Signatures of genomic adaptation to Cuckoo brood parasitism in Reed Warbler hosts across Europe

William J. Smith[1,2,3*†], Katja Rönkä[1,2*], Staffan Bensch[4], Nora Bergman[1,2], Rachael Y. Dudaniec[5], Fabrice Eroukhmanoff[6], Frode Fossøy[7], Bengt Hansson[4], Michał T. Jezierski[1,2,8], Edward Kluen[1,2], Petr Procházka[9], Camilla Lo Cascio Sætre[6], Bård G. Stokke[7], & Rose Thorogood[1,2]

1. Helsinki Institute of Life Science, University of Helsinki, Finland.
2. Research Programme in Organismal and Evolutionary Biology, Faculty of Biological and Environmental Sciences, University of Helsinki, Finland.
3. Department of Biology, University of Oxford, United Kingdom.
4. Department of Biology, Lund University, Sweden.
5. School of Natural Sciences, Macquarie University, Sydney, NSW, Australia.
6. Department of Biology, University of Oslo, Norway.
7. Norwegian Institute for Nature Research (NINA), Trondheim, Norway.
8. School Of Geography, Earth and Environmental Sciences, University of Birmingham, United Kingdom.
9. Institute of Vertebrate Biology, Czech Academy of Sciences, Czech Republic.

*Joint first authors

† Correspondence should be addressed to william.smith@helsinki.fi

Abstract

Predicting coevolutionary outcomes is critical for understanding the consequences of climate change and biodiversity loss. The geographic mosaic theory of coevolution posits that coevolutionary interactions vary among populations, facilitating the persistence of such interactions over evolutionary timescales. Characterisation of geographic mosaics of coevolution is difficult, with mobile organisms and behavioural adaptations posing temporal and spatial challenges. The Reed Warbler (*Acrocephalus scirpaceus*) and its brood parasite, the Cuckoo (*Cuculus canorus*), have been used to study coevolution for decades. By leveraging RAD-sequencing and brood parasitism data for 188 Reed Warblers across 22 European sites, we first demonstrate that the magnitude of gene flow varies across the species' distribution, allowing for variable selection which is a fundamental component of the geographic mosaic theory. We then use genotype-environment association analysis to reveal genes that are putatively involved in the Reed Warbler's response to brood parasitism, providing the first evidence for a genetic basis to this geographic mosaic of coevolution. Together, these results help to characterise this important model of host–parasite coevolution. Further work is needed to confirm a causal relationship between parasitism and genomic variation, and to identify the exact selection mechanisms underpinning this relationship.

Keywords: coevolution, brood parasitism, gene flow, genotype-environment interactions, reed warbler, selection

Introduction

Antagonistic species interactions are major drivers of evolution. Predators and prey, as well as parasites or pathogens and their hosts, directly impact one another's fitness, exerting strong selective pressure. This can result in antagonistic coevolution[1], where adaptation and counter-adaptation lead to reciprocal evolution in interacting species. Such 'arms races' are of fundamental importance, impacting both contemporary population dynamics[2] and evolutionary diversification[3]. Despite its central role in ecology and evolution, the genetic basis of many aspects of antagonistic coevolution remains underexplored, hindering our ability to understand this phenomenon[4]. Recent advances in genomic approaches offer unprecedented opportunities to address these gaps. In ecology, we are now able to identify the functional genomic underpinnings of physiological changes required to colonise new environments[5]. Genomic approaches hold similar promise in helping us to deepen our understanding of antagonistic coevolution and its impact on interacting species, including in case studies where behavioural traits are fundamental components of these interactions[6].

The geographic mosaic theory of coevolution (GMTC) is the prevailing theory explaining how antagonistic interactions can be maintained over time[7]. Without spatio-temporal variation in species interactions, such arms-races are not predicted to last for long, since sooner or later one entity 'wins': either host defences render parasitism impossible, or the parasites overcome host defences[8]. However, if this happens only locally and there is enough gene flow from other populations, interactions can continue at the landscape level. Host shifts by parasites might also play an important role. Landscape-level interactions enable antagonistic evolutionary relationships to persist over time at a metapopulation level (in many bird species, for example,

brood parasitism relationships may operate at a continental scale)[7]. The study of GMTC-related questions remains particularly challenging when host and parasite adaptations manifest as behavioural traits, given their inherent plasticity. The use of genotype-environment analyses could provide a potent tool with which to characterise geographic mosaics of coevolution in such systems, particularly in non-model and mobile species where common garden experiments would be incredibly challenging to design and execute.

Despite plenty of attention on this topic, evidence for geographic mosaics has largely remained at a descriptive level, demonstrating hot and cold spots of coevolution[9] without a detailed understanding of the underlying mechanisms, such as the strength and direction of selection within either the parasite or host, or levels of gene flow within the host and parasite (although see Kaur *et al.* 2019 for work carried out in an invertebrate system[10]). This is a particular issue in systems where coevolutionary adaptations are largely behavioural in nature – because (1) genomic data resolution has only recently advanced to the point where we can scan genomes for such genetic mechanisms, and (2) behavioural ecologists studying these systems have traditionally ignored or simplified the complex genetic basis of behavioural traits (i.e. the "phenotypic gambit"[11]). It is also debated whether plasticity explains observed variation in host responses to parasitism in particular systems[12]. Genetic signals in response to parasitism are important to identify if we are to strengthen the evidence for variable responses not being due to such (e.g. behavioural) plasticity. The successes of genomic approaches in identifying the genetic underpinnings of behavioural variation more generally e.g.[13,14], combined with an awareness that local adaptation can persist despite significant gene flow among populations[15,16], suggests that such an assessment is worthwhile.

99 Avian brood parasitism is a textbook example of antagonistic coevolution, where reciprocal
100 arms-race selection has led to the development of complex behavioural adaptations in brood
101 parasites and their hosts[17-19]. For example, the Reed Warbler (*Acrocephalus scirpaceus*), one
102 of the preferred European hosts of the brood parasitic Cuckoo (*Cuculus canorus*)[20], defends by
103 mobbing the Cuckoo at the nest site[21], or by recognising and ejecting foreign eggs[22]. For their
104 part, Cuckoos evolved host-specific female races (gentes), which mimic their preferred host's
105 eggs with increasing accuracy (see Stokke *et al.* for discussion of the role of crypsis as an
106 alternative to egg mimicry[23]), and so Cuckoo parasitism could also impact host egg colour, as
107 well as nest building and incubation behaviour. The existence of gentes perhaps explains why
108 genetic research has tended to focus on Cuckoos, perhaps at the expense of their hosts[26,27].
109 Despite over 40 years of research into this classic case study for behavioural coevolution[22,24,25],
110 our understanding of the way brood parasitism shapes host defences has been almost
111 completely limited to the phenotypic level[28], meaning that the underlying heritable mechanisms
112 remain unknown. This is a major drawback in terms of characterising a geographic mosaic of
113 coevolution[9], as it is widely recognised that knowledge of the genetic basis of adaptive traits is
114 needed to understand and ultimately predict coevolutionary trajectories[29]. The few existing
115 studies which have attempted to delineate the relationship between brood parasitism and
116 genetics have provided mixed evidence. The first such attempt, by Martín-Gálvez *et al.*[30,31],
117 detected an apparent association between an allele at microsatellite locus Ase64 and egg
118 rejection behaviour in European Magpies (*Pica pica*), parasitised by Great Spotted Cuckoos
119 (*Clamator glandarius*), and this has been used since to argue that anti-parasite behaviours have
120 a genetic basis[32]. Corroborating this, European Magpie populations across Europe vary in the
121 rate at which they reject model eggs depending on both geography and genetics[33]. However, a
122 more recent study, also using microsatellite markers, found no such association in Great Reed
123 Warblers (*Acrocephalus arundinaceus*) parasitised by Cuckoos[34]. The use of higher resolution

124 genetic data is therefore necessary to explore this question further, in light of prior mixed
125 evidence and the need for modern genomic approaches. The Reed Warbler/Cuckoo system is
126 a particularly valuable case study with which to do this, given the generally good understanding
127 of the traits involved in brood parasitism thanks to years of research. Although some studies of
128 parasite–host interactions more generally have found genomic regions putatively associated
129 with non-brood parasitism (e.g. Atlantic Salmon *Salmo salar* and ectoparasites[35]), others have
130 not (e.g. Red Grouse (*Lagopus scotica*) and the nematode *Trichostrongylus tenuis*[36]).
131

