


Please cite the Published Version

Shutt, Jack D , Nicholls, James A, Trivedi, Urmi H, Burgess, Malcolm D, Stone, Graham N, Hadfield, Jarrod D and Phillimore, Albert B (2020) Gradients in richness and turnover of a forest passerine's diet prior to breeding: a mixed model approach applied to faecal metabarcoding data. *Molecular Ecology*, 29 (6). pp. 1199-1213. ISSN 0962-1083

DOI: <https://doi.org/10.1111/mec.15394>

Publisher: Wiley

Version: Accepted Version

Downloaded from: <https://e-space.mmu.ac.uk/625378/>

Usage rights:  In Copyright

Additional Information: This is an Author Accepted Manuscript of a paper accepted for publication in *Molecular Ecology*, published by and copyright Wiley.

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

Title: Gradients in richness and turnover of a forest passerine's diet prior to breeding: a mixed model approach applied to faecal metabarcoding data

Running title: Pre-breeding dietary gradients of a passerine

Authors: Jack D Shutt^{1‡}, James A Nicholls^{1,2‡}, Urmi H Trivedi³, Malcolm D Burgess^{4,5}, Graham N Stone¹, Jarrod D Hadfield^{1#}, Albert B Phillimore^{1#*}

Affiliations:

1 Institute of Evolutionary Biology, The University of Edinburgh, The King's Buildings, Edinburgh, EH9 3FL, UK;

2 Current address: Australian National Insect Collection, CSIRO, Clunies Ross Street, Acton, ACT 2601, Australia;

3 Edinburgh Genomics, The University of Edinburgh, The King's Buildings, Edinburgh, EH9 3FL, UK;

4 Centre for Research in Animal Behaviour, University of Exeter, Exeter, EX4 4QG, UK;

5 RSPB Centre for Conservation Science, The Lodge, Sandy, Bedfordshire, SG19 2DL, UK

‡ and # Contributed equally.

Correspondence: Albert Phillimore albert.phillimore@ed.ac.uk

Orcid ID's: JDS 0000-0002-4146-8748, MDB 0000-0003-1288-1231, ABP 0000-0002-6553-1553

24 **Abstract**

25 Little is known about the dietary richness and variation of generalist insectivorous
26 species, including birds, due primarily to difficulties in prey identification. Using
27 faecal metabarcoding we provide the most comprehensive analysis of a passerine's
28 diet to date, identifying the relative magnitudes of biogeographic, habitat and temporal
29 trends in the richness and turnover in diet of *Cyanistes caeruleus* (blue tit) along a 39-
30 site, 2° latitudinal transect in Scotland. Faecal samples were collected in 2014-15 from
31 adult birds roosting in nestboxes prior to nest building. DNA was extracted from 793
32 samples and we amplified COI and 16S minibarcodes. We identified 432 molecular
33 operational taxonomic units (MOTUs) that correspond to putative dietary items. Most
34 dietary items were rare, with Lepidoptera being the most abundant and taxon-rich prey
35 order. We present a statistical approach for estimation of gradients and inter-sample
36 variation in taxonomic richness and turnover using a generalised linear mixed model.
37 We discuss the merits of this approach over existing tools and present methods for
38 model-based estimation of repeatability, taxon richness and Jaccard indices. We find
39 that dietary richness increases significantly as spring advances, but changes little with
40 elevation, latitude or local tree composition. In comparison, dietary composition
41 exhibits significant turnover along temporal and spatial gradients and among sites. Our
42 study shows the promise of faecal metabarcoding for inferring the macroecology of
43 food webs, but we also highlight the challenge posed by contamination and make
44 recommendations of laboratory and statistical practices to minimise its impact on
45 inference.

46

47

48 **Keywords**

49 Beta diversity, avian/bird, Jaccard, insectivore, prey, repeatability, blue tit, *Cyanistes*
50 *caeruleus*

51

52

53 **Introduction**

54

55 Insectivorous passerine birds in temperate environments tend to be dietary generalists
56 feeding on a broad range of invertebrate taxa (Betts, 1955; Cholewa & Wesolowski,
57 2011). There is potential for the diet of such generalists to vary over geographic
58 gradients, among habitats and seasonally within a year. Such dietary variability within
59 generalist species is poorly understood and could have profound ecological
60 consequences. Spatial variation in resource availability has implications for
61 geographic patterns in population density, breeding productivity and the degree to
62 which local adaptation in resource use may evolve. Seasonal variation in resource
63 consumption has implications for the optimal scheduling of life history events, such
64 as reproduction (Charmantier et al., 2008; Durant et al., 2005) and seasonal movements
65 (Thorup et al., 2017).

66

67 Spatiotemporal trends in diet will arise from a combination of underlying trends in
68 invertebrate resource availability and the prey preferences of the consumer. Species
69 richness – or α -diversity – of temperate invertebrate taxa generally decreases with
70 increasing latitude (Baselga, 2008) and peaks at mid-elevations (Beck et al., 2017) in
71 the summer months (Thomsen et al., 2016). Within forests, invertebrate richness can

72 vary among tree taxa by more than an order of magnitude, and in the UK has been
73 found to be highest on *Salix*, *Quercus* and *Betula* (Kennedy & Southwood, 1984; Shutt,
74 Burgess, & Phillimore, 2019). In addition to changes in species richness, species
75 composition may change from one community to the next, which is quantified as β -
76 diversity (Baselga, 2010; Whittaker, 1972). While there is evidence that forest
77 invertebrate communities show turnover over biogeographic gradients (Novotny &
78 Weiblen, 2005) and among host tree taxa (Murakami, Ichie, & Hirao, 2008), the
79 relative magnitude of turnover along different gradients has received scant attention
80 (Novotny & Weiblen, 2005). Whether diet mirrors these gradients in resource
81 availability will largely depend on how much prey selection by the consumer departs
82 from random.

83

84 Forest-dwelling hole-nesting insectivorous birds, such as blue tits (*Cyanistes*
85 *caeruleus*), have been subject to decades of intensive study (C. M. Perrins, 1979).
86 While the diet of nestlings has proven relatively straightforward to quantify, either via
87 videos/cameras at the nest (Samplonius, Kappers, Brands, & Both, 2016), or neck
88 collars on nestlings (Burger et al., 2012), much less is known about the diet of adults
89 (but see Cholewa & Wesołowski, 2011; J. A. Gibb, 1954). The paucity of information
90 about adult diet arises because these birds often forage high in trees on small prey
91 items. To date most of our taxonomic information on adult tit diet has been derived
92 from dissections of the gizzard and gut contents of euthanised birds (Betts, 1955), a
93 method that precludes the identification of soft-bodied dietary items, has relatively
94 poor taxonomic resolution (e.g. order or family level) and is destructive. These studies
95 reveal that tits consume various insects (including Lepidoptera, Hemiptera, Diptera,

96 Coleoptera, Hymenoptera) and spiders, as well as some plant matter in winter (Betts,
97 1955; Cholewa & Wesołowski, 2011; Cramp & Perrins, 1993).

98

99 The advent of next-generation sequencing and faecal DNA metabarcoding now
100 provides a non-destructive means of obtaining diet information at a fine taxonomic
101 resolution (Pompanon et al., 2012; Symondson, 2002; Taberlet, Coissac, Pompanon,
102 Brochmann, & Willerslev, 2012). Where invertebrates comprise a large proportion of
103 the diet, DNA barcodes from the rapidly evolving cytochrome oxidase I (COI)
104 mitochondrial gene have become the standard and allow identification to species-level
105 in many cases (Kress, García-Robledo, Uriarte, & Erickson, 2015). To date, most
106 published faecal metabarcoding studies have examined variation in mammalian diet
107 (Bohmann et al., 2011; Clare, Symondson, Broders, et al., 2014; Clare, Symondson,
108 & Fenton, 2014; Razgour et al., 2011). In comparison to mammals in general, and bats
109 in particular, application of faecal metabarcoding for inference of the diet of avian
110 insectivores is a small but rapidly growing field. Progress has been hampered by the
111 challenge of extracting and successfully amplifying dietary DNA from avian faeces
112 (Jedlicka, Sharma, & Almeida, 2013; Vo & Jedlicka, 2014). As such, avian faecal
113 metabarcoding studies have sampled small numbers of individuals and/or locations
114 (Table 1) and the latter limitation has precluded detailed analysis of the drivers of
115 spatial or temporal variation in the diet of avian insectivores (for an exception see
116 Sullins et al., 2018).

117

118 To date the statistical tools employed by the nascent metabarcoding field have largely
119 borrowed from community ecology. In some studies the objective has been to describe

120 the diet composition of a taxon such that statistical analysis may be unnecessary (De
121 Barba et al., 2014). Metabarcoding studies that focus on patterns in taxon richness
122 commonly apply a two-step analysis, first using rarefaction to quantify diversity at a
123 focal sampling level and then using a statistical model to examine variation in taxon
124 richness among samples (Quéméré et al., 2013). Studies interested in how taxonomic
125 composition varies among samples have tended to rely on pairwise metrics, such as
126 the Jaccard index, and non-parametric methods, such as PERMANOVA and the
127 Mantel test (Alberdi, Aizpurua, Gilbert, & Bohmann, 2018; Mata et al., 2019;
128 Trevelline, Nuttle, Hoenig, et al., 2018). Generalised linear mixed models (GLMMs)
129 and their extensions provide a method for including structure in the data collection and
130 multiple predictors into an analysis (Warton et al., 2015), but few studies have utilised
131 them in diet metabarcoding to date (for exceptions see Mata et al., 2019; Nichols,
132 Åkesson, & Kjellander, 2016).

133

134 Here we employ faecal metabarcoding using COI minibarcodes to infer the diet of an
135 insectivorous woodland passerine, the blue tit, in early spring along a 220 km transect
136 in Scotland (Appendix 1 Fig. S1). We have three main aims: (i) to quantify dietary
137 taxon richness and composition at the molecular operational taxonomic unit (MOTU)
138 level; (ii) to quantify the magnitude of changes in both measures along gradients of
139 time (day of year), latitude, elevation and tree taxon composition; and (iii) to quantify
140 gradients in the contributions that six key invertebrate orders (Araneae, Coleoptera,
141 Diptera, Hemiptera, Hymenoptera, Lepidoptera) make to diet. We show that by
142 applying a GLMM to presence/absence data it is possible to estimate changes in taxon
143 richness and turnover among points and along gradients. We also demonstrate how

144 this mixed model approach can be used to estimate repeatability and control for some
145 types of systematic contamination.

