, P = 0.4198; fit χ^2_8 = 8.118, P = 0.422, R^2 = 4.99%) and vertebrate
383 species richness (GLMM: overdispersion χ^2_{501} = 389.744, P = 0.999; fit χ^2_8 = -
384 145.12, P = 1.000, R^2 = 1.10%). GCN occupancy probability (Table 1, Fig. 3k) and
385 vertebrate species richness (Table 2, Fig. 4g) were improved in ponds with higher HSI
386 score.

387

388

389

390 **DISCUSSION**

391

392 We have ground-truthed eDNA metabarcoding as a tool for ecological hypothesis
393 testing using the community data generated by this tool in combination with
394 environmental metadata for ponds. We tested biotic and abiotic determinants of GCN
395 occupancy and wider vertebrate biodiversity as well as the appropriateness of the
396 GCN HSI for eDNA survey. GCN occupancy was higher in ponds containing smooth
397 newts, but devoid of common toads, three-spined sticklebacks, and grey squirrels.
398 Ponds inhabited by GCNs were typically small, deep, absent of outflow, and
399 surrounded by ruderal and good quality terrestrial habitat. Vertebrate species richness
400 was higher in ponds where outflow, some rough grass habitat, and poor or moderate
401 terrestrial habitat for the GCN were present. Species richness was lower at higher
402 percentages of terrestrial overhang and macrophyte cover, but greater at higher pond
403 density. The GCN HSI was appropriate for predicting both GCN occupancy and
404 vertebrate species richness. Our findings demonstrate the power of eDNA
405 metabarcoding to reboot freshwater monitoring and research by providing
406 biodiversity data en masse at low cost.

407

408

409 **Pondscape biodiversity**

410

411 eDNA metabarcoding detected six amphibian, 14 fish, 18 bird, and 22 mammal
412 species across 532 UK ponds. This diverse species inventory emphasises the
413 importance of ponds as habitat for aquatic taxa, but also as stepping stones for semi-
414 aquatic and terrestrial taxa (De Meester *et al.* 2005; Hill *et al.* 2018) through

415 provision of drinking, foraging, dispersal, and reproductive opportunities (Biggs *et al.*
416 2016; Klymus *et al.* 2017). Some species detections may be the result of eDNA
417 transport from water bodies in the surrounding area (Hänfling *et al.* 2016) to ponds
418 via inflow; however, this signifies the capacity of ponds to provide natural samples of
419 freshwater biodiversity in the wider catchment.
420
421
422 **Biotic determinants of GCN occurrence**
423
424 GCN occurrence was positively associated with vertebrate species richness, which
425 may support its status as an indicator or umbrella species for pond biodiversity
426 (Gustafson *et al.* 2006). GCNs were more likely to occur in ponds with higher
427 amphibian species richness - particularly ponds containing smooth newts and absent
428 of common toads. GCNs and smooth newts have similar habitat requirements and
429 tend to breed in the same ponds (Skei *et al.* 2006; Rannap *et al.* 2009a; Denoël *et al.*
430 2013), with >60% overlap reported (Rannap & Briggs 2006). However, smooth newts
431 can inhabit a broader range of habitat (Rannap & Briggs 2006; Skei *et al.* 2006) than
432 GCNs, which depend on larger, deeper ponds with abundant macrophytes and no fish
433 located in open, semi-rural landscapes (Denoël *et al.* 2013). Common toads inhabit
434 fish-containing ponds (Manenti & Pennati 2016) which may explain their negative
435 association with GCNs, but newts may also predate toad eggs and larvae (Langton *et*
436 *al.* 2001).
437 GCNs were negatively associated with higher fish species richness, and
438 specifically the presence of three-spined and ninespine sticklebacks - fish that are
439 common in and typical of ponds. All GCN life stages may be predated by fishes

440 (Langton *et al.* 2001) and negative effects of fish presence-absence on GCN
441 occupancy, distribution, and abundance are repeatedly reported (Joly *et al.* 2001;
442 Rannap & Briggs 2006; Skei *et al.* 2006; Denoël & Ficetola 2008; Rannap *et al.*
443 2009a, b; Hartel *et al.* 2010; Denoël *et al.* 2013). Three-spined sticklebacks predate
444 GCN eggs and larvae (McLee & Scaife 1992; Jarvis 2010), and have non-
445 consumptive effects on GCN embryos (Jarvis 2010). GCN larvae were also found to
446 alter their behaviour when exposed to the predatory three-spined stickleback but not
447 the non-predatory crucian carp (Jarvis 2012), another fish characteristic of ponds.