132 Reed Warbler and Cuckoo natural history[37,38], breeding behaviour[39,40], egg colouration[41,42],
133 predators[43,44], migration[45,46], taxonomy[47,48], and diet[49,50] are well understood and have been
134 subject to increasingly sophisticated experimental and observational research. The GMTC has
135 been proposed to explain the persistence of brood parasite / host coevolutionary dynamics[7],
136 and much of the existing body of work on these two species provides supporting evidence for
137 this[28,51]. On the other hand, certain aspects vital to the GMTC have so far not been assessed in
138 this system. These include fundamental components such as characterisation of gene flow
139 patterns at a continental scale, and assessment of genomic selection imposed by brood
140 parasitism. The putative geographic mosaic of Reed Warbler/Cuckoo interactions provides a
141 valuable opportunity to evaluate the genomic consequences of variation in brood parasitism
142 risk across a host's breeding range, where breeding habitats are scattered with varying levels
143 of isolation from one another, and varying levels of allopatry/sympatry with Cuckoos of the
144 host-specific gens (i.e. the Reed Warbler-specific race of the Cuckoo). Sites at which Reed
145 Warblers and Cuckoos regularly interact, or coevolutionary 'hot spots', have been monitored
146 for long periods of time e.g.[28]. On the other hand, there are also 'cold spots' where the two
147 species very rarely or never interact (e.g. regions in Finland colonised by Reed Warblers within
148 the past century[52]). The GMTC predicts that this could lead to relaxation or even reversal of

149 anti-parasite defences in such 'cold spots'. Although behavioural defences (such as egg
150 rejection) can rescue hosts that are in sympatry with parasitic Cuckoos, accidental rejection of
151 one's own egg is costly and decreases host fitness[53]. Thus, where they are in allopatry with
152 Cuckoos, Reed Warblers could potentially lose resistance from fitness costs or drift[54]. Re-
153 encounters are predicted to promote cyclic trajectories[7]. Cyclic changes across the Reed
154 Warbler/Cuckoo range mosaic could retain the genetic variation needed for adaptation (as in
155 sticklebacks *Gasterosteus aculeatus*, where metapopulation structuring maintains enough
156 genetic variation to fuel novel colonisations[55]).

157

158 Our work, combining behavioural and genomic information in a well-studied natural system,
159 represents an exciting application of this approach. Using a restriction site associated (RAD)
160 sequencing approach, we characterise gene flow among host populations to assess whether we
161 see patterns of variable gene flow that are a fundamental component of the GMTC.
162 Furthermore, using genotype-environment association analyses, we test for signals of selection
163 on biotic interactions by identifying SNP (single nucleotide polymorphism) loci that associate
164 with Cuckoo parasitism rates as sites across continental Europe, controlling for environmental
165 (i.e. climatic) factors. Overall, this study provides further evidence for the existence of a
166 geographic mosaic of coevolution in an avian-cuckoo system, and highlights where future
167 efforts could focus to provide a more complete characterisation of the GMTC.

168

169

170

171

172

Material and methods

Data collection

Reed Warblers' anti-Cuckoo defences vary in extent according to the level of Cuckoo parasitism the population is exposed to[54]. The species' range is also expanding to the north[52] and, at least in certain regions, to the south[56]. Our sampling sites form a latitudinal and longitudinal gradient across much of Europe, and include locations with known range expansion histories. Birds were caught and blood sampled as part of prior projects[57-59]. The inclusion of Reed Warblers from sites across much of their European breeding range enables comparisons among sites which vary in parasitism level. It also reduces the impact of other genetic or demographic effects (e.g. allele surfing[60,61]) as well as from selection imposed by other forms of environmental variation that commonly impact genomic evolution (e.g. climate[62]), on our inferences.

We collated spatial data on brood parasitism rate, making use of a combination of existing literature[63] and information provided by academic and citizen scientist collaborators. For some sites, we also had access to the rate at which Reed Warblers at that site reject a 'foreign' egg from their nest (Supplementary Table S1). All 'parasitised' sites had a parasitism rate (i.e. percentage of parasitised nests) of > 5%. All 'unparasitised' sites had an observed parasitism rate of 0% during the period in which the samples were collected (although parasitism may occur very rarely – e.g. in Helsinki, an 'unparasitised' Reed Warbler population, two nests were found to be parasitised by a Cuckoo in 2023, out of 667 nests monitored over six years (pers. obs. RT, KR, EK, WJS)). For two sampling sites, parasitism status was unknown (Figure 1A).

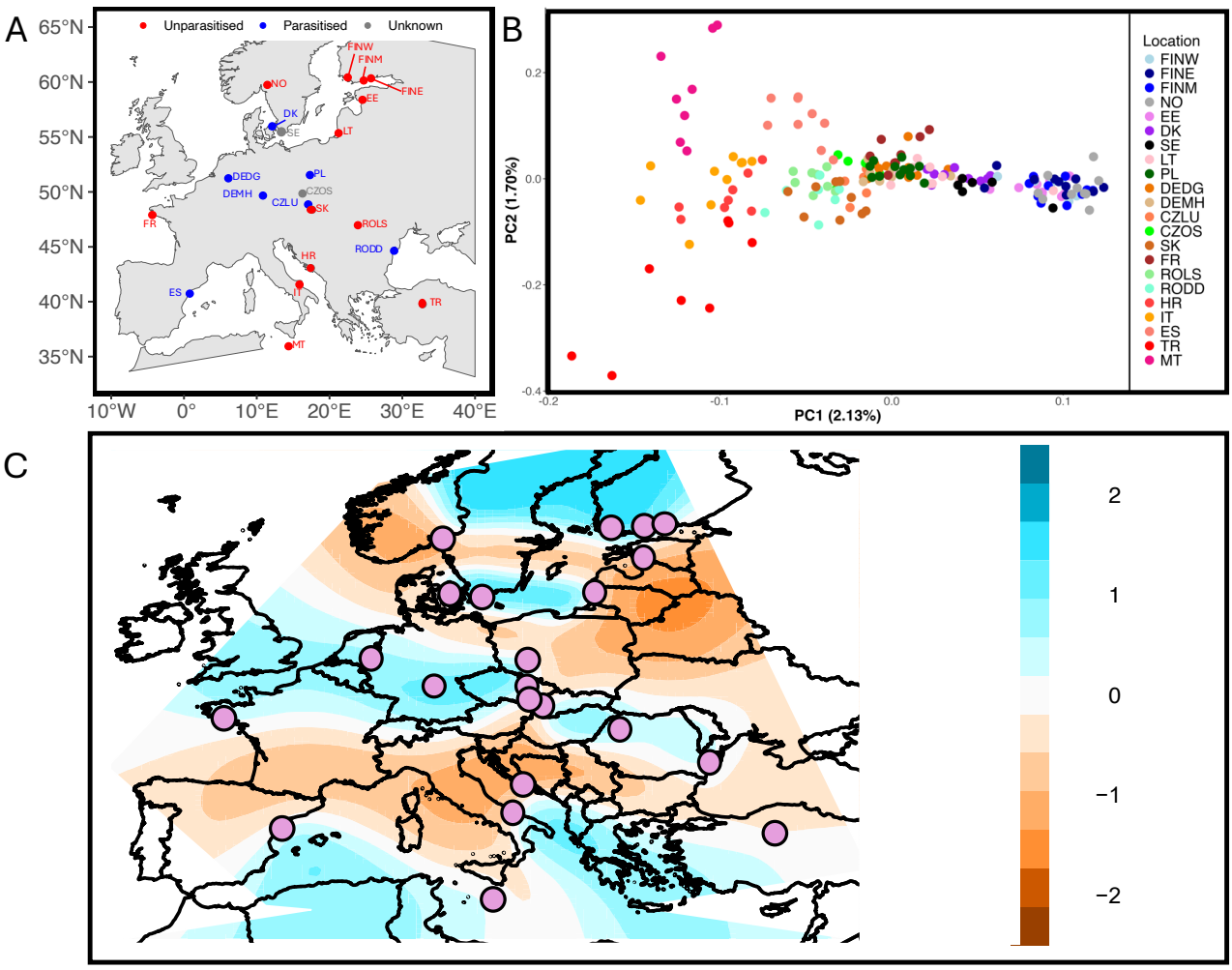


Fig. 1: (a) We used DNA samples from Reed Warblers at 22 sites. Red points represent sites where Reed Warblers are not parasitised by Cuckoos. Blue points represent locations where they are. Grey points represent sites where brood parasitism rate is not known. Location codes are indicated (see Supplementary Table S1). **(b)** Principal Component Analysis reveals genetic differences across the Reed Warbler range. PC1 shows a north-south gradient of genetic differences, and PC2 shows differentiation amongst southern European birds. The variance explained by PC1 and PC2 is shown in parentheses. Colours of points depict sampling regions. Sampling regions are ordered latitudinally in the figure legend. Location codes correspond to sites depicted in (A). **(c)** Effective migration surfaces inferred by EEMS. Orange regions have reduced gene flow, and blue regions have higher gene flow. The scale represents effective

208 migration rate on log10 scale, relative to overall migration. A value of '1' denotes effective
209 migration that is 10 times faster than average. Gene flow is higher in central Europe, with lower
210 gene flow occurring in the Mediterranean and Baltic regions. Sampling locations are indicated.