146

147

148 **Material and Methods**

149

150 **Field data collection**

151 Fieldwork was conducted during the springs of 2014 and 2015 at 39 predominantly
152 deciduous woodland sites that together comprise a 220km latitudinal transect in
153 Scotland (Shutt, Bolton, Benedicto Cabello, Burgess, & Phillimore, 2018). At each
154 site there were six Schwegler 1B 26mm-hole nestboxes distributed at approximately
155 40m intervals. From mid-March the base of each nestbox was lined with greaseproof
156 paper – with the aim of slowing DNA degradation (Oehm, Juen, Nagiller, Neuhauser,
157 & Traugott, 2011) – which was replaced when damaged or heavily soiled, and
158 removed at the onset of nest building or once a bird had attempted removal. Each
159 nestbox was inspected on alternate days and faeces on the greaseproof paper were
160 removed with sterilised tweezers (after use they were wiped with lab tissue and
161 flamed), with up to a maximum of three faeces collected in a 2mL Eppendorf tube
162 containing pure ethanol. The total number of faeces in a nestbox was recorded
163 (excluding 129 samples from early 2014). Samples were stored at -18°C within a day
164 of collection and transferred to a -20°C freezer at the end of each spring. Faecal
165 samples were collected from 35 of the 39 sites from 19 March in 2014 and 18 March
166 in 2015 until nest building, giving a median sampling range of 20 days per site in 2014
167 and 24 days in 2015 (Appendix 1 Table S1).

168

169 Latitude (site range 55.98 – 57.88°N) and elevation (10 – 433m) were obtained for
170 each nestbox (Shutt et al., 2018). Site-level habitat metrics were derived from surveys
171 of numbers of trees of different genera belonging to three size classes (based on girth
172 at breast height) within 15m radius of each nestbox, as described in Shutt et al. (2018).
173 The site-level habitat variables we considered were total foliage, tree diversity
174 (Simpson's index), the amount of oak foliage and the amount of birch foliage (Shutt
175 et al., 2018).

176

177 **Molecular protocol**

178 We balanced sampling across nestboxes and dates as far as possible by imposing an
179 upper limit of 10 samples per nestbox per year and where this maximum was exceeded
180 we subsampled such that we maximised the range of dates per nestbox. If multiple
181 faeces ($n = 2 - 3$) were present within a sample tube, part of each individual scat was
182 used for the DNA extraction with the aim of sampling a broader range of diet. This
183 protocol resulted in processing of 793 of a total of 959 faecal samples.

184

185 Thirty samples were processed in duplicate to allow us to estimate technical
186 repeatability. The selected samples were evenly distributed throughout the sampling
187 period, including samples from multiple sampling locations in both 2014 and 2015.
188 The faeces for each of the 30 duplicated samples were evenly divided into two and
189 DNA extractions were performed on each subsample; although each subsample
190 contained sections from along the entire length of the original faeces, the faeces was
191 not completely homogenised before subsampling. Each duplicate extraction was

192 subsequently treated as an independent sample for all downstream processes. All
193 aspects of the laboratory protocol (DNA extraction, PCR amplifications, PCR clean-
194 up, sequencing on a MiSeq run) were performed at different times using different
195 aliquots of reagents for each replicate within a pair of subsamples. In addition we
196 included 24 controls (including extraction negatives, PCR negatives and *Dryocosmus*
197 *israeli* as a non-native invertebrate PCR positive).

198

199 DNA was extracted from faecal samples using the QIAamp DNA Stool Mini kit,
200 following the protocol for pathogen detection with a few custom modifications
201 designed to improve DNA yields (see online protocol for details;
202 [dx.doi.org/10.17504/protocols.io.ve6e3he](https://doi.org/10.17504/protocols.io.ve6e3he)). Three loci were targeted for amplification
203 through PCR - the standard animal barcoding gene (COI), a secondary barcoding gene
204 to detect invertebrate prey DNA and confirm the faecal sample originated from a blue
205 tit and no other hole-roosting or -nesting passerine (16S rRNA), and a standard plant
206 barcoding gene (rbcL) (see online protocol for further details;
207 [dx.doi.org/10.17504/protocols.io.2jdgc6](https://doi.org/10.17504/protocols.io.2jdgc6)). Given that DNA from dietary items is
208 expected to be very degraded, the primers used amplified a small ‘minibarcodes’ region
209 of each gene (184-220 base pairs). Invertebrate primer sets were validated to ensure
210 that they would amplify DNA from the expected range of invertebrate taxa (two orders
211 of arachnids, isopods, nine insect orders).

212

213 We followed a two-stage PCR process, firstly to amplify the target regions, then
214 secondly to add indexed Illumina adaptors to the amplicons from each sample.
215 Amplicons were multiplexed into three pools, each containing between 273 and 276

216 samples (inclusive of 30 replicate samples) and 8 controls (3x PCR positives, 3x PCR
217 negatives and 2x extraction negatives; a total of 24 controls across the whole
218 experiment). Each pool was sequenced on an Illumina MiSeq, using 150 bp paired-
219 end reads.

220

221 **Bioinformatics processing**

222 Sequencing reads were initially de-multiplexed into sets corresponding to individual
223 faecal samples using the index combinations present within the adaptor sequences
224 using bcl2fastq (version v2.17.1.14). Reads were then de-multiplexed using fastq-
225 multx from ea-utils (version 1.1.2-537) with parameter ‘-m 2’ into sets corresponding
226 to each locus using the locus-specific primer sequences present at the beginning of
227 each read. Adaptor sequences, primer sequences and poor quality base calls were then
228 removed using cutadapt (version 1.8.3) with parameters: ‘-m 50’, ‘-q 30’, ‘-f fastq’,
229 leaving only sequence corresponding to the targeted gene regions. Subsequent
230 processing of the sequences applied the UPARSE pipeline (initially developed for 16S
231 metabarcoding of bacteria, (Edgar, 2013)) to data for each locus separately.

232

233 The first step in the bioinformatics pipeline was to merge the paired reads derived from
234 either end of the sequenced fragment. This process was successful for all COI and rbcL
235 reads and many 16S reads; 16S reads derived from avian DNA did not overlap, but
236 comparison with known blue tit 16S sequences indicated that these reads could be
237 combined by adding four “N”s between the forward and reverse reads to produce a
238 composite sequence of the correct length (hereafter referred to as fused reads). Reads
239 were then filtered to ensure that within a locus they were all of the same length; this

240 process removed possible pseudogenes incorporating insertions/deletions from the
241 coding COI and rbcL data. The rbcL data were not used for subsequent analyses in this
242 study, and 16S data were only used to confirm the faeces were derived from blue tits.
243 The set of filtered COI sequences was then used for two purposes. Firstly, the set of
244 unique sequences present within the full data set derived from all samples was
245 determined, with counts made of their frequencies. Unique sequences represented by
246 only a single read were removed as they most likely represent sequencing errors. The
247 unique sequences were then clustered into molecular operational taxonomic units
248 (MOTUs), grouping sequences together that had an identity of 98% or more. The most
249 frequently occurring sequence within each MOTU was designated as the reference
250 sequence for that MOTU. The second use of the filtered reads involved mapping them
251 back to this reference set of MOTU sequences on a sample by sample basis, allowing
252 a mismatch of up to 2% between filtered reads and a reference sequence, to provide a
253 more accurate assessment of the frequency of each MOTU within each faecal sample.
254 The taxonomic identity of MOTUs was determined using a BLAST search of the
255 reference set of MOTU sequences against public databases (GenBank and BOLD).

256

257 **Quality control and MOTU refinement**

258 Our analysis plan from this point on was pre-registered (osf.io/xgvm8). Some aspects
259 of our methods deviate from what was outlined in the pre-registration (see table S2 in
260 appendix 1 for an explanation of the motivation for these departures). We tested
261 whether samples were from blue tits by verifying the presence of blue tit fused 16S
262 sequences. The highest number of blue tit 16S reads from the 24 control samples was
263 58 and as a precaution all faecal samples that yielded fewer than 100 blue tit 16S

reads were excluded from further analyses as they were not conclusively confirmed to be blue tit faeces ($n = 9$). Of the remaining avian faecal samples, blue tit was the commonest of the fused 16S MOTU in all but one sample, but this sample still had sufficient ($n = 1465$) blue tit reads to confirm its identity. No other avian DNA was present in any sample.

COI reads were checked from control samples to confirm the presence of positive control species and provide a baseline for background contamination. All nine PCR positive control samples contained MOTUs attributable to *Dryocosmus israeli* (range of reads = 7796 - 19115) and no more than 16 reads of any other MOTU identified as belonging to the Metazoan kingdom. Eight out of nine PCR negative controls contained no more than 19 reads of any MOTU. The ninth was highly contaminated and contained 6798 reads arising from more than 20 MOTUs. Therefore, we checked for contamination along rows or columns within plates by estimating Spearman's correlations in the number of MOTU reads between samples in neighbouring cells in the same PCR column or row. The row containing the contaminated negative sample was found to have a substantially higher mean level of within row correlation ($r = 0.37$) than other row and column correlations (mean $r = 0.04$). This was considered to be most likely a systematic contamination event and this row ($n = 11$ focal samples + 1 negative control) was excluded from all analyses. In addition, closer inspection of the contaminated plate revealed two wells (both faecal samples) in the neighbouring row to the contamination event containing very similar MOTUs with the contaminated row and these were also removed from further analysis. Of the six extraction negative controls, four contained no MOTU at a higher read frequency than 3. The remaining

288 two contained contamination (maximum reads = 10037 and 1611) but on further
289 inspection there was no evidence for this being systematic. As there were few cases
290 where a control (positive or negative) had > 20 reads for any non-target MOTU, we
291 adopted 20 reads as the cut-off for identifying MOTU presence.

292

293 The above steps reduced the number of samples from 847 to 824 (772 focal) containing
294 2524 MOTUs. All MOTUs with fewer than 20 reads in any single sample were
295 removed as possible false positives (remaining n = 1432 MOTUs). All MOTUs
296 without any BLAST match, or identified as environmental contamination, were
297 removed (remaining n = 1323). Then, a full taxonomy was obtained for each remaining
298 MOTU and taxonomic reduction of the dataset began to eliminate non-prey items.
299 Firstly, only MOTUs belonging to the Metazoan kingdom were considered possible
300 prey items (remaining n = 1078). Then, all MOTUs not belonging to the phyla
301 Annelida, Arthropoda and Mollusca were discarded (remaining n = 1005). Finally, all
302 mites in the dataset of orders Astigmata (11), Mesostigmata (56), Oribatida (1),
303 Siphonoptera (2) and Trombidiformes (24) were removed, as they were likely to be
304 ectoparasites rather than actively foraged prey (remaining n = 911). For the MOTU
305 identification we required that the percentage match was at least 90% (remaining n =
306 785). Taxa identified to an identification match of 90% or more are considered correct
307 to a minimum of order level, and this is the level that is important to the analyses in
308 this study. Several MOTUs identified as '*Arachnida* sp' were removed on finding that
309 these MOTUs were most closely matched to fungi (remaining n = 778). All
310 *Dryocosmus* (positive control) and waxworm (*Galleria mellonella* – from a feeding
311 experiment in 2014 that provided 10 waxworms in a plastic cup adjacent to two

312 nestboxes per site) MOTUs were removed (remaining $n = 757$). Then, all remaining
313 MOTUs belonging to the same best-hit taxon were merged (remaining $n = 432$).
314 Finally, due to the importance of Lepidoptera to tit diet we assessed the biological
315 plausibility of *Lepidoptera* identifications, which was possible due to comprehensive
316 UK occurrence data for this order (Sterling & Parsons, 2012; Waring & Townsend,
317 2017). Nineteen of 131 Lepidopteran MOTUs assigned species names were reassigned
318 to a British species when this species was within a 1% match of a geographically
319 implausible top hit. We assigned species status to taxa with a 99% or greater identity
320 match with the BLAST hit and a histogram of identity matches is provided (Fig S2).