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

464 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 (Perrow *et al.* 1997; Paillisson & Marion 2001; 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 *et al.* 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 (Langton *et al.* 2001; Gustafson *et al.* 2011). 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.

488 **Abiotic determinants of GCN occurrence**

489

490 GCNs were more likely to occupy deeper ponds, but less likely to inhabit larger
491 ponds, with outflow present, no ruderal habitat, and some other good terrestrial
492 habitat. Consistent with our results, pond depth was previously found to positively
493 influence GCN occupancy (Denoël & Ficetola 2008). Shallow ponds can be
494 inhospitable due to drying or freezing and may contain less prey, but detrimental
495 effects are often observed in open farmland (Denoël & Ficetola 2008). Although our
496 results indicate GCNs prefer larger ponds, pond area does not always influence
497 occurrence (Maletzky *et al.* 2007; Denoël & Ficetola 2008; Gustafson *et al.* 2011) and
498 was deemed a poor predictor of reproductive success (Vuorio *et al.* 2013). GCNs have
499 been found to utilise small and large ponds (Rannap & Briggs 2006; Skei *et al.* 2006);
500 however, very small ponds (less than 124 m^2) may be unable to support all life stages,
501 and larger ponds may contain fish and experience eutrophication from agricultural or
502 polluted run-off (Rannap & Briggs 2006). Effects of pond outflow (facilitated by
503 drains, pipes or streams) are understudied, whereas inflow impacts biodiversity via
504 polluted agricultural run-off and connections to streams and rivers containing large,
505 predatory fish. Based on our findings that show outflow presence deters GCN
506 occupation, we suggest outflow minimises fluctuations in pond depth (Freshwater
507 Habitats Trust 2015) and alters patterns of colonisation and community structure to
508 those of ponds which fluctuate with precipitation.

509 Our results support the importance of good terrestrial habitat to GCNs for
510 shelter, foraging and dispersal (Langton *et al.* 2001). We observed higher GCN
511 occurrence in ponds surrounded by ruderal or other good quality terrestrial habitat.
512 However, the majority of terrestrial habitat variables were not retained by our model

513 selection. Hartel *et al*. (2010) also found landscape variables, excluding urbanisation,
514 were inadequate predictors of GCN distribution, although their study area was mostly
515 rural and thus optimal for amphibians. Conversely, research on more diverse
516 landscapes found GCN occupancy was lower in coniferous forest, yet higher in
517 deciduous or herb-rich forest and pasture (Gustafson *et al.* 2011; Vuorio *et al.* 2013).
518 In our study, the metadata available were qualitative, preventing detailed analyses on
519 terrestrial habitat quality in relation to GCN occupancy. Better understanding of
520 occupancy and interactions with terrestrial species could be achieved with
521 quantitative data on terrestrial habitat type, density, distance to ponds, and species
522 utilisation, but this is a phenomenal task for large numbers of ponds across a vast
523 landscape (Denoël & Ficetola 2008).
524
525
526 **Abiotic determinants of vertebrate species richness**
527
528 Vertebrate species richness was higher in ponds where outflow was present, with poor
529 or moderate overall terrestrial habitat for GCNs, and more ponds nearby. Conversely,
530 species richness was lower in ponds with higher percentages of macrophyte cover and
531 terrestrial overhang, and absent of or surrounded by some rough grass habitat. We
532 compare our results to previous studies of aquatic species richness but these largely
533 focus on species assemblages or guilds, primarily macrophytes, macroinvertebrates,
534 and amphibians. Outflow and inflow have been understudied in relation to pond
535 biodiversity, although outflow may release harmful pollutants and pathogens (Beutel
536 & Larson 2015) that would accumulate in a closed pond system. Species richness
537 increased as pond density increased, which echoes the positive relationship between