211

212 *Genomic data generation and pre-processing*

213

214 Genomic DNA extraction and RAD-sequencing was carried out as detailed in prior work[57,58].
215 We generated and filtered two genotype files, one for use in population genetic structure and
216 gene flow analyses, and another for use in genotype-environment analyses (GEA). Both files
217 included both autosomes and sex chromosomes. For both datasets, we scanned for individuals
218 with high relatedness (see[64]) (none of the pairs of individuals had relatedness of > 0.1, and so
219 no individuals were excluded). To calculate this information, we used VCFtools v0.1.17[65] --
220 relatedness. We split Finnish samples into three groups called 'West', 'Middle', and 'East'
221 Finland. The subsequent 'Dataset 1' included 188 individuals from 22 sampling sites. For the
222 genotype file used in GEA pipelines, 'Dataset 2', we excluded sampling sites for which there
223 was no brood parasitism data (Sweden and one of the Czech sites) and split the Finnish samples
224 in the same way as in Dataset 1. Dataset 2 included 177 individuals from 20 sampling sites.
225 Full details on the genomic datasets are available in Supplementary Table S1.

226

227 For both Dataset 1 and Dataset 2, we used STACKS v2.65[66] to sort, filter, and demultiplex
228 reads with the *process_radtags* script. We aligned reads to the bAcrSci1 reference genome[67]
229 with bwa v0.7.17[68]. We used SAMTOOLS v1.18 to sort files[69]. Following this, we used the
230 *ref_map.pl* pipeline to build the initial catalogue to be used as an input for *populations*. Using
231 VCFtools, we filtered by depth, retaining sites with mean depth of more than or equal to 10,

232 and less than or equal to 100. We then assessed sites for Hardy-Weinberg Equilibrium (using
233 an exact test (--hwe), see[70]) and filtered to exclude SNPs out of HWE in all populations with
234 more than ten individuals. Using *populations*, we selected sites where all individuals were
235 represented (-R 1), where there was a minimum minor allele count of 3 (--min-mac 3), and
236 where there was a maximum observed heterozygosity of 0.75 (--max-obs-het 0.75). We
237 restricted analysis to the first SNP of each locus (--write-single-snp). For Dataset 1, we used
238 PLINK v2.00[71] to detect SNP pairs with linkage disequilibrium of 0.5, and then exclude one
239 of them (--indep-pairwise 50 5 0.5), retaining 13,208 SNPs ('Dataset 1') for our analyses. For
240 Dataset 2, we did not filter for linkage disequilibrium as this would risk removing parts of the
241 genome under strong selection. Hence, we retained 13,346 SNPs.

242

243 *Population genetic structure and gene flow analyses*

244

245 We first carried out PCA with PLINK v2.00[71], as well as ADMIXTURE v1.3.0[72], ranging the
246 number of assumed populations between 1 and 22. We then examined gene flow among
247 sampling sites. To do so, we used EEMS v0.0.0.9, which makes use of geo-referenced genetic
248 samples to detect regions of higher- and lower-than-average gene flow and has been simulated
249 to accurately recover effective migration surfaces (visualised gene flow) in scenarios with
250 complicated histories of gene flow[73]. We first converted the Dataset 1 VCF into PLINK input
251 files using the –make-bed command. We then used the bed2diffs_script to convert these into a
252 matrix of average pairwise genetic dissimilarities, which was used as the input file for EEMS.
253 Following this, we carried out 8 runs each with an MCMC chain of 10,000,000 iterations, with
254 a 20,000-iteration burn-in and 1800 thinning iterations. The number of geographic demes was
255 set to 100, 300, 500, 700, 900, 1100, 1300 and 1500, and runs were combined using

256 rEEMSplots in R. Data is presented on a $\log_{10}$ scale, where an increase of one represents a
257 tenfold increase in migration rate relative to average levels. We also used VCFtools to output
258 mean pairwise F_{ST} values for each population pair.

259

260 *Genotype-environment analysis*

261

262 To explore genotype-environment associations we used three different GEA analytical
263 frameworks: univariate latent factor mixed modelling (LFMM)[74], multivariate redundancy
264 analysis (RDA) and partial redundancy analysis (pRDA)[75], and the F_{ST}-based method
265 BayeScEnv[76]. We used the vcf2geno command from the R package *LEA*[77] to convert the
266 relevant VCF files into suitable input files. Finally, we estimated additive polygenic scores to
267 determine the cumulative signal of selection across candidate loci identified across all GEA
268 methods as in[78,79]. Parasitism rate tends to be highest in the centre of Europe, and lowest in its
269 northern and southern parts[80]. To summarise climatic variation, we downloaded the 19 BioClim
270 climate variables[81]. To avoid high correlations amongst the variables, we calculated Pearson's
271 correlation coefficient in R and selected four variables which had correlation coefficients of <
272 0.7 across our sites. These were temperature seasonality (BIO4), mean temperature of the
273 wettest quarter (BIO8), precipitation of the wettest month (BIO13), and precipitation
274 seasonality (BIO15). The latter was particularly important to include as it was identified as an
275 important contributor to genetic variation in European populations of the Reed Warbler in prior
276 work[57]. Measures of precipitation are important to include given the significance of rainfall in
277 determining water level and therefore reedbed growth (i.e. Reed Warbler habitat availability
278 and quality[82]). We also included latitude in our models, to account for neutral geographical

279 variation. Both LFMM and RDA/pRDA are run with parasitism rate, climatic variables, and
280 latitude together. BayeScEnv is run individually for each variable.

281

282 We carried out LFMM with the *lfmm* v2 package in R[83]. This uses an MCMC algorithm to
283 explore associations between allele frequency and environmental variables. We included
284 parasitism status, our four selected BioClim variables, and latitude as our environmental
285 variables. We then ran LFMM using *lfmm_ridge* to estimate the model, and *lfmm_test* to assess
286 the significance of the SNPs. The number of latent factors was chosen based on our population
287 genetic structure analyses (allowing us to control for neutral genomic processes e.g. due to
288 colonisation history). We selected K = 1 as we inferred limited differentiation across Europe
289 from the PCA, ADMIXTURE, and F_{ST} results. We then calculated the genomic inflation factor
290 (GIF)[84,85]. We confirmed that GIF values were close to 1, demonstrating that population
291 structure was adequately controlled for. LFMM then adjusts the p-values by the GIF values,
292 and we applied the Benjamini-Hochberg procedure[86] to tolerate a false discovery rate of 10%,
293 to return significant candidate loci.

294

295 Following LFMM, we carried out RDA, which models relationships between environmental
296 predictors and genomic variation (i.e. the SNPs). RDA is a constrained ordination and has been
297 demonstrated to outperform other GEA analyses when selection acts over multiple loci[75]. We
298 carried out RDA using the *vegan* R package's *rda* function. We calculated adjusted R^2 (based
299 on the number of predictors), before assessing variance explained by each eigenvalue. We
300 assessed the RDA model for significance using F-statistics via an analysis of variance
301 (ANOVA) with 999 permutations. We assessed both the full model and each constrained axis.
302 We also assessed the Variance Inflation Factors of each predictor variable, confirming that they

303 are below ten, and therefore that collinearity amongst predictors is not a significant problem.
304 We identified candidate SNPs on significant constrained axes based on their loadings in
305 ordination space, using a cut-off of ±3 SD (see[87]), and removed duplicate candidates. We
306 repeated this in a 'partial RDA' context, which accounts for population genomic structure.
307 Applying both RDA approaches provides a more holistic assessment of SNPs under selection
308 according to variation in levels of brood parasitism.

309

310 Finally, we used BayeScEnv v1.1[76]. BayeScEnv does not directly assess associations between
311 allele frequencies and environmental variables, but rather environmental and genetic
312 differentiation (i.e. F_{ST}), providing an additional GEA test. We used VCFR[88] to convert the
313 VCF file to a genind object (of the adegenet R package[89]), and then used hierfstat[90] to convert
314 this into an input file. We divided parasitism rate at each site by the standard deviation, as
315 recommended by the BayeScEnv authors[76]. We then ran BayeScEnv with twenty pilot runs of
316 length 5000, a thinning interval size of ten, 5000 outputted iterations, a burn-in length of
317 50,000, prior probability against the neutral model of 0.5, and prior preference for the locus-
318 specific effect model of 0.5. We identified SNPs with a posterior error probability (see[91,92]) of
319 < 0.01 as candidate loci. We repeated BayeScEnv using the four Bioclim variables and latitude
320 as the focal parasitism variable, to identify any common candidate loci.