321

322 **Statistical analyses**

323 Analyses focussed on the presence/absence of MOTUs in a sample, as read numbers
324 are not considered a reliable measure of the amount of a MOTU in a sample due to
325 biases in primer binding and amplification (Clare, 2014; Yu et al., 2012). Control
326 samples were excluded from analyses. DNA within a sample was often derived from
327 multiple faeces, and the effect of this on MOTU presence was controlled for by
328 including number of faeces as a four-level categorical fixed effect (1, 2, 3, unknown).

329

330 To examine geographic, habitat and temporal variation in blue tit diet (Shutt, Nicholls,
331 et al., 2019), we included the presence or absence of each MOTU in each sample as
332 the response variable in a Bayesian generalized linear mixed model (GLMM) with a
333 probit error structure (Hadfield, 2010). This analysis excluded the replicate samples
334 (for reasons discussed in Appendix 2). The effects of year and number of faeces in the
335 sample (treated as categorical) and the effects of ordinal date, latitude, elevation, total

336 foliage, birch foliage, oak foliage and tree diversity (treated as continuous) were
337 treated as fixed. These fixed effects quantify trends in dietary richness. After
338 accounting for these trends, variation in richness amongst sites, nest-boxes, days
339 within year (categorical) and faecal samples were modelled by fitting each term as
340 random. MOTU effects were fitted as random in order to capture differences amongst
341 MOTUs in their overall prevalence. Variation in the prevalence of individual MOTUs
342 amongst sites, nest-boxes, days within years (categorical) and faecal samples was
343 modelled by interacting each term with MOTU. In the core model we also allowed the
344 prevalence of individual MOTUs to vary with ordinal date, latitude and elevation
345 effects, again by interacting each term with MOTU to form random regressions. The
346 three slope terms were allowed to covary with each other and the main MOTU effect
347 (the intercept). We also include plate by MOTU random interaction term to control for
348 any plate-wide contamination by particular MOTUs present. To estimate and correct
349 for any tendency for contamination of rows or columns within a plate we ran an
350 additional model with row (within plate) and column (within plate) interacted with
351 MOTU as random terms and this is the main model that we present in the results.

352

353 In addition to the core model, we also ran four additional models, each of which
354 allowed the prevalence of individual MOTUs to vary across one of the four habitat
355 variables. The additional random slope terms were allowed to covary with the original
356 three slope terms and the intercept. However, because of the length of time that the
357 core model took to run (three months) we excluded the day within year term and its
358 interaction with MOTU. The importance of these effects are minor relative to other
359 terms in the model (day in year variance = 0.003, day in year:MOTU variance = 0.036,

360 table S4A) and the interaction in particular contributed a lot to computation time
361 because with 91 days and 432 MOTUS there are nearly 40,000 effects. All models
362 were run for 260,000 iterations, with the first 60,000 removed as burn-in and thinning
363 every 100. These models took two months to run on an iMac 10.13.6 with 3.4 Ghz
364 Intel core i7, 16GB RAM and 4 cores.

365

366 To examine trends in the presence/absence of prey orders in blue tit diet, the dataset
367 was reduced down to presence/absence of the six most common orders (Araneae,
368 Coleoptera, Diptera, Hemiptera, Hymenoptera and Lepidoptera), termed ‘focal
369 orders’, which together comprise over 91% of all prey taxa identified. A similar
370 GLMM to that described above was then employed, but with focal order and date,
371 latitude, elevation and tree diversity individually and interacted with focal order as
372 fixed effects. Site, nest-box, day and faecal samples were fitted as random main effects
373 and as random interactions with focal order. These models were run for 195,000
374 iterations, with the first 45,000 removed as burn-in and thinning every 75.

375

376 To assess the repeatability of the approach we used a similar analysis to that described
377 above with the presence/absence of each MOTU as a response for the faecal samples
378 for which extraction, PCR and metabarcoding had been replicated (29 samples x 432
379 MOTUs). Fixed effects were year and the number of faeces in the sample, both as
380 factors, with random terms limited to MOTU, faecal sample ID, faecal sample ID by
381 MOTU interaction, extraction sample ID and residual. This model was run for 13
382 million generations with the first 3 million removed as burn-in and thinning every
383 5000.

384

385 All numeric predictor variables in all analyses were scaled to have a mean of 0 and a
386 variance of 1 to provide direct comparability of results. We used parameter expanded
387 priors for the variances such that the marginal priors on all variances followed a scaled
388 (1000) $F_{1,1}$ distribution. Traces of posteriors were visually inspected to check for
389 convergence and adequate sampling. For the main model, the effective sample sizes
390 (ESS) were a bit low for some variances (< 500), but in all cases the ESS were adequate
391 to provide a reliable point estimate (> 100) even if in some instances the accuracy of
392 the credible intervals is poor. As a test of model adequacy we conducted posterior
393 predictive simulations to assess whether key features of the data were captured (Fig.
394 S3). We opted to use an MCMCglmm approach rather than much faster numerical
395 integration approaches, such as lme4 (Bates, Maechler, & Bolker, 2012) or glmmTMB
396 (Brooks et al., 2017), because posterior predictions revealed that parameter estimates
397 from MCMCglmm provided an accurate description of the data, whereas those from
398 lme4 and glmmTMB were highly inconsistent (Appendix 2). Additional simulations
399 confirmed that parameter estimates from lme4 were highly biased, most likely because
400 with rare-outcome data the approximations used for integrating over the random
401 effects break down.

402

403 In order to get a quantitative understanding of how α and β diversity change across
404 different levels of biological organisation (e.g., nestbox or site) and as a function of
405 continuous biogeographic variables (e.g., elevation or tree diversity) we develop a
406 framework for *focussing* repeatability metrics at the appropriate biological level (see
407 Appendix 2). The two-way dichotomy into between-group and within-group that

408 forms the basis of standard repeatability calculations (see Nakagawa & Schielzeth,
409 2010 for a review) can be seen as a special case. The quantities required for these
410 calculations also appear in many indices developed by ecologists to quantify similarity
411 in community structure. Given this, we show how such indices can also be derived
412 directly from a GLMM which has the advantages that credible intervals can easily be
413 computed, incomplete sampling is naturally dealt with (Chao, Chazdon, Colwell, &
414 Shen, 2006) and changes in the indices as a function of differences in a continuous
415 variable (such as latitude) can be handled. The main disadvantage of the approach is
416 that between-species correlation structures may typically be richer than what a fitted
417 GLMM assumes, such that variation in community structure may be greater than the
418 model allows. However, posterior predictive checking allows model inadequacies to
419 be detected, and richer correlation structures are available, for example through a
420 phylogeny (Hadfield & Nakagawa, 2010) or through factor analysis (Niku, Hui,
421 Taskinen, & Warton, 2019; Warton et al., 2015).

422

423 In Appendix 2 we also present methods for using model outputs to generate
424 expectations for the taxon richness of a faecal sample and Jaccard index (often used in
425 studies of β -diversity) that quantifies the similarity of faecal samples. This allows us
426 to relate model coefficients back to effect sizes that are more often used in community
427 ecology. However, as the Jaccard index captures both turnover and community
428 nestedness (Baselga, 2010), in the results we mainly use repeatability to quantify
429 turnover.

430

431

432 **Results**

433

434 **Read quality**

435 The three MiSeq runs combined generated 34.04 million raw paired-end reads, of
436 which 9.8 million were classified as COI amplicons after de-multiplexing based on the
437 primer sequences. Amplicons for 16S and rbcL were also generated, but our diet
438 analysis focuses only on COI. 8.9 million merged sequences passed all the quality
439 filters. Out of these, 8.7 million sequences were retained after alignment against the
440 reference OTU sequences.

441

442 **Diet Composition**

443 After identifying samples that tested positive for blue tit 16S DNA, excluding non-
444 prey taxa and collapsing similar sequences, we identified 432 prey MOTUs across
445 772 faecal samples. Of these MOTUs, 57% could be matched to candidate species on
446 the basis of > 99% sequence identity and a voucher/reference specimen identified to
447 species level. A further 4% were >99% matched and therefore identifiable to species
448 level, but lacking a reference initially identified to species level. The remainder of
449 MOTUs are not identifiable to species level but are diagnostically distinct dietary
450 items at minimum within the order identified by the best hit. (Appendix 1 Fig. S2,
451 Table S3). The mean number of MOTUs per sample was 5.06, with mode = 3,
452 median = 5 and range = 0 - 20. The MOTU abundance distribution was highly right-
453 skewed, with 42.4% recorded in only one sample and 74.3% recorded in five or
454 fewer samples (Fig. 1A).

455

Only 15 MOTUs were recorded in more than 50 samples (five Lepidoptera, four Hemiptera, three Diptera and one each of Collembola, Coleoptera and Hymenoptera). Eleven of these MOTUs were identified to species level, with *Argyresthia goedartella* (Lepidoptera: Yponomeutidae) most common (34.6% of samples, Fig. 1A inset). Most of these species are associated with resources available early in spring (Table S4), such as catkins on birch (*Betula pendula/pubescens*) or alder (*Alnus glutinosa*) or buds of birch or sycamore (*Acer pseudoplatanus*). We also found winter moth (*Operophtera brumata*) in 27 samples (3.5%), the larvae of which comprise a major component of nestling diet later in the spring but were not known to occur in the diet in early spring.

Eighteen invertebrate orders were encountered in at least one sample, with Insecta contributing 86.1% of MOTUs. Within insects, MOTUs matched to the order Lepidoptera were the most commonly recorded (present in 73.6% of samples, Fig. 1B) and taxon-rich (131 taxa, Fig. 1C). Other commonly recorded orders were Hemiptera, Diptera, Hymenoptera, Coleoptera, Araneae and Collembola.