538 pond density and macrophyte and macroinvertebrate richness observed by Gledhill *et*
539 *al*. (2008). These findings combined again highlight the importance of pond networks
540 for aquatic and non-aquatic taxa (Hill *et al.* 2018). Shade was identified as a principal
541 driver of macroinvertebrate and macrophyte diversity in ponds (Sayer *et al.* 2012).
542 Yet, canopy and macrophyte cover positively influence amphibian species richness
543 (Piha *et al.* 2007). Plentiful rough grass habitat may create more ecological niches and
544 foraging opportunity for different vertebrates, but quantitative data on type and
545 abundance of terrestrial habitat are needed to understand species preferences.
546
547
548 **Applicability of the HSI to GCN and community eDNA surveys**
549
550 We found the HSI can predict eDNA-based GCN occupancy at the pondscape. This
551 contradicts previous conventional studies which deemed the index inappropriate for
552 predicting GCN occupancy or survival probabilities (Unglaub *et al.* 2015). Our results
553 also suggest some indices comprising the GCN HSI, for example, outflow and
554 terrestrial habitat, represent suitability criteria for other biodiversity. An adapted HSI,
555 designed to predict species richness, could help select areas for management and
556 enhancement of aquatic and non-aquatic biodiversity. Nonetheless, the GCN HSI also
557 confers protection to pond biodiversity by identifying optimal habitat for pond
558 creation and restoration to encourage populations of this threatened amphibian. The
559 HSI is not without issue due to qualitative data used for score calculation and
560 subjective estimation of indices (Oldham *et al.* 2000). For future application of this
561 index in GCN eDNA survey, we recommend metabarcoding to quantify some
562 qualitatively assessed indices (e.g. water quality via macroinvertebrate diversity, fish

563 and waterfowl presence) alongside GCN detection. Provided rigorous spatial and
564 temporal sampling are undertaken, eDNA metabarcoding can also generate site
565 occupancy data to estimate relative species abundance (Hänfling *et al.* 2016;
566 Valentini *et al.* 2016).
567
568
569 **Prospects of eDNA metabarcoding for conservation, management, and research**
570
571 We have demonstrated the effectiveness of eDNA metabarcoding for landscape-scale
572 biodiversity monitoring and ecological hypothesis testing. We investigated
573 associations between aquatic and non-aquatic vertebrates, and combined
574 metabarcoding with environmental metadata to revisit important ecological
575 hypotheses at an unprecedented scale. Our findings indicate preferred habitat of a
576 threatened amphibian. GCNs were more likely to occupy ponds where smooth newts
577 were present, but common toads, three-spined sticklebacks, and grey squirrels were
578 absent. GCNs prefer small but deep ponds that are absent of outflow, and surrounded
579 by ruderal and good quality terrestrial habitat. These findings will guide management
580 in the face of increasing land-use and habitat fragmentation - a poignant issue as
581 protective legislation for the GCN in the UK is changing. Whilst conservation of
582 threatened species and their habitat should be a priority, the bigger picture should not
583 be ignored. eDNA metabarcoding could enhance our understanding of freshwater
584 networks - particularly pondscapes - to enable more effective monitoring, protection,
585 and management of aquatic and non-aquatic biodiversity. We are only now beginning
586 to realise and explore these opportunities.
587

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).

REFERENCES

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, 60, 255–265.

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. *et al.* (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. *Conserv. Genet.*, 17, 1–17.

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.*, 67, 1–48.

Beutel, M.W. & Larson, L. (2015). Pathogen removal from urban pond outflow using rock biofilters. *Ecol. Eng.*, 78, 72–78.

Biggs, J., Ewald, N., Valentini, A., Gaboriaud, C., Dejean, T., Griffiths, R.A., *et al.* (2015). Using eDNA to develop a national citizen science-based monitoring programme for the great crested newt (*Triturus cristatus*). *Biol. Conserv.*, 183, 19–28.

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.

Bjørnstad, O.N. (2016). 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. *et al.* (2017). Environmental DNA metabarcoding: transforming how we survey animal and plant communities. *Mol. Ecol.*, 26, 5872–5895.

De Meester, L., Declerck, S., Stoks, R., Louette, G., Van De Meutter, F., De Bie, T. *et al.* (2005). Ponds and pools as model systems in conservation biology, ecology and evolutionary biology. *Aquat. Conserv.*, 15, 715–725.

Denoël, M. & Ficetola, G.F. (2008). Conservation of newt guilds in an agricultural landscape of Belgium: the importance of aquatic and terrestrial habitats. *Aquat. Conserv.*, 18, 714–728.

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.

Evans, N.T., Li, Y., Renshaw, M.A., Olds, B.P., Deiner, K., Turner, C.R. *et al.* (2017). Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. *Can. J. Fish. Aquat. Sci.*, 74, 1362-1374.

Ficetola, G.F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J. *et al.* (2010). An *In silico* approach for the evaluation of DNA barcodes. *BMC Genomics*, 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. *Landsc. Ecol.*, 23, 1219–1230.