321

322 We filtered the annotation file for the bAcrSci1 genome to identify candidate genes from our
323 outlier SNPs (i.e. SNPs that fall within a gene in the annotation file). Then, we (1) performed
324 Gene Ontology Enrichment Analysis (geneontology.org, Panther v18.0), with *Gallus gallus*
325 (the most complete avian genome annotation) as the reference set to identify whether genes
326 associated with specific biological processes were overrepresented, and (2) delineated the

cumulative impact of selection at the individual level, by calculating per-individual additive polygenic scores (estimates of the cumulative signal of selection across candidate loci; referred to henceforth as 'polygenic scores') (see[78,79]). We assessed the correlation between the polygenic scores and parasitism rate to determine the extent to which the candidate alleles vary with parasitism. We tested linear and quadratic models and selected the model with the lowest Bayesian information criterion (BIC) value. We also assessed the correlation of polygenic scores with the bioclimatic variables and latitude, as well as the rate at which Reed Warblers at each site reject a foreign egg, again testing both linear and quadratic models.

349 **Results**

350

351 ***Population structure and gene flow analyses***

352

353 PCA reveals genetic differentiation with geography, albeit without discrete groupings (Figure
354 1B) and overall low differences between populations (corroborating F_{ST} results; Supplementary
355 Table S2). PC1 reflects differentiation along a north-south gradient, whereas PC2 shows
356 differentiation within the southern European individuals. In particular, Turkish and Maltese
357 birds differ from those elsewhere in southern Europe. ADMIXTURE highlights similar
358 patterns. Although K = 1 is best supported, higher K values highlight some population
359 structure, with K = 2 splitting birds from Nordic regions from those from southern Europe, and
360 characterising birds from within the range core as genetically admixed individuals (Figure S1).
361 Higher K values identify genetic structure amongst the southern European birds. Gene flow
362 analysis depicts a varying degree of gene flow throughout Europe, with higher gene flow across
363 the core of the continent flanked by areas of lower gene flow in both the north and south (Figure
364 1C).

365

366 ***Genotype-environment associations***

367

368 LFMM identified 11 candidate SNPs associated with parasitism rate (Supplementary Table S3)
369 and 129 SNPs associated with the other environmental variables (i.e. latitude or one of the four
370 bioclimatic variables). Whilst there was some SNP overlap amongst the five environmental
371 variables, the candidate parasitism SNPs were not identified as associated with any of these
372 other variables. The RDA model, as a global model, was significant to $p = 0.001$, and explained

373 0.9% of variance in the dataset (n.b. variance explained is typically low in RDA models
374 assessing the genetic impacts of environmental variation). The variance inflation factor (VIF)
375 between predictor variables was < 5, suggesting limited impact of collinearity on the RDA.
376 The first five axes were significant at p = 0.001 and together explained 86.03% of variation.
377 Using these five significant axes for outlier detection, we identified 41 SNPs as being
378 associated with parasitism, 86 with latitude, 30 with BIO4, 14 with BIO8, and 17 with BIO13
379 (Figure 2A). Partial RDA was also significant as a global model (p = 0.001) and explained
380 0.05% of the variance. The first four axes were significant to p = 0.001 and together explained
381 69.91% of variation. From this method, we identified 43 SNPs associated with parasitism, 78
382 with latitude, 29 with BIO4, 10 with BIO8, and 9 with BIO13 (Figure 2B). A total of 37
383 parasitism candidate SNPs were identified by both RDA and pRDA. Four parasitism candidate
384 SNPs were unique to RDA, and six to pRDA. Whilst the RDA/pRDA may be somewhat
385 confounded by BIO8 and BIO13 (Figure 2), all of the 11 candidate SNPs identified by LFMM
386 (which tested each variable in a univariate manner and did not identify overlap with the
387 parasitism SNPs) were also identified by both RDA and pRDA. Finally, BayeScEnv detected
388 2 parasitism candidate SNPs, and 26 SNPs associated with latitude or a bioclimatic variable.
389 One of the two parasitism candidate SNPs was also a candidate associating with BIO8 and
390 BIO15 (and there was extensive overlap amongst SNPs associating with the five environmental
391 variables, with 10 of the 26 being candidates for more than one environmental variable).
392 Neither of the BayeScEnv parasitism loci were identified by either LFMM or RDA/pRDA.
393 Together, LFMM, RDA/pRDA, and BayeScEnv detected 49 candidate SNPs (Figure 2C). We
394 identified a known gene associated with 24 of these candidate SNPs.

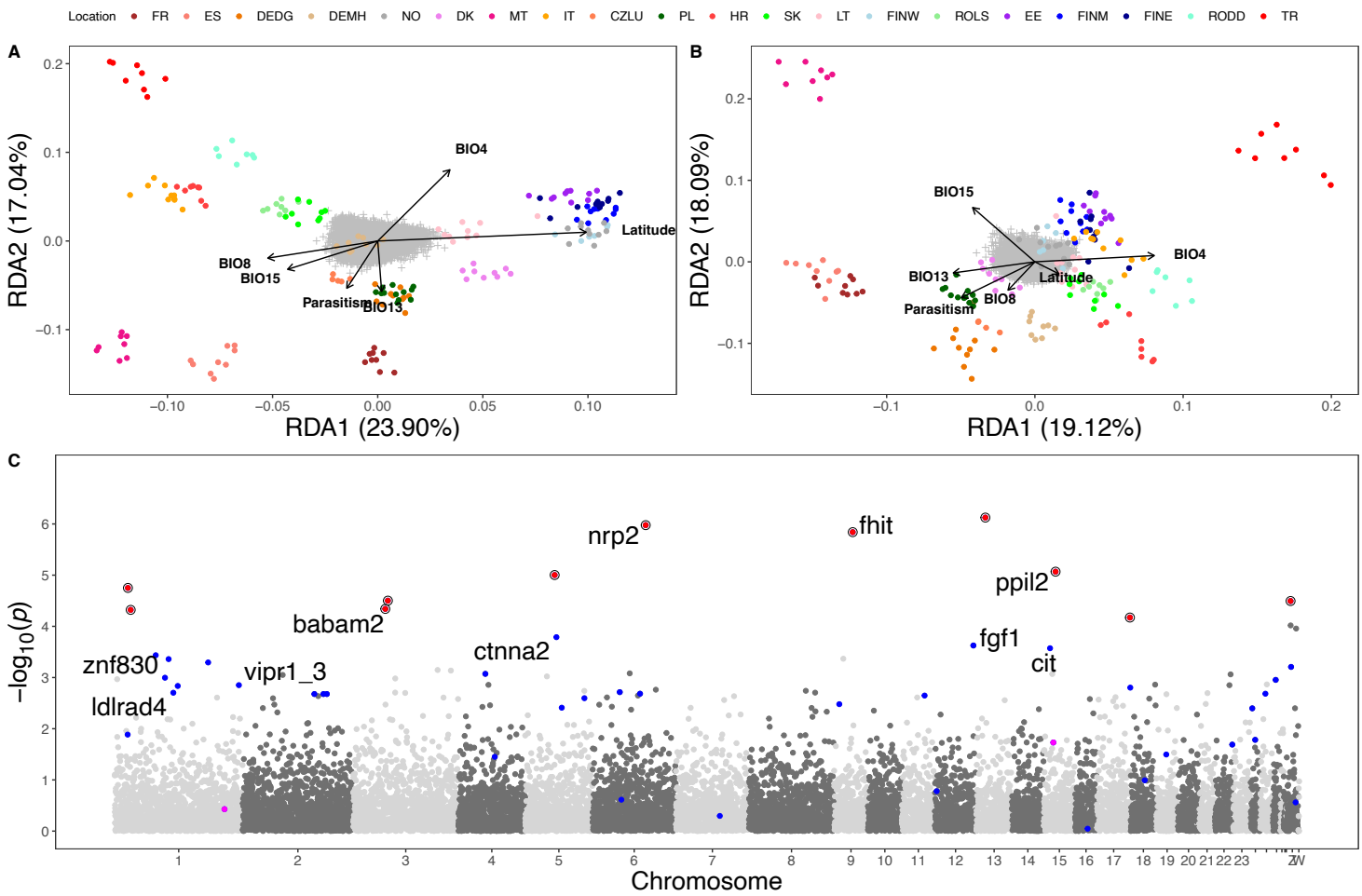
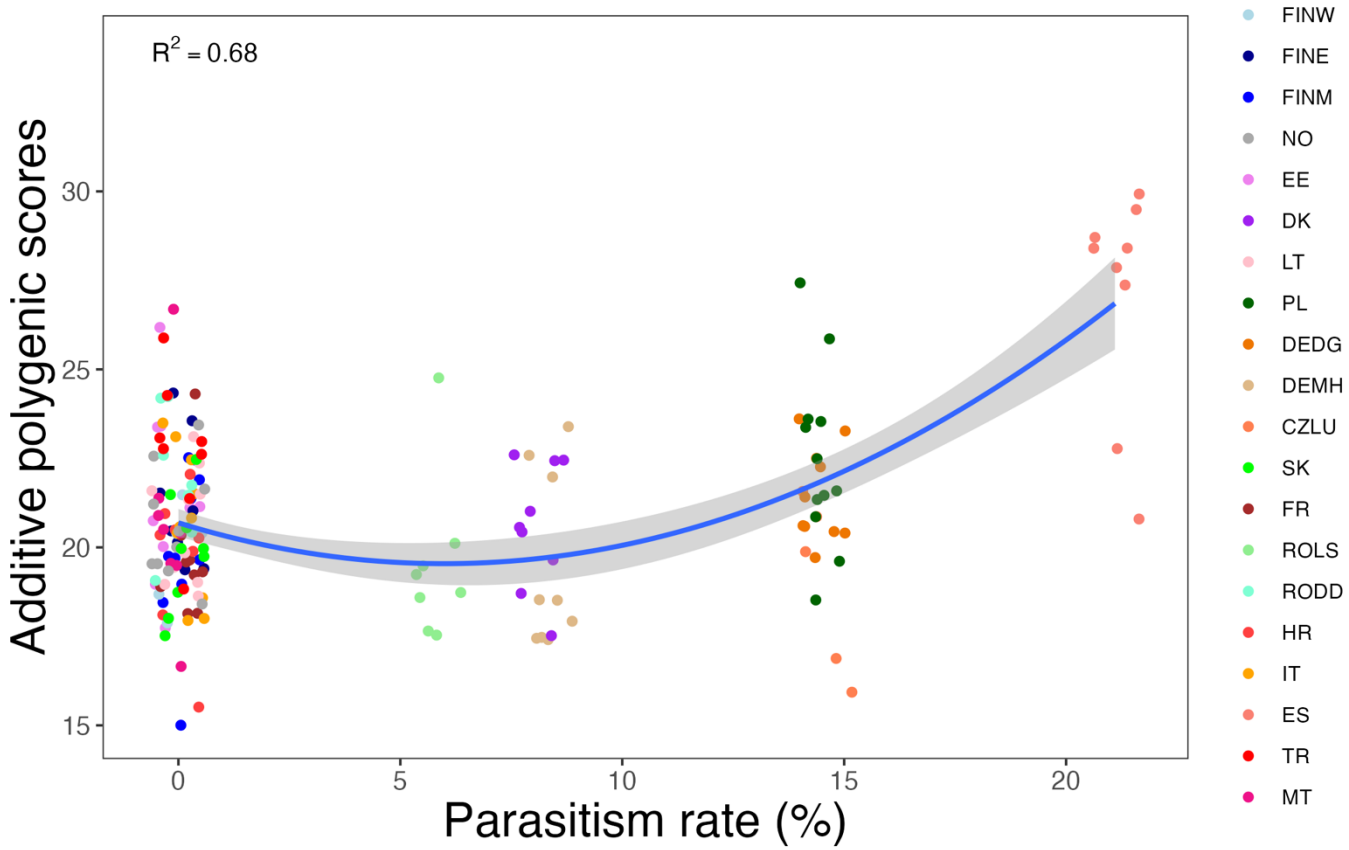