Technical Repeatability

The value of faecal metabarcoding as a tool to infer diet depends on how reliable it proves to be and a key measure of this is repeatability. Our protocol included 30 paired replicate extractions from a different portion of the same faecal sample (although note that the sample was not homogenised prior to extraction), 29 of which remained after quality control and which we used to estimate technical repeatability (Appendix 1 Table S5G). The repeatability estimate is highly sensitive to the quantity being measured (measurand), the definition of within and between group, the reference

population and whether it is considered on the latent or data scale (Appendix 2). The technical repeatability of a MOTU within a faeces (with only faeces and faeces:MOTU contributing to the between-group variance) had a posterior mode of 0.305 (95% credible interval = 0.223 – 0.408) on the data (0,1) scale and 0.783 (0.712 – 0.845) on the latent (threshold) scale. Variation in MOTU richness at the sample level was reasonable (0.325 (0.118 – 0.770)) but the richness of samples within faeces are not strongly correlated and so the technical repeatability of richness for a faeces is low (Data; 0.003 (0 – 0.714), Latent; 0.003 (0 – 0.676)). However, the credible intervals are large, and the main analysis (see below) shows non-zero correlations between the richness of faeces from the same nestbox suggesting the true technical repeatability of richness must be non-zero.

Dietary MOTU richness

We used a generalized linear mixed model (GLMM) with a binary (threshold) response to examine the predictors of MOTU presence. From the main effects we can gain insights into how dietary MOTU richness (related to α -diversity) varies across time and space. Day of year predicted a small but significant increase in dietary richness over the course of the spring ($b = 0.082$ (0.024 – 0.135), Fig. 2C), with the expected number of MOTUs per faecal sample increasing from 1.981 to 3.933 from the first to last date (Table 2). For elevation ($b = -0.022$ (-0.131 – 0.104)) and latitude ($b = 0.058$ (-0.015 – 0.144)) gradients in dietary richness were non-significant (Fig. 2A-B, Table 2), as were the metrics describing among-site variation in woodland habitat (total foliage, foliage diversity, amount of oak, amount of birch, Table S5B). The repeatability of species richness within nestboxes at a site was moderate (Data; 0.140

(0.041-0.264), Latent; 0.158 (0.046-0.296), Appendix 2) but we found little evidence that richness varied among sites or among days within a year (after controlling for the linear increase). The effect of including more than one faeces in the sample was positive, but non-significant.

508

509 **Dietary MOTU turnover**

510 The probability of being present in a sample varied substantially across MOTUs
511 (variance on probit scale = 0.574 (0.475 – 0.696), Appendix 1 Table S5B). From the
512 interactions between MOTU identity and other terms we can gain insights into how
513 the probability of sampling individual MOTUs changes over time and space, providing
514 a measure of turnover and its significance. There was significant among MOTU
515 variation in the slope of presence/absence on day of year, elevation and latitude
516 (Appendix 1 Table S5B, Fig. 2D-F), with MOTU turnover more pronounced over
517 elevation and day of year. However, the predicted repeatabilities for MOTUs in faeces
518 sampled at the same elevation (but at different sites) were rather low (Data; 0.002
519 (0.001-0.003), Latent; 0.041 (0.028-0.059)). Due to the substantial between- faeces
520 and between nest-box variation in MOTU presence the repeatability for the site-level
521 probability of a MOTU at the same elevation was higher (Data; 0.066 (0.041 – 0.095),
522 Latent; 0.148 (0.106 – 0.215)), but still modest. The effect of date was similarly low
523 and even within nestboxes the repeatability of a MOTU in faeces from the same day
524 was small (Data; 0.002 (0.001-0.004), Latent; 0.081 (0.057-0.114)). See Appendix 2
525 for further analysis of repeatabilities. As an alternative measure of how environmental
526 variables affect community composition we calculated the expectation for the Jaccard
527 index and standardised Jaccard index (Appendix 2) between two sites at (i) the mean

528 and (ii) sampled extremes of latitude, elevation and day of year (Table 2). For all three
529 environmental variables communities are less similar (lower Jaccard index) at the
530 extremes than they are at the mean, but this effect is most pronounced for elevation
531 and day of year.

532

533 We considered among-MOTU variation in the relationship between the four
534 continuous habitat variables and probability of occurrence in four additional models
535 (Tables S5C–F). For three habitat metrics (total foliage, tree diversity and oak
536 availability) among-MOTU variation in habitat slopes was small and non-significant,
537 implying no discernible MOTU turnover along these gradients. The slope of MOTU
538 presence/absence on birch availability exhibited significant among-MOTU variation,
539 but turnover along this gradient is less than found for biogeographic and temporal
540 gradients (Appendix 1 Table S5F, Fig. S4) indicating a weak relationship.

541

542 The variance in the MOTU identity by site effects was large (0.474 (0.394 – 0.551)),
543 revealing that even after controlling for biogeographic trends in turnover gradients
544 there is substantial MOTU turnover among sites (Table S5B). Indeed, the
545 biogeographic and habitat variables in aggregate only explained a small fraction of the
546 between site variance (Data; 0.101 (0.069 – 0.142), Latent; 0.236 (0.174 – 0.296),
547 Appendix 2). The total within-site (due to both biogeographic variation and random
548 site variation) repeatability was small if assessed at the level of faeces (Data; 0.023
549 (0.016 – 0.029), Latent; 0.275 (0.242 – 0.306)) but larger if assessed at the level of
550 nestboxes (Data; 0.270 (0.223 – 0.334), Latent; 0.568 (0.520 – 0.628)). This arises
551 because of the considerable variance amongst faeces within a nestbox. The variance

552 in MOTU identity by nestbox effects was comparable to the site effects (0.434 (0.376
553 –0.529)), but the within-nestbox repeatability at a single site was small (Data; 0.016
554 (0.012 – 0.022), Latent; 0.275 (0.247 – 0.312)), again because of the large between-
555 faeces variance. The within-nestbox repeatability across sites (where site and nestbox
556 effects contribute to the between group variance) was greater (Data; 0.069 (0.057 –
557 0.083), Latent; 0.474 (0.445 – 0.502)).

558

559 Interactions between MOTU and plate, plate-row and plate-column were also
560 significant (Appendix 1 Table S5B), which may reflect within plate contamination.
561 However, our placing of samples on the plate in the order in which samples were
562 collected in the field (spatially and temporally structured) could also contribute to this
563 signature if there is spatiotemporal structure in MOTU presence/absence that is not
564 accounted for by the day of year:MOTU and site:MOTU terms.

565

566 **Order level trends**

567 Lepidoptera showed a significant increase in probability of occurrence with increasing
568 latitude ($b = 0.236$ (0.044 – 0.430)) and elevation ($b = 0.309$ (0.073 – 0.583), Fig.
569 3AB, Appendix 1 Table S6). Other than Lepidoptera, only Diptera also showed a
570 significant increase with latitude ($b = 0.252$ (0.058 – 0.446)). Hymenoptera showed a
571 significant increase in probability of occurrence with increasing elevation ($b = 0.319$
572 (0.061 – 0.557)), with positive trends also apparent for Diptera, Hemiptera and
573 Coleoptera.

574

575 The probability of sampling a hemipteran increases very steeply through time over the
576 course of the spring ($b = 0.422$ ($0.259 - 0.590$)), with significant positive relationships
577 also apparent for Lepidoptera ($b = 0.174$ ($0.006 - 0.341$)) and Coleoptera ($b = 0.269$
578 ($0.113 - 0.424$)) (Fig. 3C). Increasing site level tree diversity had a significant positive
579 effect on the probability of sampling Diptera ($b = 0.344$ ($0.095 - 0.586$)) and a
580 significant negative effect on the probability of sampling Hymenoptera ($b = -0.283$ ($-$
581 $0.528 - -0.037$), Fig. 3D).

582

583

584 **Discussion**

585

586 We demonstrate that faecal metabarcoding can provide deep insights into the diet of a
587 generalist woodland bird, and provide the first in-depth analysis of the natural diet of
588 a passerine bird prior to breeding. We show that across Scottish woodlands in early
589 spring - when overall food availability is low - blue tits are able to locate and harvest
590 over 400 prey taxa. Further, we show strong temporal patterns in the taxonomic
591 richness and composition of the invertebrate prey items.

592

593 **Diet Composition**

594 Our findings on blue tit diet composition broadly agree with previous work on this
595 species (Betts, 1955; J. Gibb & Betts, 1963). As for previous faecal metabarcoding
596 studies on generalist insectivores (Clare, Fraser, Braid, Fenton, & Hebert, 2009;
597 Jedlicka, Vo, & Almeida, 2016; Sedlock, Krüger, & Clare, 2014), we found most
598 dietary taxa to be rare. The six most common orders were also detected using

599 morphology-based identification of gizzard contents by Betts (1955). For a fuller
600 discussion of the commonest taxa see the extended discussion in Appendix 1.

601

602 One surprise in our data was the prevalence of winter moth early in the spring. The
603 larvae of this species are one of the main foods provisioned to nestling tits (Betts, 1955;
604 C. Perrins, 1991) and whilst they are the most common spring Lepidopteran larvae on
605 our transect, their availability peaks in late May/early June (Shutt, Burgess, et al.,
606 2019), and so we did not anticipate finding them in the diet in March/April. A *post hoc*
607 analysis (GLMM with threshold response, site and nestbox effects as random and year
608 effects as fixed) revealed that the probability of occurrence in a sample increases
609 significantly in the days running up to the site-average first egg laying date ($b = 0.039$,
610 $CI = 0.023 - 0.055$), from around a 2% chance at 30 days prior to laying to 17% at the
611 average site-level blue tit first egg date. This increase in the incidence of winter moth
612 in the diet most likely corresponds with a change in the availability of early instar
613 larvae, rather than eggs, which would be available throughout the period (Waring &
614 Townsend, 2017). This finding raises the possibility that tits might use early instars of
615 winter moth and other foliar caterpillar larvae as a cue of when to breed.

616

617 **Dietary Richness and Turnover**

618 The biogeographic variables that we considered, latitude and elevation, had no
619 significant effect upon dietary MOTU richness, but a significant effect upon dietary
620 turnover. This reveals that whilst the total richness of prey eaten may not vary
621 geographically (see also the very low site variance), the taxa comprising the diet vary
622 along biogeographic clines (more so over elevation than latitude) and also from site to

623 site, as revealed by the significant site by MOTU interaction component. These
624 findings are consistent with those from faecal metabarcoding of insectivorous bats
625 (Clare, Symondson, Broders, et al., 2014; Sedlock et al., 2014) and could indicate local
626 dietary specialisation. However, we suspect that a more likely explanation for this
627 apparent specialisation is that it arises from patterns in prey availability (V. Moran &
628 Southwood, 1982) and that the birds are flexible in their prey.