Griffith, D., Veech, J. & Marsh, C. (2016). cooccur: Probabilistic Species Co-Occurrence Analysis in R. *J. Stat. Softw.*, 69, 1–17.

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*). *Ann. Zool. Fennici*, 48, 295–307.

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

Hänfling, B., Lawson Handley, L., Read, D.S., Hahn, C., Li, J., Nichols, P. *et al.* (2016). Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. *Mol. Ecol.*, 25, 3101–3119.

Harper, L.R., Lawson Handley, L., Hahn, C., Boonham, N., Rees, H.C., Gough, K.C. *et al.* (2018). Needle in a haystack? A comparison of eDNA metabarcoding and targeted qPCR for detection of the great crested newt (*Triturus cristatus*). *Ecol. Evol.*, 8, 6330-6341.

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. *Environ. Conserv.*, 37, 195–200.

Hering, D., Borja, A., Jones, J.I., Pont, D., Boets, P., Bouchez, A., *et al.* (2018). Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive. *Water Res.*, 138, 192–205.

Hervé, M. (2018) RVAideMemoire: Diverse basic statistical and graphical functions. R package version 0.9-69.

Hill, M.J., Hassall, C., Oertli, B., Fahrig, L., Robson, B.J., Biggs, J. *et al.* (2018). New policy directions for global pond conservation. *Conserv. Lett.*, e12447.

Hosmer, D.W. & Lemeshow, S. (2000). *Applied Logistic Regression*. Second Edition. John Wiley & Sons, Inc., New York, USA, pp. 31–46.

Jarvis, L.E. (2010). Non-consumptive effects of predatory three-spined sticklebacks (*Gasterosteus aculeatus*) on great crested newt (*Triturus cristatus*) embryos. *Herpetol. J.*, 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. *Conserv. Biol.*, 15, 239–248.

Kelly, R.P., O'Donnell, J.L., Lowell, N.C., Shelton, A.O., Samhouri, J.F., Hennessey, S.M. *et al.* (2016). Genetic signatures of ecological diversity along an urbanization gradient. *PeerJ*, 4, e2444.

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, 54.

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. *Aquat. Bot.*, 46, 1–15.

Lawson Handley, L. (2015). How will the "molecular revolution" contribute to biological recording? *Biol. J. Linn. Soc. Lond.*, 115, 750–766.

Lele, S.R., Keim, J.L. & Solymos, P. (2016). ResourceSelection: Resource Selection (Probability) Functions for Use-Availability Data. R package version 0.3-2.

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. *Biol. Conserv.*, 212, 74–85.

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). *Ann. Limnol. – Int. J. of Lim.*, 43, 107–115.

Manenti, R. & Pennati, R. (2016). Environmental factors associated with amphibian breeding in streams and springs: effects of habitat and fish occurrence. *Amphib-reptil.*, 37, 237–242.

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). R package version 2.1-1.

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*). *Herpetol. Bull.*, 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. *Front. Ecol. Evol.*, 5, 158.

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*). *Herpetol. J.*, 10, 143–155.

Olds, B.P., Jerde, C.L., Renshaw, M.A., Li, Y., Evans, N.T., Turner, C.R. *et al.* (2016). Estimating species richness using environmental DNA. *Ecol. Evol.*, 6, 4214–4226.

Paillisson, J.-M. & Marion, L. (2001). Interaction between coot (*Fulica atra*) and waterlily (*Nymphaea alba*) in a lake: the indirect impact of foraging. *Aquat. Bot.*, 71, 209–216.

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. In: *Shallow Lakes'* 95. Springer, Dordrecht, pp. 241–255.

Piha, H., Luoto, M. & Merilä, J. (2007). Amphibian occurrence is influenced by current and historic landscape characteristics. *Ecol. Appl.*, 17, 2298–2309.

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. *Divers. Distrib.*, 15, 692–700.

Rannap, R., Lõhmus, A. & Briggs, L. (2009b). Restoring ponds for amphibians: a success story. *Hydrobiologia*, 634, 87–95.

Rees, H.C., Maddison, B.C., Middleditch, D.J., Patmore, J.R.M. & Gough, K.C. (2014). REVIEW: The detection of aquatic animal species using environmental DNA–a review of eDNA as a survey tool in ecology. *J. Appl. Ecol.*, 51, 1450–1459.

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 Res.*, 39, e145.

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. *et al.* (2012). The role of pond management for biodiversity conservation in an agricultural landscape. *Aquat. Conserv.*, 22, 626–638.