Fig. 2: Genotype–environment association with brood parasitism. **(a)** Redundancy analysis (RDA) using 13,346 SNPs, depicting the first two axes and using the following variables: parasitism, latitude, BIO4 (temperature seasonality), BIO8 (mean temperature of the wettest quarter), BIO13 (precipitation of the wettest month), and BIO15 (precipitation seasonality). SNP/individual scores are scaled by the square root of the eigenvalues to give appropriate weight to the axes based on their contribution to the model. SNP genotypes are represented as grey crosses, and the direction of variation in each constraining variable is indicated with black arrows (arrow length represents the strength of the correlation of that variable). Individuals are represented by different colours according to location. Percentages on the axes indicate the proportion of total variance explained. **(b)** Partial redundancy analysis (pRDA), accounting for population genetic structure. **(c)** Manhattan plot showing genotype–parasitism associations,

407 estimated with LFMM, across the genome. The red dot indicates the candidate SNPs identified
408 by LFMM itself (as well as RDA and pRDA). Blue dots represent the candidate SNPs identified
409 by RDA and/or pRDA. Pink dots represent candidate SNPs identified by BayeScEnv. Amongst
410 the candidate SNPs, those with the top 11 LFMM p-values are circled, and the top ten candidate
411 SNPs which are associated with known genes are annotated. Full gene names are presented in
412 Supplementary Table S3.

413

414 Gene ontology analysis of the SNPs associated with parasitism rate highlighted genes with a
415 diverse array of functions and associations, with overrepresentation of genes related to cell
416 communication, luteinising hormone secretion, tissue morphogenesis, and the cell surface
417 receptor protein tyrosine kinase signalling pathway. Specific genes (Supplementary Table S3)
418 included those with neurological (e.g. *nrp2*, *ctnna2,* rtn4rl1), metabolic (e.g. *cit*), and cell-cycle
419 related (e.g. *babam2, znf830*) associations. The additive polygenic scores for the 49 candidate
420 loci increased with parasitism rate at the location at which the individual was sampled (n = 177
421 individuals). Both the linear model ($F_{(1,175)}$ = 178.9, adjusted R^2 = 0.5027, $p < 0.05$) and the
422 quadratic model were significant ($F_{(2,174)}$ = 188, adjusted R^2 = 0.68, $p < 0.05$). The quadratic
423 model explained more variation than the linear model and had a lower BIC score (linear model
424 = 868.0, quadratic model = 794.2, ΔBIC = 73.86) (Figure 3).



425

426 **Fig. 3:** Correlation of brood parasitism rate and individual polygenic scores (an estimate of the
427 cumulative signal of selection across all the candidate loci). Each dot represents the polygenic
428 score for an individual Reed Warbler and the parasitism rate at its location. The solid line
429 depicts the regression line from the quadratic model. Points are jittered for clarity. Sampling
430 location is indicated by point colour (see Figure 1 for more information on sampling sites).

431

432 For all other models (looking at the variables other than parasitism, including rejection rate)
433 (Figure S2), the linear model was chosen as the BIC difference could not separate the quadratic
434 and the linear models, so the simpler was chosen. R-squared values for all of them were an
435 order of magnitude or more lower than for the relationship for parasitism.

436

437

438 **Discussion**

439

440 Our analyses provide two novel pieces of information which contribute to our understanding
441 of the Reed Warbler–Cuckoo mosaic of coevolution. Firstly, we show that the level of gene
442 flow varies within the European breeding range of the Reed Warbler. Generally, gene flow is
443 higher within the range core and lower in the north and south of our sampling area – which is
444 probably attributable to the Baltic and Adriatic seas acting as barriers to gene flow. This
445 variation in gene flow also corresponds to general patterns in the intensity of brood parasitism,
446 wherein parasitism rate is generally higher in the host's range core when compared to the range
447 edge. Principal component analysis suggests some isolation by distance between core and edge
448 populations, however the edge populations cluster as three connected groups in the PCA,
449 largely corresponding to their geographical location. Differentiation of Turkish individuals
450 could be related to subspecies-level admixture in this region[93]. Prior work, using microsatellite
451 data, also identified restricted gene flow in southern Europe[94], corroborating our results. Low
452 gene flow to the west of Romania might be attributable to the known migratory divide in this
453 species[95]. Whilst Reed Warblers lack significant population genetic structure, they are known
454 to have some geographic differentiation[47,57,96] (in comparison to many Palearctic bird species,
455 which are panmictic[97]). This may facilitate the loss, gain or maintenance of genetic variation
456 involved in responses to Cuckoo brood parasitism, allowing adaptation to parasitism to exhibit
457 spatial variation at a continental scale. This would allow for a geographic mosaic of coevolution
458 to be maintained.