629

630 The increase in dietary MOTU richness as spring progresses parallels seasonal
631 increases in the abundance and availability of herbivorous insects in European forests
632 (Bale et al., 2002; Southwood, Wint, Kennedy, & Greenwood, 2004). Whilst dietary
633 richness generally increases during spring, some taxa become less likely to occur and
634 others more so, arising from the distinct phenologies of individual prey taxa (Forrest,
635 2016; Southwood et al., 2004). All of the main orders showed a tendency toward
636 increasing as spring progressed, though on the data scale the increase was steepest for
637 Hemiptera, which may be attributable to a pronounced spring phenology in the
638 abundance of aphids on buds and leaves (Bell et al., 2015).

639

640 The habitat indices that we consider were non-significant predictors of blue tit dietary
641 richness, and MOTU turnover along such gradients was much weaker than estimated
642 for the biogeographic and temporal variables. One potential explanation for our low
643 estimate of turnover along such habitat gradients is that most invertebrate prey species
644 may not be entirely restricted to a particular tree species. Alternatively, perhaps our
645 ‘territory’ based habitat metrics are inadequate measures of the availability of different
646 tree species to each bird at this time. At face value our results are consistent with the

647 greater importance of larger-scale geographic clines (i.e. latitude, elevation) as
648 determinants of prey presence/absence, presumably because they act as a proxy for
649 other environmental variables that limit invertebrate distributions, such as temperature.
650 However, substantial spatial turnover remained even after controlling for
651 spatiotemporal gradients, which suggests that there are important drivers of prey
652 turnover that we have overlooked.

653

654 **Model based inference of richness and turnover**

655 Describing and explaining temporal and geographical variation in components of
656 diversity is a mainstay of community ecology (Dornelas et al., 2014; Li et al., 2018;
657 Magurran, 2013). α -diversity can be calculated for the sampled community scale (be
658 that a location or point in time), which has made its statistical analysis relatively
659 straightforward. In comparison, β -diversity is often calculated as a pairwise
660 similarity between communities (Koleff, Gaston, & Lennon, 2003), and where
661 multiple communities are considered the non-independence of comparisons presents
662 a challenge to statistical inference (Baselga, 2010). In an important development
663 Baeten et al. (2014) explained how a generalized linear model with taxon
664 presence/absence as a binomial response could be used to estimate changes in
665 richness and turnover between points and crucially determine statistical significance.
666 Here we have extended their framework to a generalized linear mixed model and we
667 show that the interaction of taxon (MOTU) with categorical (random intercepts) and
668 continuous (random slopes) variables estimates turnover between points (in space or
669 time) and along gradients, respectively. We also show that it is possible to predict the
670 Jaccard index (measure of β -diversity) between a pair of communities sampled at

671 points in space or time as a measure of effect size (Appendix 2). The principal
672 benefits of this new model-based approach over existing pair-wise approaches are
673 that (i) it allows estimation of confidence intervals and p values for turnover and
674 richness along gradients and among samples without such calculations being
675 complicated by non-independence; (ii) hierarchical structure in the sampling can be
676 included, and turnover can be assessed at each level explicitly taking into account
677 heterogeneity in sampling effort at lower levels; (iii) multiple covariates can be
678 included; (iv) inferences can be made including or excluding a control for taxon
679 abundance and (v) model based inference of repeatability is possible (see Appendix
680 2). The model coefficients can also be used to derive predictions of the total number
681 of taxa in a community and the Jaccard index (or alternative β -diversity metric)
682 between communities. Our model is defined in the context of the probability of a
683 taxon being present in a faecal sample, and as the number of samples (n) increases
684 total taxon richness is predicted to increase monotonically (with a decelerating
685 slope), such that when $n = \infty$, every taxon will be present. There are similarities
686 between this curve and rarefaction curves that are often used to standardise for
687 heterogeneity in sampling in ecology (Gotelli & Colwell, 2011), with both methods
688 requiring inference of the probability of each taxon being in a sample. In addition,
689 the Jaccard index will increase monotonically and with an accelerating function with
690 increasing species richness (Appendix 2) and monotonically and with a decelerating
691 function with sampling effort. Given that community level diversity metrics are
692 highly sensitive to the choice of n , we suggest that when using our framework an $n =$
693 1 represents the most natural level at which to report community-level metrics (see
694 Appendix 2) and requires no extrapolation.

695

696 A limitation of our approach is that by imposing a parametric correlation structure on
697 the data, that correlation structure is relatively simple and probably doesn't catch the
698 full complexity of species associations. For example, if there was a patchily
699 distributed species of herb on which three prey taxa were specialised on, then these
700 three species would co-occur with higher probability than our model would suggest.
701 Rectifying these problem would require a) identifying the herb that generates these
702 correlations, measuring its prevalence and incorporating that data into the model b)
703 use more complex correlation structures to be modelled in situations where the
704 number of taxa is large (Runcie & Mukherjee, 2013; Warton et al., 2015) or c)
705 develop sandwich type estimators (Huber, 1967; Zeger, Liang, & Albert, 1988) that
706 would allow robust inferences to be made even when unmodelled correlations exist.

707

708 **Methodological Considerations**

709 In this study we have demonstrated that faecal metabarcoding can provide a robust and
710 powerful method for assessing passerine diet, allowing greater sample sizes and
711 taxonomic resolution than direct assessment (Betts, 1955). Inclusion of positive and
712 negative controls and repeat samples are part of the standard laboratory practice
713 (Alberdi et al., 2018) – though few previous metabarcoding studies have included any
714 of these (but see De Barba et al., 2014; Jedlicka et al., 2016) – and have proven
715 invaluable in informing this work. Our protocol yielded fourteen MOTUs for the
716 positive control taxon, suggesting that the 2% divergence rule of thumb used in early
717 barcoding studies to group conspecific COI barcode sequences in Metazoa (Hebert,
718 Cywinska, & Ball, 2003 and <http://www.barcodinglife.com>) is likely to produce

719 spurious taxa, potentially misleading naïve analyses and underlining the necessity for
720 subsequent quality control steps. Negative controls (extraction and PCR) allowed us
721 to identify a case of systematic contamination and also informed our cut-off number
722 of reads (but see Deagle et al., 2018 for a critique of thresholds). After strict removal
723 of samples that appeared likely to have been affected by systematic contamination,
724 some residual contamination on plates was evident and we were able to control for this
725 to some degree by including row:MOTU, column:MOTU and plate:MOTU as random
726 terms. We recommend that future studies adopt the plate:MOTU random term and
727 randomise samples across plates, such that samples from a single year, site or time of
728 year do not all appear on one plate. Although the maximum number of taxa in a sample
729 was high ($n = 20$), PCR competition and the methodological maximum reads per
730 metabarcoding plate presumably place a limit on detecting very rare dietary items.
731 Reducing the number of target loci (three in this study, see methods) or level of
732 multiplexing (i.e. the number of samples per sequencing run) could increase the reads
733 available per locus per sample and increase detectability. However, reducing
734 multiplexing may come at an increased financial cost for sequencing.

735

736 From our repeat samples we were able to estimate technical repeatability and several
737 measures of biological repeatability (Appendix 2). Repeatability of MOTU
738 presence/absence was rather low, consistent with low repeatability estimates found by
739 another study that subsampled avian faecal samples (Jedlicka et al., 2016). An
740 implication is that if the focus of an avian faecal metabarcoding study is on the
741 detection of the presence/absence of a specific taxon, then multiple repeat DNA
742 extractions, amplifications and metabarcoding runs are advisable. Homogenisation of

743 faecal samples prior to DNA extraction may increase both the ability to detect a
744 particular taxon and repeatability given the possible heterogeneity within single faeces.

745

746 **Conclusion**

747 Using a metabarcoding approach, we reveal the diet of a generalist passerine at a finer
748 resolution than any previous study and quantify dietary richness and turnover across
749 space and time. At the scale of our study, blue tit dietary richness increases as spring
750 progresses, but is unaffected by latitude, elevation and habitat, whilst dietary turnover
751 is most pronounced over temporal (day of year) and elevational gradients.

752

753

754 **Acknowledgements**

755

756 The authors thank Irene Benedicto Cabello and Ed Ivimey-Cook for assistance in the
757 field, Mark Blaxter for molecular advice, two anonymous reviewers for comments on
758 the ms and Andrés Baselga for discussion of beta-diversity. We are indebted to the
759 landowners who allowed us access to their land. JDS was funded by a NERC
760 doctoral training studentship (NE/1338530), ABP by a NERC Advanced fellowship
761 (NE/I020598/1) and JDH by a Royal Society URF.

762

763

764 **References**

765 Alberdi, A., Aizpurua, O., Gilbert, M. T. P., & Bohmann, K. (2018). Scrutinizing key steps
766 for reliable metabarcoding of environmental samples. *Methods in Ecology and*
767 *Evolution*, 9(1), 134-147.

Baeten, L., Warton, D. I., Van Calster, H., De Frenne, P., Verstraeten, G., Bonte, D., . . . Eriksson, O. (2014). A model-based approach to studying changes in compositional heterogeneity. *Methods in Ecology and Evolution*, 5(2), 156-164.

Bale, J. S., Masters, G. J., Hodkinson, I. D., Awmack, C., Bezemer, T. M., Brown, V. K., . . . Farrar, J. (2002). Herbivory in global climate change research: direct effects of rising temperature on insect herbivores. *Global change biology*, 8(1), 1-16.

Baselga, A. (2008). Determinants of species richness, endemism and turnover in European longhorn beetles. *Ecography*, 31(2), 263-271.

Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity. *Global ecology and biogeography*, 19(1), 134-143.

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes.

Beck, J., McCain, C. M., Axmacher, J. C., Ashton, L. A., Bärtschi, F., Brehm, G., . . . Fiedler, K. (2017). Elevational species richness gradients in a hyperdiverse insect taxon: a global meta-study on geometrid moths. *Global ecology and biogeography*, 26(4), 412-424.

Bell, J. R., Alderson, L., Izera, D., Kruger, T., Parker, S., Pickup, J., . . . Harrington, R. (2015). Long-term phenological trends, species accumulation rates, aphid traits and climate: five decades of change in migrating aphids. *Journal of Animal Ecology*, 84(1), 21-34.

Betts, M. M. (1955). The food of titmice in oak woodland. *The Journal of Animal Ecology*, 24, 282-323.