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. *Amphib-Reptil.*, 27, 309–324.

Stefanoudis, P.V., Sayer, C.D., Greaves, H.M., Davidson, T.A., Robson, H., Almeida, D. *et al.* (2017). Consequences of fish for cladoceran, water beetle and macrophyte communities in a farmland pond landscape: implications for conservation. *Fund. Appl. Limnol.*, 190, 141–156.

Unglaub, B., Steinfartz, S., Drechsler, A. & Schmidt, B.R. (2015). Linking habitat suitability to demography in a pond-breeding amphibian. *Front. Zool.*, 12, 9.

Ushio, M., Fukuda, H., Inoue, T., Makoto, K., Kishida, O., Sato, K. *et al.* (2017). Environmental DNA enables detection of terrestrial mammals from forest pond water. *Mol. Ecol. Resour.*, 17, e63-e75.

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P.F. *et al.* (2016). Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Mol. Ecol.*, 25, 929–942.

Vuorio, V., Heikkinen, R.K. & Tikkanen, O.-P. (2013). Breeding Success of the Threatened Great Crested Newt in Boreal Forest Ponds. *Ann. Zool. Fennici*, 50, 158–169.

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 Ethol.*, 13, 127–135.

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. *J. Comput. Biol.*, 7, 203–214.

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.

794 **Table 1** Summary of established and novel biotic and abiotic determinants of GCN occupancy.
795 Reported effects on GCN occupancy in the literature and hypothesised effects on eDNA-based GCN
796 occurrence are given for each determinant. Any determinants not reported in the literature are listed as
797 UNK. Direction of observed effects on eDNA-based GCN occupancy determined by each analysis
798 (GLMM assessing number of species in each vertebrate group, *N* = 532; co-occur analysis, *N* = 532;
799 GLMM combining abiotic and biotic factors *N* = 504; and GLMM assessing HSI score, *N* = 504) are
800 given. No, negative and positive effects are listed as 0, - and + respectively. For categorical variables
801 with more than one level, effect size and standard error (SE) are only given for levels reported in the
802 model summary. Test statistic is for LRT used and significant P-values (<0.05) are in bold. Variables
803 included for model selection but not retained in the final model are listed as NR. Co-occur analysis was
804 not applicable (NA) to abiotic factors.
805

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**

806

807 **Table 2** Summary of analyses testing for variation in eDNA-based vertebrate species richness across
808 ponds (*N* = 504), attributable to aquatic and terrestrial habitat. Results of the separate GLMM assessing
809 variation explained by HSI score are italicised. For categorical variables with more than one level,
810 effect size and standard error (SE) are only given for levels reported in the model summary. Test
811 statistic is for LRT used and significant P-values (<0.05) are in bold.
812