459

460 Our second novel piece of evidence for a geographic mosaic of coevolution is the identification
461 of genetic associations with Cuckoo parasitism. Using three different types of genotype-
462 environment association analysis, we have identified candidate genes under selection that

463 highly correlate with parasitism. We have therefore described, for the first time, a continental
464 signature of genomic adaptation to brood parasitism pressure. Four of the identified genes (and
465 11 of the identified SNPs) were detected by multiple method types. The candidate genes
466 represent a diverse suite of physiological functions, as assessed by gene ontology analysis.
467 Among those previously studied in birds, several genes stand out as particularly interesting.
468 Notably, *nrp2*, which was identified as correlated with parasitism rate by RDA, partial RDA,
469 and LFMM, is associated with neurodevelopmental functions (potentially relating to vocal
470 learning in birds [98] and also to memory and learning in laboratory mice[99]). More generally,
471 *ctnna2* is known to have a major role in the function and growth of neurons[100]. These findings
472 suggest a potential role for these genes in learning or communication, or at least some cognitive
473 response, in the defence against Cuckoo parasitism (perhaps relating to mobbing for example,
474 which is a key defence against Cuckoos at Reed Warbler nest sites[21]). Other candidate genes
475 have less obvious links to parasite responses. For example, *ppil2* is associated with immunity
476 (and was identified as involved in immune system changes during chicken moult[101]). Note that
477 there were no SNPs near the Ase64 microsatellite locus (found to be associated with egg
478 rejection capability in magpies[30]), precluding us from testing whether this region is associated
479 with parasitism rate in our study species. The closest SNP in our dataset was 7670 bp away,
480 and was not among the parasitism candidate SNPs.
481
482 Despite the associations found here, there are several caveats which are important to highlight.
483 Firstly, correlation does not necessarily mean causation. There will undoubtedly be various
484 abiotic and biotic correlates of parasitism in the Reed Warbler–Cuckoo system. Whilst we have
485 explored the potential confounding role of climatic variation (e.g. none of focal environmental
486 variables correlate with the parasitism polygenic scores), a well-known biotic correlate of
487 parasitism in this system is the increase in parasitism rate with host population density[63]. Our

488 candidate SNPs might therefore be impacted by selection according to local conspecific
489 density, or some correlate of this (e.g. disease, intraspecific competition). It is also known that
490 habitat structure plays a role in determining host defences and parasite adaptations[102,103]. The
491 identification of genes which are putatively involved in immunity and growth provides support
492 for this. Disentangling such factors represents a major challenge. Common garden experiments,
493 which have proven useful in studying the geographic mosaic theory of coevolution in plants
494 (e.g.[104,105]) are impractical for wild bird populations. A feasible alternative might involve
495 characterising the responses of individual Reed Warblers to Cuckoos (e.g. extent of mobbing
496 behaviour or egg rejection), and then using genomic data for these individuals to assess
497 differentiation at our candidate regions and across the genome. This would also allow for
498 characterisation of differential selection on male and female hosts. If carried out both within
499 and between different sites that differ in parasitism rates across the Reed Warbler's European
500 range, this could provide additional evidence for the role of these candidate genes in mediating
501 Reed Warbler–Cuckoo interactions. A second caveat is that, given that RAD sequencing data
502 covers only a small fraction of the genome, many of the candidate genes are likely to be in
503 linkage with genes that are adaptively responding to parasitism rate, rather than having direct
504 mechanistic roles. A cautionary example is provided by the gene responsible for the black coat
505 of some North American Grey Wolves (*Canis lupus*), which does not determine immunological
506 differences as was previously thought, but is linked to immunological genes[106]).
507
508 Whilst there is limited value in speculating about the exact functions of our candidate genes
509 and their potential links to parasitism, it is worth outlining potential pathways or adaptations in
510 which we would expect parasitism to underpin genetic differences. For example, genetic
511 differences could relate to the visual systems of the Reed Warbler and therefore contribute to
512 egg-rejection capacity, which is known to be a major component of the Reed Warbler's

513 antiparasitic defences[107]. There might also be a genetic underpinning to the willingness of Reed
514 Warblers to mob Cuckoos (perhaps relating to varying aggression amongst individuals). The
515 extent of mobbing is known to vary across Europe[21]. Nesting behaviour (i.e. how well-hidden
516 nests are) and phenology (i.e. timing of laying) are also likely to have a genetic component and
517 would have an impact on parasitism likelihood. Our parasitism polygenic scores do not
518 correlate with known rates of rejection of foreign eggs across sampling sites (Figure S2), so it
519 seems unlikely that these genes impact this particular anti-cuckoo behaviour. Overall, it is very
520 probable that all of these axes of phenotypic variation and others would be highly polygenic
521 and therefore will require further research to parse the contributions of different genes, and,
522 more broadly, the respective contributions of genetics and environment to parasitism response.
523
524 Whilst our results provide support for a genomic association with brood parasitism in the Reed
525 Warbler–Cuckoo system, a more complete characterisation of the geographic mosaic of
526 coevolution will require further research. Although our reduced-representation based
527 assessment provides a significant advance on attempts using microsatellite markers, analysis
528 of whole-genome sequence data in the Reed Warbler and other hosts would be invaluable in
529 providing a more fine-grain assessment of the architecture of genomic selection. This is
530 because it is well-known that using RAD-sequence data to infer local adaptation may lead to
531 candidate loci being missed[108]. As well as this, we have highlighted a need for better
532 understanding of parasitism status within the range of this host and others, given that some
533 locations do not have a known parasitism rate. Furthermore, the impact of human activity has,
534 in recent centuries, changed landscapes across Europe[109,110], likely affecting distributions and
535 abundance of the Cuckoo and its hosts. It would be fascinating to delineate the genomic impacts
536 of such changes, perhaps using museum specimens. Furthermore, attempting to extend this
537 type of study to the Cuckoo itself would be worthwhile. The existence of gentes, each

538 potentially subject to differing selection pressures, adds complexity to the Reed Warbler–
539 Cuckoo coevolutionary mosaic. Might we expect different genomic regions to be impacted by
540 selection in different Cuckoo lineages (gentes), due to variation in parasitism pressures across
541 host populations? Sampling of Cuckoos and their hosts across varying habitats and regions, at
542 continental scale, will be needed to resolve this.
543
544 Our characterisation of this antagonistic coevolutionary interaction represents the most in-
545 depth assessment of the genetic underpinnings of brood parasitism carried out to date. A similar
546 approach – using estimates of interaction strength – could hold promise in elucidating biotic
547 interactions more generally. For example, in tight predator–prey relationships, predation rate
548 could be used in much the same way as we have used brood parasitism. Whilst studies of
549 parasite–host interactions have made use of in-depth population genomic assessments since the
550 advent of widely available whole-genome data[35,36], population genomic research that has
551 involved the study of predation (and competition) has historically focused on understanding
552 changes in overall levels of genetic variation[111]. Genotype–environment analysis would,
553 however, allow for the identification of candidate genes involved in this ecologically
554 fundamental biotic interaction, if combined with a recognition that genes involved in
555 behavioural expression are likely to work in a polygenic fashion, meaning that their individual
556 effects are small. Such approaches are being increasingly used as part of wider population
557 genetic studies (e.g.[112]) (as well as gene expression-based assessments[113] or experimental
558 approaches[114] which have been deployed in laboratory settings). More broadly, the application
559 of genomic research to study behaviour in well-studied systems such as humans[115] and mice[116]
560 is developing rapidly. Nevertheless, genotype–environment analyses deserve specific attention
561 in non-model systems in the wild, especially in tightly linked predator–prey species pairs which
562 are known to experience cyclical demographic changes. Overall, our work highlights that, even

with traits which are likely to be highly polygenic or involve significant plastic elements, genotype–environment analyses offer opportunities to significantly improve our understanding of the genetic underpinnings of biotic interactions, especially when deployed alongside a thorough grounding in the natural history and field biology of the species in question.

In conclusion, this investigation represents a significant step forward in understanding the geographic mosaic of coevolution in the Reed Warbler–Cuckoo system. We not only provide support for variable gene flow across the European range of the Reed Warbler, but also, for the first time, identify putative signature of genomic selection exerted by brood parasitism. This work highlights the need for additional characterisation of the genetic and environmental factors responsible for this antagonistic coevolutionary interaction.

Acknowledgements: We are grateful to those who supplied information on egg rejection and brood parasitism rates in prior works. These include Geir Rudolfsen, Inge Hafstad, Bruno Bargain, Josef Beier, David Bigas Campàs, Bernd Leisler, Péter L. Pap, Ričardas Patapavičius, Karl Schulze-Hagen, Johannes Bang, Bernd Leisler, Jesper J. Madsen, Anders Pape Møller, Andrzej Dyrcz, Marcel Honza, Vytautas Jusys, Eivin Røskaft, and Arne Moksnes. We acknowledge the use of the CSC – IT Center for Science, Finland, for computational resources. BGS was funded by the Research Council of Norway, grant no. 151641/432. The work was funded by Academy of Finland project grant number 333801 and HiLIFE Fellows grant to RT.