Bohmann, K., Monadjem, A., Noer, C. L., Rasmussen, M., Zeale, M. R., Clare, E., . . . Gilbert, M. T. P. (2011). Molecular diet analysis of two African free-tailed bats (Molossidae) using high throughput sequencing. *PLoS One*, 6(6), e21441.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., . . . Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R journal*, 9(2), 378-400.

Burger, C., Belskii, E., Eeva, T., Laaksonen, T., Mägi, M., Mänd, R., . . . Both, C. (2012). Climate change, breeding date and nestling diet: how temperature differentially affects seasonal changes in pied flycatcher diet depending on habitat variation. *Journal of Animal Ecology*, 81(4), 926-936. doi:10.1111/j.1365-2656.2012.01968.x

Chao, A., Chazdon, R. L., Colwell, R. K., & Shen, T. J. (2006). Abundance-based similarity indices and their estimation when there are unseen species in samples. *Biometrics*, 62(2), 361-371.

Charmantier, A., McCleery, R. H., Cole, L. R., Perrins, C., Kruuk, L. E. B., & Sheldon, B. C. (2008). Adaptive Phenotypic Plasticity in Response to Climate Change in a Wild Bird Population. *Science*, 320(5877), 800-803. doi:10.1126/science.1157174

Cholewa, M., & Wesołowski, T. (2011). Nestling food of European hole-nesting passerines: do we know enough to test the adaptive hypotheses on breeding seasons? *Acta Ornithologica*, 46(2), 105-116.

Clare, E. L. (2014). Molecular detection of trophic interactions: emerging trends, distinct advantages, significant considerations and conservation applications. *Evolutionary applications*, 7(9), 1144-1157.

Clare, E. L., Fraser, E. E., Braid, H. E., Fenton, M. B., & Hebert, P. D. (2009). Species on the menu of a generalist predator, the eastern red bat (*Lasiurus borealis*): using a molecular approach to detect arthropod prey. *Molecular ecology*, 18(11), 2532-2542.

Clare, E. L., Symondson, W. O., Broders, H., Fabianek, F., Fraser, E. E., MacKenzie, A., . . . Martinez-Nuñez, F. (2014). The diet of *Myotis lucifugus* across Canada: assessing foraging quality and diet variability. *Molecular ecology*, 23(15), 3618-3632.

Clare, E. L., Symondson, W. O., & Fenton, M. B. (2014). An inordinate fondness for beetles? Variation in seasonal dietary preferences of night-roosting big brown bats (*Eptesicus fuscus*). *Molecular ecology*, 23(15), 3633-3647.

- 820 Cramp, S., & Perrins, C. M. (1993). *Handbook of the Birds of Europe the Middle East and*
821 *North America*. Oxford, UK: OUP.
- 822 Crisol-Martínez, E., Moreno-Moyano, L. T., Wormington, K. R., Brown, P. H., & Stanley, D.
823 (2016). Using next-generation sequencing to contrast the diet and explore pest-
824 reduction services of sympatric bird species in macadamia orchards in Australia. *PLoS*
825 *One*, 11(3), e0150159.
- 826 De Barba, M., Miquel, C., Boyer, F., Mercier, C., Rioux, D., Coissac, E., & Taberlet, P. (2014).
827 DNA metabarcoding multiplexing and validation of data accuracy for diet assessment:
828 application to omnivorous diet. *Molecular Ecology Resources*, 14(2), 306-323.
- 829 Deagle, B. E., Thomas, A. C., McInnes, J. C., Clarke, L. J., Vesterinen, E. J., Clare, E. L., . . .
830 Eveson, J. P. (2018). Counting with DNA in metabarcoding studies: how should we
831 convert sequence reads to dietary data? *Molecular ecology*.
- 832 Dornelas, M., Gotelli, N. J., McGill, B., Shimadzu, H., Moyes, F., Sievers, C., & Magurran,
833 A. E. (2014). Assemblage time series reveal biodiversity change but not systematic
834 loss. *Science*, 344(6181), 296-299.
- 835 Durant, J. M., Hjermann, D. Ø., Anker-Nilssen, T., Beaugrand, G., Myrsetrud, A., Pettorelli,
836 N., & Stenseth, N. C. (2005). Timing and abundance as key mechanisms affecting
837 trophic interactions in variable environments. *Ecology Letters*, 8(9), 952-958.
- 838 Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon
839 reads. *Nature methods*, 10(10), 996.
- 840 Forrest, J. R. (2016). Complex responses of insect phenology to climate change. *Current*
841 *opinion in insect science*, 17, 49-54.
- 842 Gibb, J., & Betts, M. M. (1963). Food and food supply of nestling tits (Paridae) in Breckland
843 pine. *The Journal of Animal Ecology*, 489-533.
- 844 Gibb, J. A. (1954). Feeding ecology of tits, with notes on treecreeper and goldcrest. *Ibis*, 96(4),
845 513-543.
- 846 Gotelli, N. J., & Colwell, R. K. (2011). Estimating species richness. *Biological diversity:*
847 *frontiers in measurement and assessment*, 12, 39-54.
- 848 Hadfield, J. D. (2010). MCMC Methods for Multi-Response Generalized Linear Mixed
849 Models: The MCMCglmm R Package. *Journal of Statistical Software*, 33, 1-22.
- 850 Hadfield, J. D., & Nakagawa, S. (2010). General quantitative genetic methods for comparative
851 biology: phylogenies, taxonomies and multi-trait models for continuous and
852 categorical characters. *Journal of evolutionary biology*, 23, 494-508.
- 853 Hebert, P. D., Cywinska, A., & Ball, S. L. (2003). Biological identifications through DNA
854 barcodes. *Proceedings of the Royal Society of London B: Biological Sciences*,
855 270(1512), 313-321.
- 856 Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard
857 conditions. *Proceedings of the fifth Berkeley symposium on mathematical statistics*
858 *and probability*, 1(1), 221-233.
- 859 Jedlicka, J. A., Sharma, A. M., & Almeida, R. P. (2013). Molecular tools reveal diets of
860 insectivorous birds from predator fecal matter. *Conservation Genetics Resources*,
861 5(3), 879-885.
- 862 Jedlicka, J. A., Vo, A.-T. E., & Almeida, R. P. (2016). Molecular scatology and high-
863 throughput sequencing reveal predominately herbivorous insects in the diets of adult
864 and nestling Western Bluebirds (*Sialia mexicana*) in California vineyards. *The Auk*,
865 134(1), 116-127.
- 866 Kennedy, C., & Southwood, T. (1984). The number of species of insects associated with
867 British trees: a re-analysis. *The Journal of Animal Ecology*, 455-478.
- 868 King, R., Symondson, W., & Thomas, R. (2015). Molecular analysis of faecal samples from
869 birds to identify potential crop pests and useful biocontrol agents in natural areas.
870 *Bulletin of entomological research*, 105(3), 261-272.
- 871 Koleff, P., Gaston, K. J., & Lennon, J. J. (2003). Measuring beta diversity for presence-
872 absence data. *Journal of Animal Ecology*, 72(3), 367-382.

873 Kress, W. J., García-Robledo, C., Uriarte, M., & Erickson, D. L. (2015). DNA barcodes for
874 ecology, evolution, and conservation. *Trends in Ecology & Evolution*, 30(1), 25-35.

875 Li, Y., Evans, N. T., Renshaw, M. A., Jerde, C. L., Olds, B. P., Shogren, A. J., . . . Pfrender,
876 M. E. (2018). Estimating fish alpha-and beta-diversity along a small stream with
877 environmental DNA metabarcoding. *Metabarcoding and Metagenomics*, 2, e24262.

878 Magurran, A. E. (2013). *Measuring biological diversity*: John Wiley & Sons.

879 Mata, V. A., Rebelo, H., Amorim, F., McCracken, G. F., Jarman, S., & Beja, P. (2019). How
880 much is enough? Effects of technical and biological replication on metabarcoding
881 dietary analysis. *Molecular ecology*, 28(2), 165-175.

882 Moran, A. J., Prosser, S. W., & Moran, J. A. (2019). DNA metabarcoding allows non-invasive
883 identification of arthropod prey provisioned to nestling Rufous hummingbirds
884 (*Selasphorus rufus*). *PeerJ*, 7, e6596.

885 Moran, V., & Southwood, T. (1982). The guild composition of arthropod communities in trees.
886 *The Journal of Animal Ecology*, 289-306.

887 Murakami, M., Ichie, T., & Hirao, T. (2008). Beta-diversity of lepidopteran larval
888 communities in a Japanese temperate forest: effects of phenology and tree species.
889 *Ecological research*, 23(1), 179-187.

890 Nakagawa, S., & Schielzeth, H. (2010). Repeatability for Gaussian and non- Gaussian data: a
891 practical guide for biologists. *Biological Reviews*, 85(4), 935-956.

892 Nichols, R. V., Åkesson, M., & Kjellander, P. (2016). Diet assessment based on rumen
893 contents: A comparison between DNA metabarcoding and macroscopy. *PLoS One*,
894 11(6).

895 Niku, J., Hui, F. K., Taskinen, S., & Warton, D. I. (2019). gllvm: Fast analysis of multivariate
896 abundance data with generalized linear latent variable models in r. *Methods in Ecology
897 and Evolution*, 10(12), 2173-2182.

898 Novotny, V., & Weiblen, G. D. (2005). *From communities to continents: beta diversity of
899 herbivorous insects*. Paper presented at the Annales Zoologici Fennici.

900 Oehm, J., Juen, A., Nagiller, K., Neuhauser, S., & Traugott, M. (2011). Molecular scatology:
901 how to improve prey DNA detection success in avian faeces? *Molecular ecology
902 resources*, 11(4), 620-628.

903 Perrins, C. (1991). Tits and their caterpillar food supply. *Ibis*, 133(s1), 49-54.

904 Perrins, C. M. (1979). *British Tits*. Glasgow: Collins.

905 Pompanon, F., Deagle, B. E., Symondson, W. O., Brown, D. S., Jarman, S. N., & Taberlet, P.
906 (2012). Who is eating what: diet assessment using next generation sequencing.
907 *Molecular ecology*, 21(8), 1931-1950.

908 Quéméré, E., Hibert, F., Miquel, C., Lhuillier, E., Rasolondraibe, E., Champeau, J., . . .
909 Gautier, L. (2013). A DNA metabarcoding study of a primate dietary diversity and
910 plasticity across its entire fragmented range. *PLoS One*, 8(3), e58971.

911 Razgour, O., Clare, E. L., Zeale, M. R., Hanmer, J., Schnell, I. B., Rasmussen, M., . . . Jones,
912 G. (2011). High- throughput sequencing offers insight into mechanisms of resource
913 partitioning in cryptic bat species. *Ecology and Evolution*, 1(4), 556-570.