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***

813

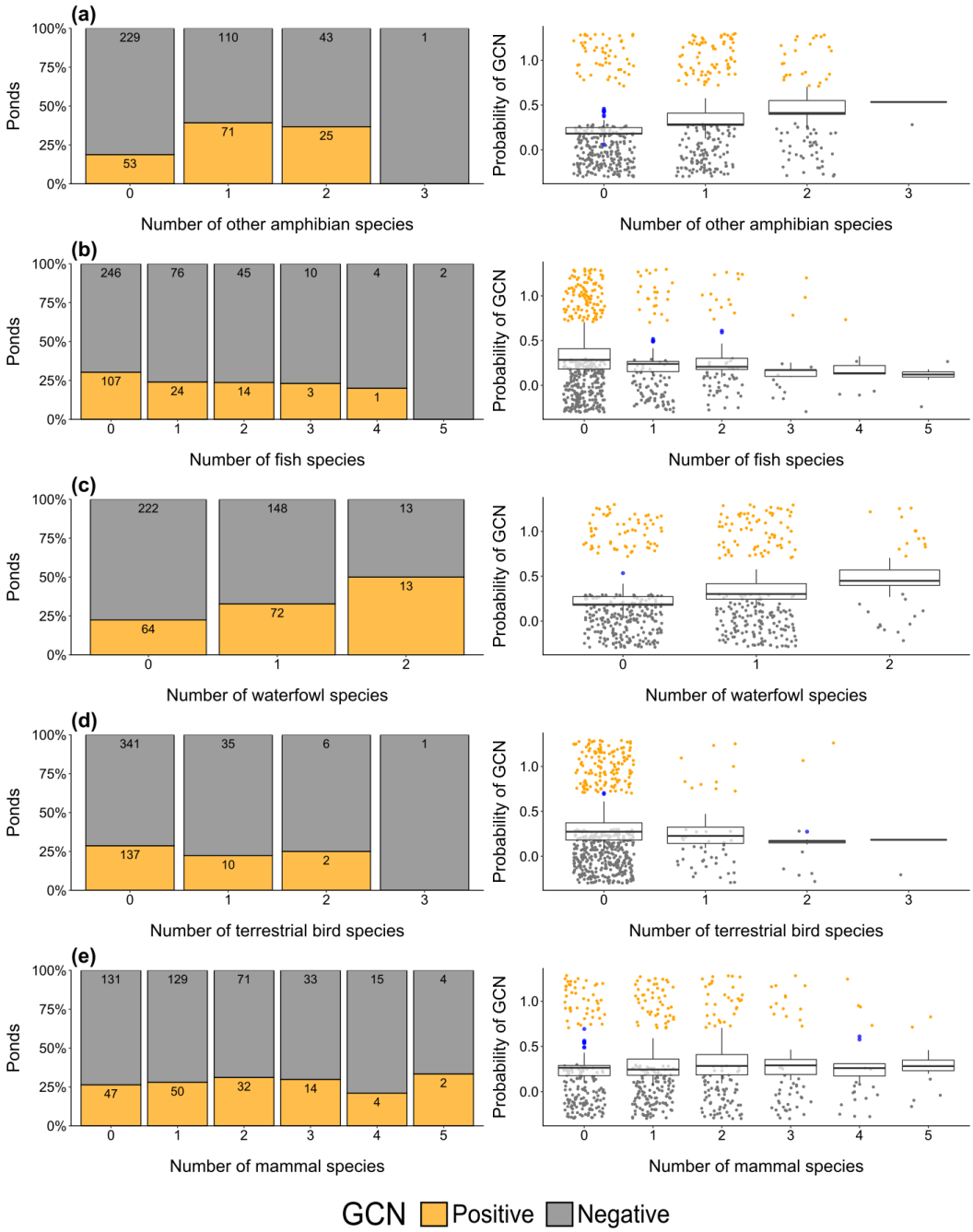

Figure 1 GCN presence (orange) and absence (grey) in relation to number of species from different vertebrate groups detected by eDNA (N = 532 ponds): (a) other amphibians, (b) fish, (c) waterfowl, (d) terrestrial birds, and (e) mammals. Observed proportion of ponds with and without GCN (left) is plotted 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.

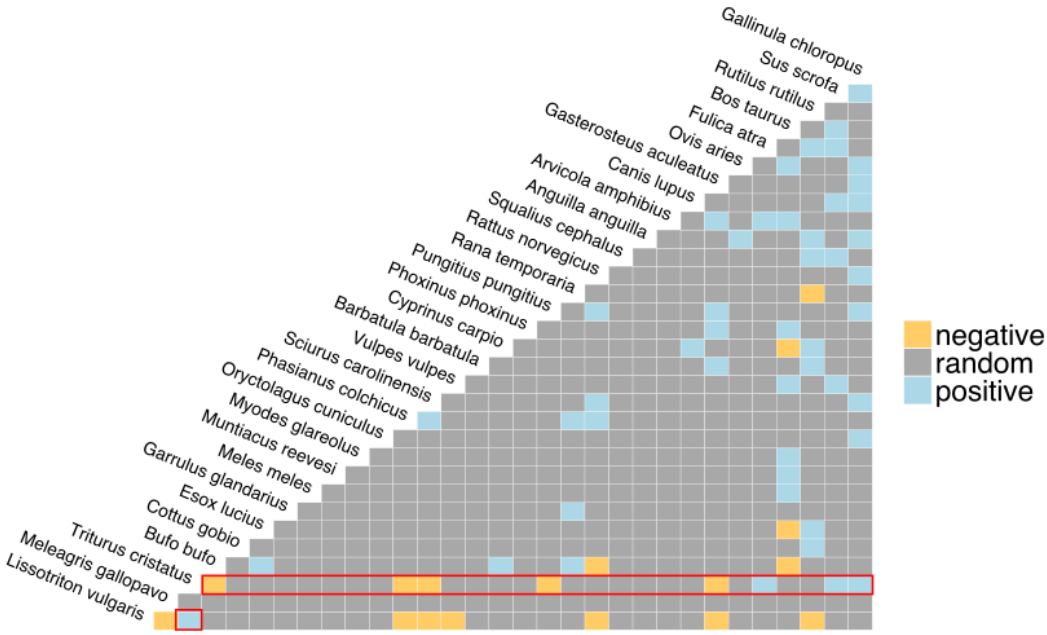

Figure 2 Heat map showing significant (P < 0.05) positive and negative species associations determined by the probabilistic co-occurrence model for the eDNA metabarcoding presence-absence data (N = 532 ponds). Species names are positioned to indicate the columns and rows that represent their pairwise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