Ethics statement: All blood sampling was conducted under permission from animal ethics experimentation boards and with permission to derogate from wildlife protection laws in respective countries. In Finland these were granted by the Project Authorization Board of the Regional State Administrative Agency (ESAVI/7857/2018) and the Centre for Economic

Development, Transport and the Environment (VARELY/758/2018). See references 57-59 for a list of permissions for other locations.

Data availability statement:

Data and code will be made available on publication.

Author contributions:

Conceptualisation: Rose Thorogood, Katja Rönkä, William J. Smith. **Data curation**: William J Smith, Katja Rönkä, Rose Thorogood, Nora Bergman. **Formal analysis**: William J. Smith, Katja Rönkä. **Funding acquisition**: Rose Thorogood. **Investigation**: William J. Smith, Katja Rönkä, Michał T. Jezierski. **Methodology**: William J. Smith, Kata Rönkä, Michał T. Jezierski. **Project administration**: William J. Smith, Katja Rönkä, Rose Thorogood. **Resources**: Rose Thorogood, Katja Rönkä, Nora Bergman, Edward Kluen, Petr Procházka, Bård Stokke, Bengt Hansson. **Supervision**: Rose Thorogood. **Validation**: William J. Smith, Katja Rönkä, Michał T. Jezierski. **Visualisation**: William J. Smith, Michał T. Jezierski, Katja Rönkä, Rose Thorogood. **Writing – original draft**: William J. Smith, Katja Rönkä, Rose Thorogood. **Writing – review & editing**: William J. Smith, Katja Rönkä, Staffan Bensch, Nora Bergman, Rachael Y. Dudaniec, Fabrice Eroukhmanoff, Frode Fossøy, Bengt Hansson, Michał T. Jezierski, Edward Kluen, Petr Procházka, Camilla Lo Cascio Sætre, Bård G. Stokke, & Rose Thorogood.

References

1 Paterson, S. *et al.* Antagonistic coevolution accelerates molecular evolution. *Nature* **464**, 275-278, doi:10.1038/nature08798 (2010).

2 Yamamichi, M. & Ellner, S. P. Antagonistic coevolution between quantitative and Mendelian traits. *Proc Biol Sci* **283**, 20152926, doi:10.1098/rspb.2015.2926 (2016).

3 Laine, A-L. Role of coevolution in generating biological diversity: spatially divergent selection trajectories. *Journal of Experimental Botany* **60**, 2957-2970, doi:10.1093/jxb/erp168 (2009).

4 Ebert, D. Open questions: what are the genes underlying antagonistic coevolution? *BMC Biol* **16**, 114, doi:10.1186/s12915-018-0583-7 (2018).

5 Thomas, M. A. & Klaper, R. Genomics for the ecological toolbox. *Trends in Ecology & Evolution* **19**, 439-445, doi:10.1016/j.tree.2004.06.010 (2004).

6 Živković, D., John, S., Verin, M., Stephan, W. & Tellier, A. Neutral genomic signatures of host-parasite coevolution. *BMC Evolutionary Biology* **19**, 230, doi:10.1186/s12862-019-1556-3 (2019).

7 Thompson, J. N. *The Geographic Mosaic of Coevolution*. (University of Chicago Press, 2005).

8 Gomulkiewicz, R., Thompson, J., Holt, R., Nuismer, S. & Hochberg, M. Hot Spots, Cold Spots, and the Geographic Mosaic Theory of Coevolution. *The American naturalist* **156**, 156-174, doi:10.1086/303382 (2000).

9 Gomulkiewicz, R. *et al.* Dos and don'ts of testing the geographic mosaic theory of coevolution. *Heredity (Edinb)* **98**, 249-258, doi:10.1038/sj.hdy.6800949 (2007).

10 Kaur, R. *et al.* Ant behaviour and brain gene expression of defending hosts depend on the ecological success of the intruding social parasite. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, 20180192, doi:10.1098/rstb.2018.0192 (2019).

11 Springer, S. A., Crespi, B. J. & Swanson, W. J. Beyond the phenotypic gambit: molecular behavioural ecology and the evolution of genetic architecture. *Molecular Ecology* **20**, 2240-2257 (2011).

12 Lindholm, A. K. & Thomas, R. J. Differences between Populations of Reed Warblers in Defences against Brood Parasitism. *Behaviour* **137**, 25-42 (2000).

13 Lank, D. B., Smith, C. M., Hanotte, O., Burke, T. & Cooke, F. Genetic polymorphism for alternative mating behaviour in lekking male ruff *Philomachus pugnax*. *Nature* **378**, 59-62, doi:10.1038/378059a0 (1995).

14 Saino, N. *et al.* Polymorphism at the Clock gene predicts phenology of long-distance migration in birds. *Molecular Ecology* **24**, 1758-1773 (2015).

15 Recuerda, M. *et al.* Adaptive phenotypic and genomic divergence in the common chaffinch (*Fringilla coelebs*) following niche expansion within a small oceanic island. *Journal of Evolutionary Biology* **36**, 1226-1241, doi:10.1111/jeb.14200 (2023).

16 Clark, J. D., Benham, P. M., Maldonado, J. E., Luther, D. A. & Lim, H. C. Maintenance of local adaptation despite gene flow in a coastal songbird. *Evolution* **76**, 1481-1494, doi:10.1111/evo.14538 (2022).

17 Rothstein, S. I. A Model System for Coevolution: Avian Brood Parasitism. *Annual Review of Ecology and Systematics* **21**, 481-508 (1990).

18 Thorogood, R., Spottiswoode, C. N., Portugal, S. J. & Gloag, R. The coevolutionary biology of brood parasitism: a call for integration. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, 20180190, doi:10.1098/rstb.2018.0190 (2019).

19 Langmore, N. E. *et al.* Coevolution with hosts underpins speciation in brood-parasitic cuckoos. *Science* **384**, 1030-1036, doi:10.1126/science.adj3210 (2024).

20 Rothstein, S. I. A Model System for Coevolution: Avian Brood Parasitism. *Annual Review of Ecology and Systematics* **21**, 481-508, doi:10.1146/annurev.es.21.110190.002405 (1990).

21 Tolman, D., Campobello, D., Rönkä, K., Kluen, E. & Thorogood, R. Reed Warbler Hosts Do Not Fine-Tune Mobbing Defenses During the Breeding Season, Even When Cuckoos Are Rare. *Frontiers in Ecology and Evolution* **9**, doi:10.3389/fevo.2021.725467 (2021).

22 Soler, M. *Avian Brood Parasitism: Behaviour, Ecology, Evolution and Coevolution*. (Springer International Publishing, 2018).

23 Stokke, B. G., Fossøy, F., Røskaft, E. & Moksnes, A. in *Avian Brood Parasitism: Behaviour, Ecology, Evolution and Coevolution* (ed. Manuel Soler) 363-384 (Springer International Publishing, 2017).

24 Davies, N. B. & Quinn, D. *Cuckoos, Cowbirds and Other Cheats*. (Bloomsbury Publishing Plc, 2011).

25 Gibbs, H. L. *et al.* Genetic evidence for female host-specific races of the common cuckoo. *Nature* **407**, 183-186, doi:10.1038/35025058 (2000).

26 Fossøy, F. *et al.* Ancient origin and maternal inheritance of blue cuckoo eggs. *Nature Communications* **7**, 10272, doi:10.1038/ncomms10272 (2016).

27 Merondun, J. *et al.* Evolution and genetic architecture of sex-limited polymorphism in cuckoos. *Science Advances* **10**, eadl5255, doi:10.1126/sciadv.adl5255.

28 Stoddard, M. C. & Kilner, R. M. The past, present and future of 'cuckoos versus reed warblers'. *Animal Behaviour* **85**, 693-699 (2013).

29 Bay, R. A. *et al.* Predicting Responses to Contemporary Environmental Change Using Evolutionary Response Architectures. *The American Naturalist* **189**, 463-473, doi:10.1086/691233 (2017).

30 MartÍN-GÁLvez, D. *et al.* A quantitative trait locus for recognition of foreign eggs in the host of a brood parasite. *Journal of Evolutionary Biology* **19**, 543-550 (2006).

31 Martín-Gálvez, D. *et al.* Cuckoo parasitism and productivity in different magpie subpopulations predict frequencies of the 457bp allele: a mosaic of coevolution at a small geographic scale. *Evolution* **61**, 2340-2348, doi:10.1111/j.1558-5646.2007.00194.x (2007).