914 Runcie, D. E., & Mukherjee, S. (2013). Dissecting high-dimensional phenotypes with
915 Bayesian sparse factor analysis of genetic covariance matrices. *Genetics*, 194(3), 753-
916 767.

917 Rytkönen, S., Vesterinen, E. J., Westerduin, C., Leviäkangas, T., Votka, E., Mutanen, M., . . .
918 Orell, M. (2019). From feces to data: A metabarcoding method for analyzing
919 consumed and available prey in a bird- insect food web. *Ecology and evolution*, 9(1),
920 631-639.

921 Samplonius, J. M., Kappers, E. F., Brands, S., & Both, C. (2016). Phenological mismatch and
922 ontogenetic diet shifts interactively affect offspring condition in a passerine. *Journal
923 of Animal Ecology*, 85(5), 1255-1264.

- 924 Sedlock, J. L., Krüger, F., & Clare, E. L. (2014). Island bat diets: does it matter more who you
925 are or where you live? *Molecular ecology*, 23(15), 3684-3694.
- 926 Shutt, J. D., Bolton, M., Benedicto Cabello, I., Burgess, M. D., & Phillimore, A. B. (2018).
927 The effects of woodland habitat and biogeography on blue tit *Cyanistes caeruleus*
928 territory occupancy and productivity along a 220 km transect. *Ecography*, 41, 1967-
929 1978.
- 930 Shutt, J. D., Burgess, M. D., & Phillimore, A. B. (2019). A spatial perspective on the
931 phenological distribution of the spring woodland caterpillar peak. *American*
932 *Naturalist*, 194, E000.
- 933 Shutt, J. D., Nicholls, J. A., Trivedi, U. H., Burgess, M. D., Stone, G. N., Hadfield, J. D., &
934 Phillimore, A. B. (2019). *MOTU presence absence data*. Retrieved from:
935 <https://doi.org/10.5061/dryad.hhmgqnkd3>
- 936 Southwood, T. R. E., Wint, G. W., Kennedy, C. E., & Greenwood, S. R. (2004). Seasonality
937 abundance, species richness and specificity of the phytophagous guild of insects on
938 oak (*Quercus*) canopies. *European Journal of Entomology*, 101(1), 43-50.
- 939 Sterling, P., & Parsons, M. (2012). *Field guide to the micro-moths of Great Britain and*
940 *Ireland*: British Wildlife Publishing.
- 941 Sullins, D. S., Haukos, D. A., Craine, J. M., Lautenbach, J. M., Robinson, S. G., Lautenbach,
942 J. D., . . . Sandercock, B. K. (2018). Identifying the diet of a declining prairie grouse
943 using DNA metabarcoding. *The Auk: Ornithological Advances*, 135(3), 583-608.
- 944 Symondson, W. (2002). Molecular identification of prey in predator diets. *Molecular ecology*,
945 11(4), 627-641.
- 946 Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards
947 next-generation biodiversity assessment using DNA metabarcoding. *Molecular*
948 *ecology*, 21(8), 2045-2050.
- 949 Thomsen, P. F., Jørgensen, P. S., Bruun, H. H., Pedersen, J., Riis-Nielsen, T., Jonko, K., . . .
950 Karsholt, O. (2016). Resource specialists lead local insect community turnover
951 associated with temperature—analysis of an 18-year full-seasonal record of moths and
952 beetles. *Journal of Animal Ecology*, 85(1), 251-261.
- 953 Thorup, K., Tøttrup, A. P., Willemoes, M., Klaassen, R. H., Strandberg, R., Vega, M. L., . . .
954 Rahbek, C. (2017). Resource tracking within and across continents in long-distance
955 bird migrants. *Science Advances*, 3(1), e1601360.
- 956 Trevelline, B. K., Latta, S. C., Marshall, L. C., Nuttle, T., & Porter, B. A. (2016). Molecular
957 analysis of nestling diet in a long-distance Neotropical migrant, the Louisiana
958 Waterthrush (*Parkesia motacilla*). *The Auk*, 133(3), 415-428.
- 959 Trevelline, B. K., Nuttle, T., Hoenig, B. D., Brouwer, N. L., Porter, B. A., & Latta, S. C.
960 (2018). DNA metabarcoding of nestling feces reveals provisioning of aquatic prey and
961 resource partitioning among Neotropical migratory songbirds in a riparian habitat.
962 *Oecologia*, 187(1), 85-98.
- 963 Trevelline, B. K., Nuttle, T., Porter, B. A., Brouwer, N. L., Hoenig, B. D., Steffensmeier, Z.
964 D., & Latta, S. C. (2018). Stream acidification and reduced aquatic prey availability
965 are associated with dietary shifts in an obligate riparian Neotropical migratory
966 songbird. *PeerJ*, 6, e5141.
- 967 Vo, A. T., & Jedlicka, J. (2014). Protocols for metagenomic DNA extraction and Illumina
968 amplicon library preparation for faecal and swab samples. *Molecular ecology*
969 *resources*, 14(6), 1183-1197.
- 970 Waring, P., & Townsend, M. (2017). *Field guide to the moths of Great Britain and Ireland*:
971 Bloomsbury Publishing.
- 972 Warton, D. I., Blanchet, F. G., O'Hara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C., &
973 Hui, F. K. (2015). So many variables: joint modeling in community ecology. *Trends*
974 *in Ecology & Evolution*, 30(12), 766-779.
- 975 Whittaker, R. H. (1972). Evolution and measurement of species diversity. *Taxon*, 213-251.

976 Yu, D. W., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., & Ding, Z. (2012). Biodiversity
 977 soup: metabarcoding of arthropods for rapid biodiversity assessment and
 978 biomonitoring. *Methods in Ecology and Evolution*, 3(4), 613-623.
 979 Zeger, S. L., Liang, K.-Y., & Albert, P. S. (1988). Models for longitudinal data: a generalized
 980 estimating equation approach. *Biometrics*, 1049-1060.
 981

982 **Data Accessibility Statement**

983 The MOTU presence/absence data for COI are available
984 from <https://doi.org/10.5061/dryad.hhmgqnkd3>. The data and model outputs used in
985 appendix 2 are available from
986 <https://github.com/allyphillimore/faecalmetabarcoding-adults>

987

988

989 **Author contributions**

990 JDS, JAN, ABP and JDH were the main contributors to study conceptualization and
991 methodology, with JDS and ABP responsible for fieldwork, JAN responsible for
992 designing and conducting the molecular work and JDH designing the statistical
993 methods, developing the theory and writing Appendix 2. JDS and UHT contributed to
994 data curation. Statistical analysis was conducted by JDS, ABP and JDH. ABP was
995 responsible for project administration and ABP and JDH for funding acquisition. JDS
996 wrote the original draft, and all authors contributed to further writing and editing.

997

998

999 **Tables and Figures**

1000

1001 Table 1. Sampling and laboratory protocols employed by published faecal barcoding
 1002 studies focusing on the invertebrate component of diet. An entry of ‘None’ means that
 1003 while steps may have been taken in the study, no specific method was detailed.

1004

Number of study species (most common species)	Total number of samples (maximum number of samples per species)	Number of sites (region)	Controls	Measures taken to assess repeatability	Reference
1 (Lesser Prairie-Chicken)	314	4 (Kansas and Colorado, USA)	None	None	(Sullins et al., 2018)
1 (Western Bluebird)	210	3 (neighbouring vineyards, California, USA)	None	Ten faeces subsampled.	(Jedlicka et al., 2016)
3 (Wood Thrush)	137 (51)	1 (Pennsylvania, USA)	PCR negatives and positives	None	(Trevelline, Nuttle, Hoenig, et al., 2018)
1 (Louisiana Waterthrush)	130	2 (Arkansas and Pennsylvania, USA)	None	None	(Trevelline, Latta, Marshall, Nuttle, & Porter, 2016)
1 (Louisiana Waterthrush)	92	3 (headwater streams, Pennsylvania, USA)	None	None	(Trevelline, Nuttle, Porter, et al., 2018)
(Rufous hummingbird)	30	1 (Vancouver Island, Canada)	1 x extraction negative	None	(A. J. Moran, Prosser, & Moran, 2019)
13 (Lewin’s Honeyeater)	82 (29)	1 (Bundaberg, Australia)	Extraction negatives	PCR run twice to test amplification repeatability	(Crisol-Martínez, Moreno-Moyano, Wormington, Brown, & Stanley, 2016)

1 (Western Bluebird)*	16	2 (neighbouring vineyards, California, USA)	None	Faecal sample was subdivided and run on two extraction kits.	(Jedlicka et al., 2013)
4 (Blue tit, Great Tit, Willow Tit)	14 (4)	2 (Oulu and Kuusamo, Finland)	Extraction negative	None	(Rytönen et al., 2019)
3 (Sedge Warbler)‡	11 (6)	3 (South Wales, UK)	None	None	(King, Symondso n, & Thomas, 2015)

1005 ‡ Study employed Sanger sequencing rather than metabarcoding.

1006

Table 2. Expectations for the MOTU richness of - and Jaccard indices between - samples of communities at (i) the same and (ii) extreme points along latitude, elevation and day of year gradients. Expectations are calculated for a random sample, nestbox, day and site averaging over variation in other predictor variables (for further details see Appendix 2). Expectations were generated for 2014 and a single faecal sample.

Predictor	Sampling position	MOTU richness at mean	MOTU richness at minimum	MOTU richness at maximum	Jaccard index	Standardised Jaccard index ¹
Latitude	Mean	2.339 (1.665 - 3.271)			0.011 (0.008 - 0.015)	5.255 (4.238 - 6.296)
Latitude	Extremes		2.607 (1.476 - 3.831)	2.213 (1.266 - 3.444)	0.009 (0.006 - 0.012)	3.569 (2.776 - 4.423)
Elevation	Mean	2.340 (1.669 - 3.277)			0.013 (0.009 - 0.017)	5.966 (4.814 - 7.179)
Elevation	Extremes		2.435 (1.524 - 3.647)	2.138 (1.084 - 3.700)	0.008 (0.005 - 0.011)	2.344 (1.727 - 3.047)
Day of year	Mean	2.464 (1.653 - 3.252)			0.013 (0.008 - 0.017)	5.922 (4.811 - 7.103)
Day of year	Extremes		1.981 (1.352 - 2.848)	3.933 (2.459 - 5.603)	0.007 (0.005 - 0.010)	1.973 (1.431 - 2.592)

¹ The standardised Jaccard index is the ratio of the observed index to that expected if the same number of species were sampled at random from two communities (see Appendix 2). It will tend to be > 1 as common/widespread species will be over-represented in both communities. The expectation for the Jaccard index and standardised index for two samples taken entirely at random from the transect is

1019 0.01 (0.007 – 0.014) and 4.727 (3.891 – 5.732), respectively, and these values can be taken as a
1020 baseline that captures the effect of common/widespread species on measures of community similarity.