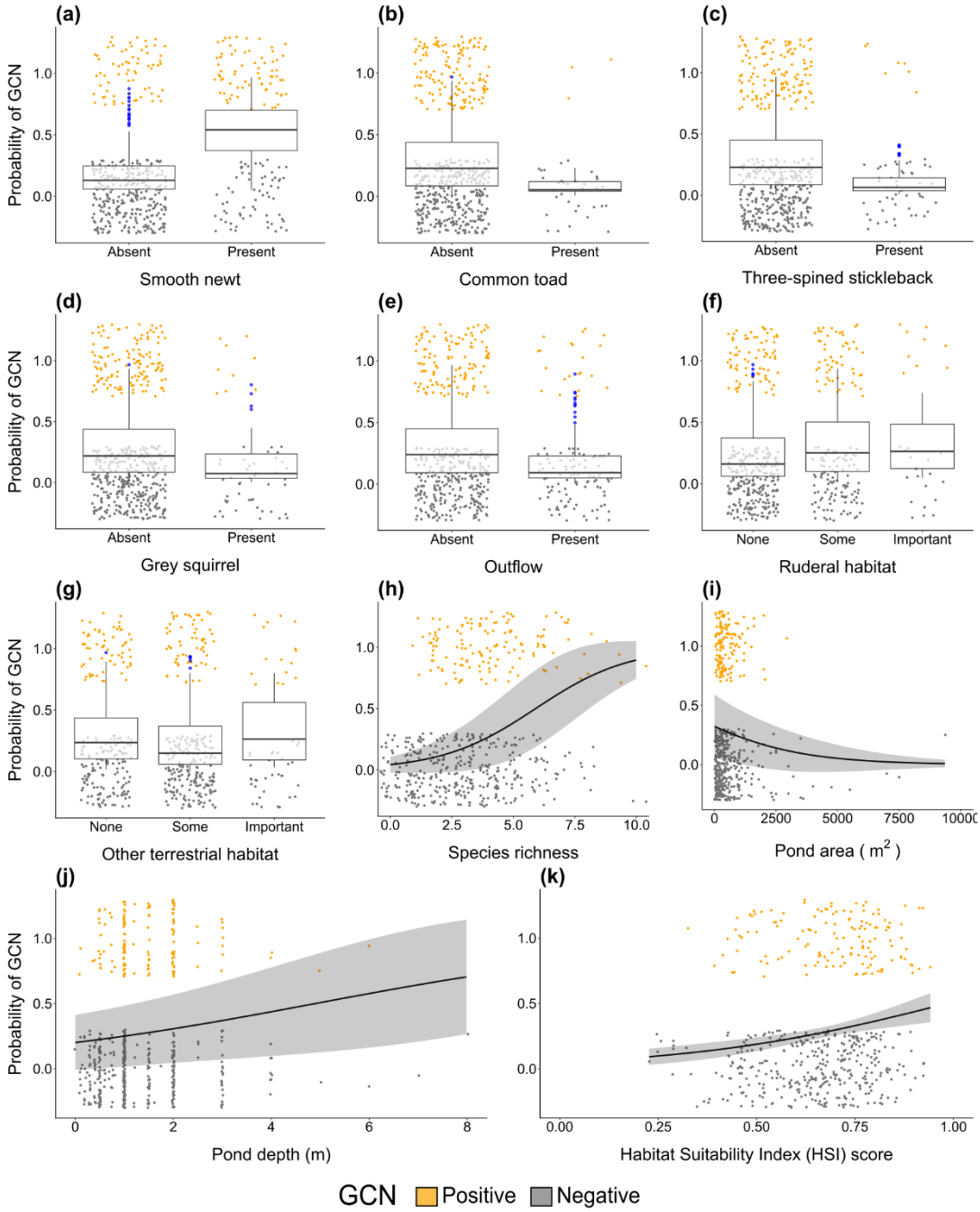

Figure 3 Biotic and abiotic determinants of GCN occurrence, as predicted by the binomial GLMMs (*N* = 504 ponds): (a) smooth newt occurrence, (b) common toad occurrence, (c) three-spined stickleback occurrence, (d) grey squirrel occurrence, (e) pond outflow, (f) ruderal habitat, (g) other good quality 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.