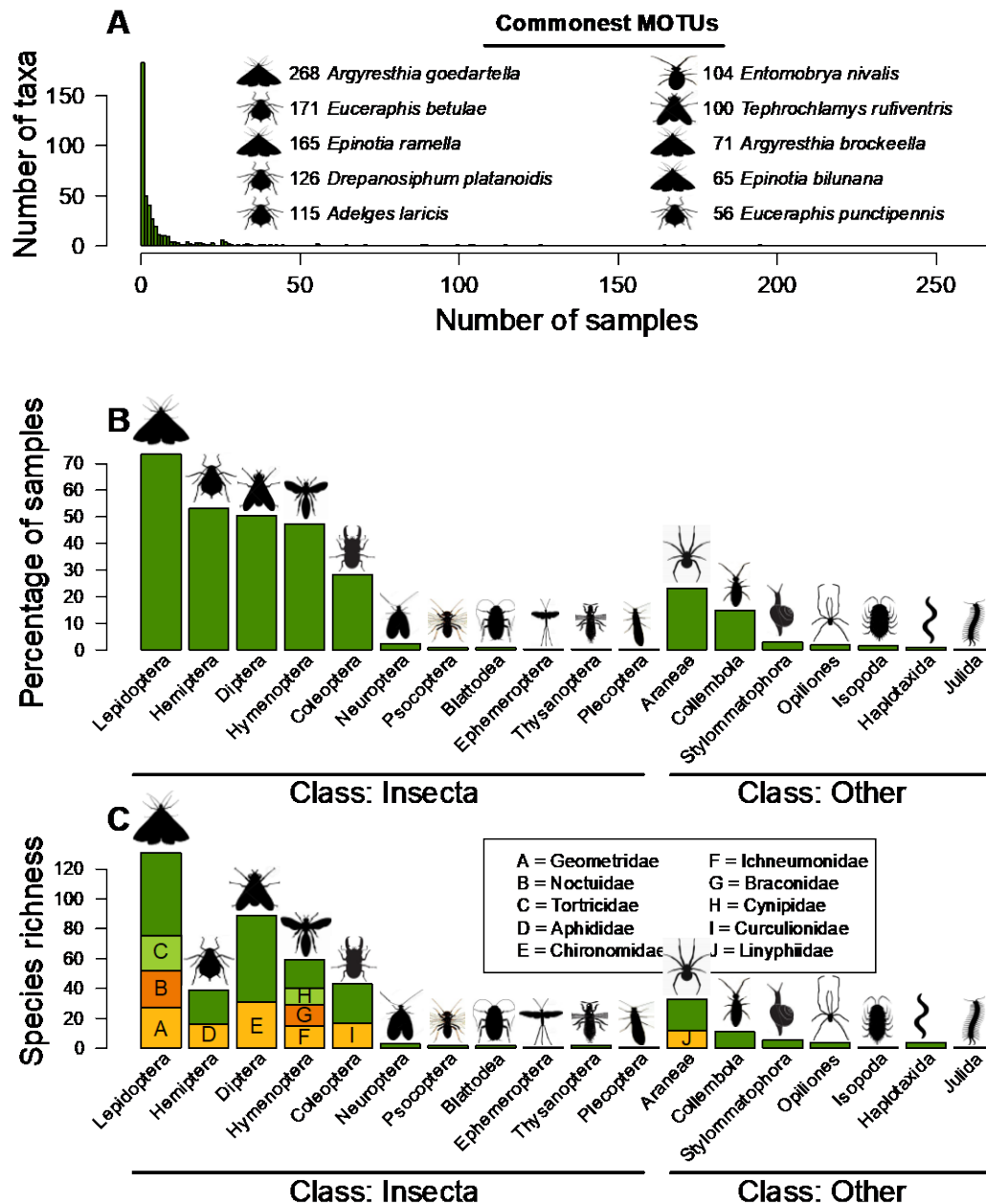


Fig. 1 **A** Histogram of the abundance distribution of prey MOTUs. Inset details the most prevalent MOTUs identified to species level (those recorded in more than 50 samples), with the number of samples they were recorded in. **B** Relative abundance of prey orders in the spring diet of blue tits. **C** Number of MOTUs within prey orders (families comprising > 10 MOTUs are highlighted individually within their respective orders). In **B** and **C** orders within Insecta (left) are split from orders within other classes (right). Images are used to indicate taxonomic order rather than the life-stage or species that is preyed upon.

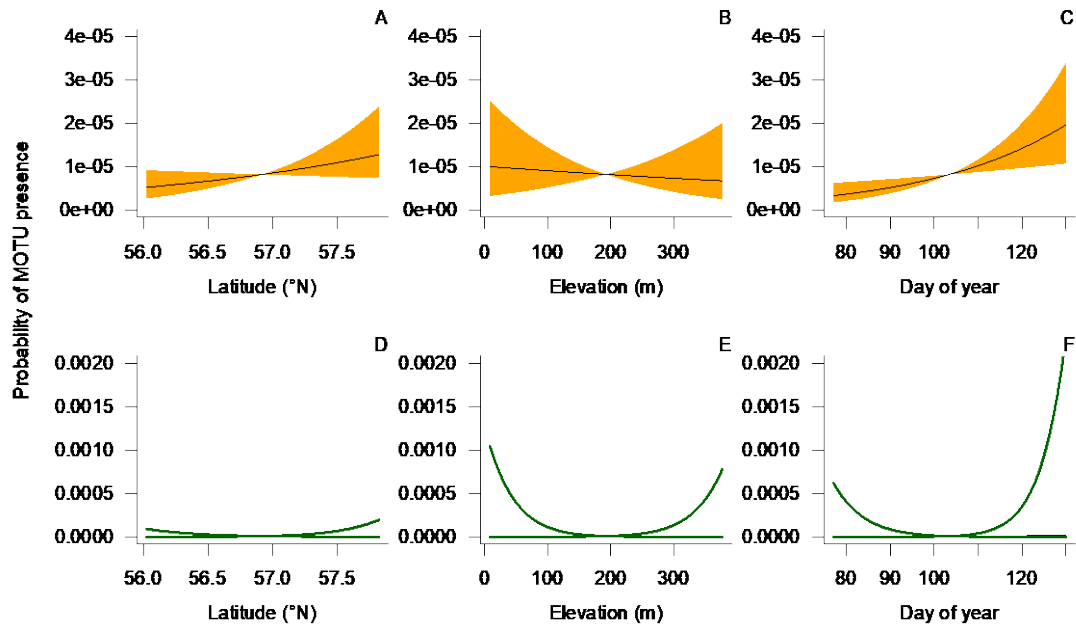
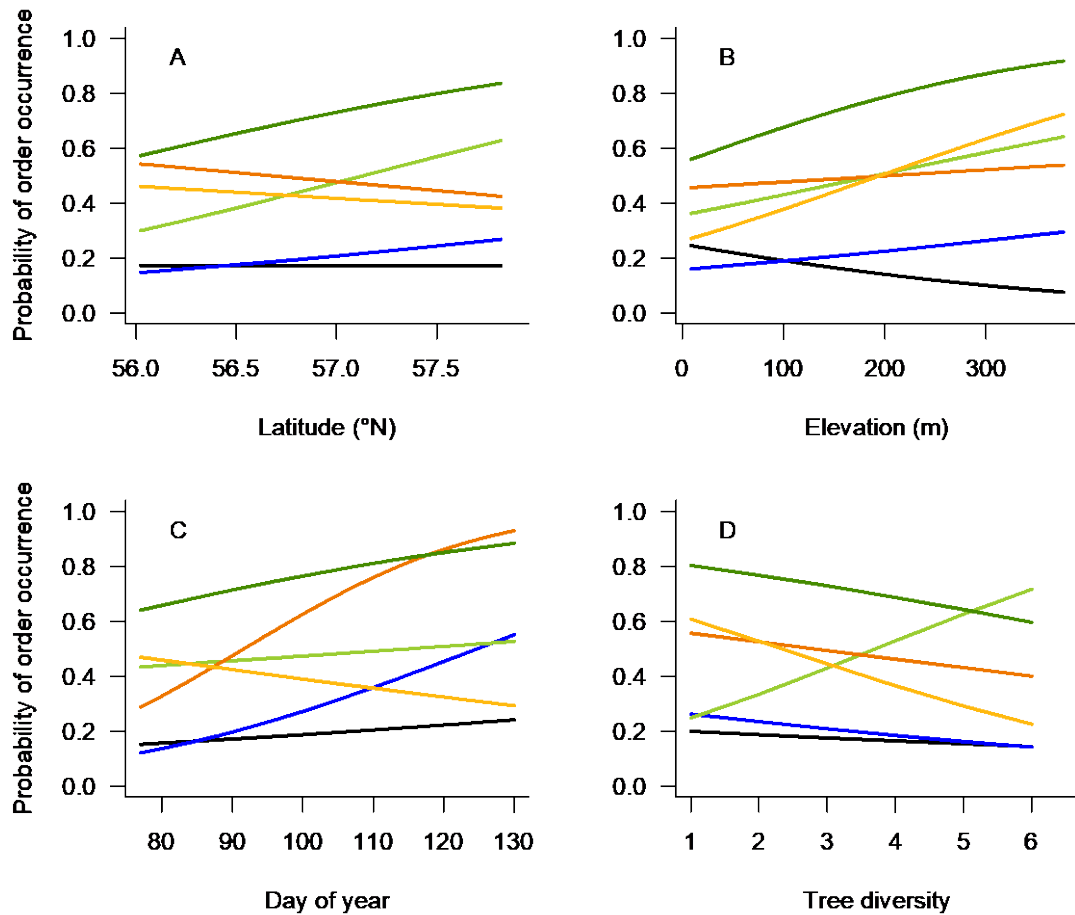


Fig. 2 Dietary richness (A – C) and turnover (D – F) along latitudinal (A, D), elevational (B, E) and temporal (C, F) gradients. In A - C the solid black lines indicates the model prediction of dietary MOTU occurrence (related to richness), with the solid orange area illustrating the 95% credible intervals in the slope. In D – F the green lines correspond to the 95% upper and lower bounds of the estimated distribution of among-MOTU slopes. The wider the difference between the upper and lower line the greater the turnover along the gradient. Predictions are made from the core model (Table S4B).



— Araneae — Coleoptera — Diptera — Hemiptera — Hymenoptera — Lepidoptera

Fig. 3 Model predictions for the occurrence of six prey orders across A. latitude, B. elevation, C. day of year and D. tree diversity. Predictions are made based on the intercept of the model reported in Table S5.