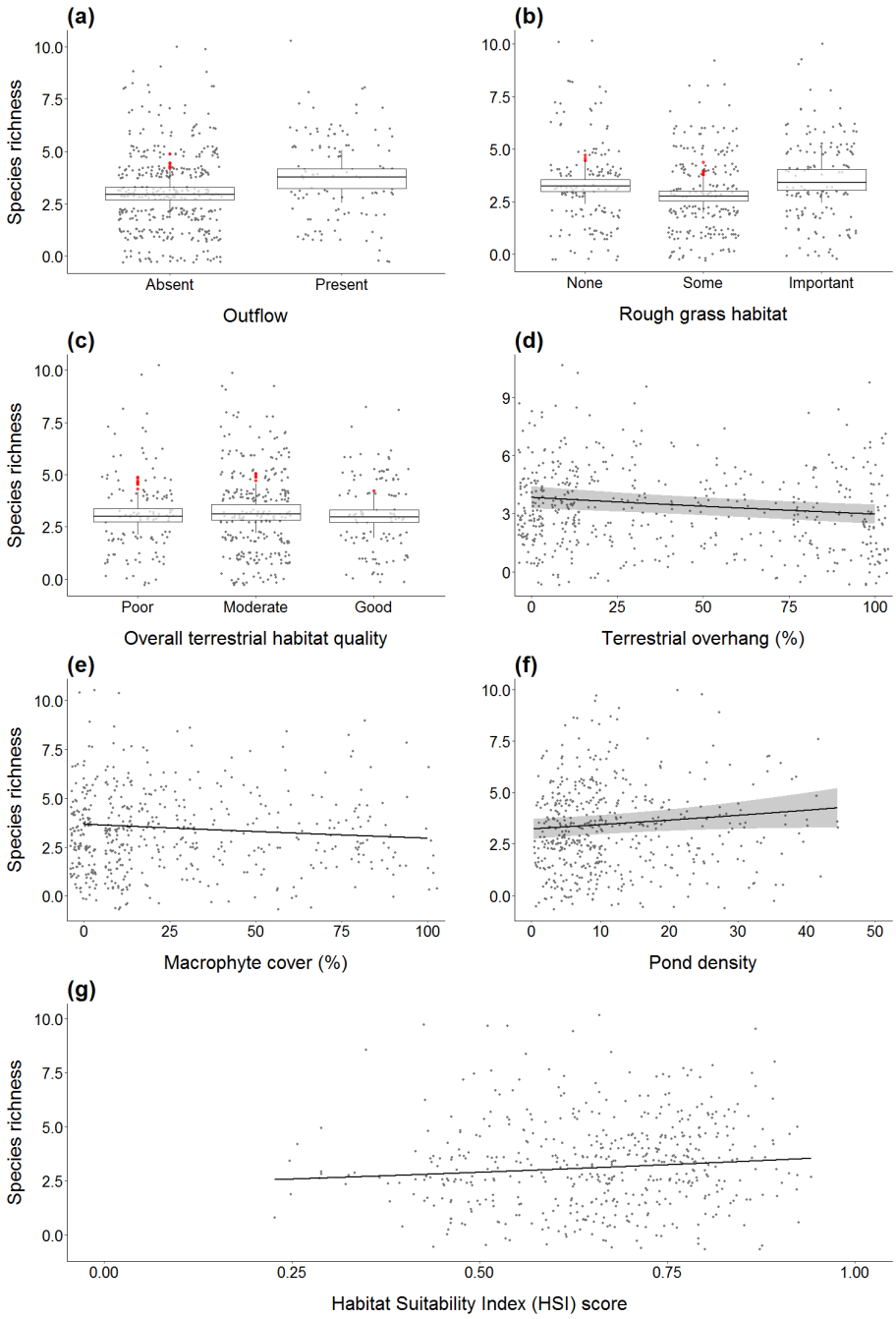

Figure 4 Abiotic and biotic determinants of vertebrate species richness, as predicted by the Poisson GLMMs (N = 504 ponds): (a) outflow, (b) rough grass habitat, (c) overall quality of terrestrial habitat, (d) percentage of terrestrial overhang, (e) percentage of macrophyte cover, (f) pond density, and (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.