32 Ruiz-Raya, F. Ecophysiology of egg rejection in hosts of avian brood parasites: new insights and perspectives. *Current Zoology* **67**, 631-638, doi:10.1093/cz/zoab042 (2021).

33 Soler, J. J., Martinez, J. G., Soler, M. & Møller, A. P. Genetic and geographic variation in rejection behavior of cuckoo eggs by European Magpie populations: an experimental test of rejector-gene flow. *Evolution* **53**, 947-956, doi:10.1111/j.1558-5646.1999.tb05388.x (1999).

34 Procházka, P. *et al.* Host genotype and age have no effect on rejection of parasitic eggs. *Naturwissenschaften* **101**, 417-426, doi:10.1007/s00114-014-1171-5 (2014).

35 Zueva, K. J. *et al.* Footprints of directional selection in wild Atlantic salmon populations: evidence for parasite-driven evolution? *PLoS One* **9**, e91672, doi:10.1371/journal.pone.0091672 (2014).

36 Wenzel, M. A., Douglas, A., James, M. C., Redpath, S. M. & Piertney, S. B. The role of parasite-driven selection in shaping landscape genomic structure in red grouse (*Lagopus lagopus scotica*). *Molecular Ecology* **25**, 324-341 (2016).

37 Moksnes, A., Fossøy, F., Røskaft, E. & Stokke, B. Reviewing 30 years of studies on the Common Cuckoo - accumulated knowledge and future perspectives. *Chinese Birds* **4**, 3-14, doi:10.5122/cbirds.2013.0001 (2013).

38 Pearson, D. J., Small, B. J. & Kennerley, P. R. Eurasian Reed Warbler: The characters and variation associated with the Asian form fuscus. **95**, 42-61 (2002).

39 Moksnes, A. *et al.* Behavioural Responses of Potential Hosts Towards Artificial Cuckoo Eggs and Dummies. *Behaviour* **116**, 64-89 (1991).

40 Wang, L., Yang, C., He, G., Liang, W. & Møller, A. P. Cuckoos use host egg number to choose host nests for parasitism. *Proceedings of the Royal Society B: Biological Sciences* **287**, 20200343, doi:doi:10.1098/rspb.2020.0343 (2020).

41 Avilés, J. M., Stokke, B. G., Moksnes, A., Røskaft, E. & Møller, A. P. Environmental conditions influence egg color of reed warblers *Acrocephalus scirpaceus* and their

parasite, the common cuckoo *Cuculus canorus*. *Behavioral Ecology and Sociobiology* **61**, 475-485, doi:10.1007/s00265-006-0275-0 (2007).

42 Stoddard, M. C. & Stevens, M. Avian vision and the evolution of egg color mimicry in the common cuckoo. *Evolution* **65**, 2004-2013 (2011).

43 Jelínek, V., Karasová, T., Weidinger, K., Procházka, P. & Honza, M. Do common cuckoo chicks suffer nest predation more than host nestlings? *Behavioral Ecology and Sociobiology* **70**, 1975-1987, doi:10.1007/s00265-016-2203-2 (2016).

44 Kameníková, M., Navrátil, J. & Rajchard, J. Dependence of clutch predation rate of Eurasian reed warbler *Acrocephalus scirpaceus* on nesting site selection: a model study. *Biologia* **71**, 452-456, doi:10.1515/biolog-2016-0053 (2016).

45 Buchan, C. *et al.* Combining remote sensing and tracking data to quantify species' cumulative exposure to anthropogenic change. *Global Change Biology* **29**, 6679-6692 (2023).

46 Procházka, P. *et al.* Delineating large-scale migratory connectivity of reed warblers using integrated multistate models. *Diversity and Distributions* **23**, 27-40 (2017).

47 Olsson, U. *et al.* Mitochondrial phylogeny of the Eurasian/African reed warbler complex (*Acrocephalus*, Aves). Disagreement between morphological and molecular evidence and cryptic divergence: A case for resurrecting *Calamoherpe ambigua* Brehm 1857. *Molecular Phylogenetics and Evolution* **102**, 30-44 (2016).

48 Wei, C. *et al.* Geographic variation in the calls of the Common Cuckoo (*Cuculus canorus*): isolation by distance and divergence among subspecies. *Journal of Ornithology* **156**, 533 - 542 (2015).

49 Ientile, R. *et al.* Analysis of Prey Composition in Eurasian Reed Warblers *Acrocephalus scirpaceus* Droppings at Four Breeding Sites in Italy. *Diversity* **14**, 1134 (2022).

50 Mills, L. J. *et al.* Using molecular and crowd-sourcing methods to assess breeding ground diet of a migratory brood parasite of conservation concern. *Journal of Avian Biology* **51** (2020).

51 Lindholm, A. K. Brood parasitism by the cuckoo on patchy reed warbler populations in Britain. *Journal of Animal Ecology* **68**, 293-309 (1999).

52 Fransson, T. & Stolt, B.-O. Migration routes of North European Reed Warblers *Acrocephalus scirpaceus*. *Ornis Svecica* **15**, 153-160, doi:10.34080/os.v15.22735 (2005).

53 Stokke, B. G., Takasu, F., Moksnes, A. & Røskaft, E. The Importance of Clutch Characteristics and Learning for Antiparasite Adaptations in Hosts of Avian Brood Parasies. *Evolution* **61**, 2212-2228 (2007).

54 Thorogood, R. & Davies, N. B. Reed warbler hosts fine-tune their defenses to track three decades of cuckoo decline. *Evolution* **67**, 3545-3555, doi:10.1111/evo.12213 (2013).

55 Lescak, E. A. *et al.* Evolution of stickleback in 50 years on earthquake-uplifted islands. *Proceedings of the National Academy of Sciences* **112**, E7204-E7212, doi:10.1073/pnas.1512020112 (2015).

56 Lo Cascio Sætre, C. *et al.* Rapid adaptive phenotypic change following colonization of a newly restored habitat. *Nature Communications* **8**, 14159, doi:10.1038/ncomms14159 (2017).

57 Sætre, C. L. C. *et al.* Genomic consequences of range expansion and colonisation in the reed warbler (*Acrocephalus scirpaceus*). *bioRxiv*, 2022.2011.2028.518135, doi:10.1101/2022.11.28.518135 (2022).

58 Bergman, N. *et al.* Conserved genetic diversity and no persisting founder effects following rapid range expansion of a long-distance migrant (*Acrocephalus scirpaceus*). *bioRxiv* (2025)

59 Procházka, P. *et al.* Low genetic differentiation among reed warbler *Acrocephalus scirpaceus* populations across Europe. *Journal of Avian Biology* **42**, 103-113 (2011).

60 Travis, J. M. J. *et al.* Deleterious Mutations Can Surf to High Densities on the Wave Front of an Expanding Population. *Molecular Biology and Evolution* **24**, 2334-2343, doi:10.1093/molbev/msm167 (2007).

61 Rougemont, Q., Leroy, T., Rondeau, E. B., Koop, B. & Bernatchez, L. Allele surfing causes maladaptation in a Pacific salmon of conservation concern. *PLOS Genetics* **19**, e1010918, doi:10.1371/journal.pgen.1010918 (2023).

62 Lancaster, L. T. *et al.* Understanding climate change response in the age of genomics. *Journal of Animal Ecology* **91**, 1056-1063 (2022).

63 Stokke, B. G. *et al.* Host density predicts presence of cuckoo parasitism in reed warblers. *Oikos* **116**, 913-922 (2007).

64 Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* **42**, 565-569, doi:10.1038/ng.608 (2010).

65 Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158, doi:10.1093/bioinformatics/btr330 (2011).

66 Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. Stacks: an analysis tool set for population genomics. *Molecular ecology* **22**, 3124-3140, doi:10.1111/mec.12354 (2013).

67 Sætre, C. L. C. *et al.* A Chromosome-Level Genome Assembly of the Reed Warbler (*Acrocephalus scirpaceus*). *Genome Biology and Evolution* **13**, evab212, doi:10.1093/gbe/evab212 (2021).

68 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

69 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

70 Wigginton, J. E., Cutler, D. J. & Abecasis, G. R. A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum Genet* **76**, 887-893, doi:10.1086/429864 (2005).

71 Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* **81**, 559-575, doi:10.1086/519795 (2007).

72 Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655-1664, doi:10.1101/gr.094052.109 (2009).

73 Petkova, D., Novembre, J. & Stephens, M. Visualizing spatial population structure with estimated effective migration surfaces. *Nat Genet* **48**, 94-100, doi:10.1038/ng.3464 (2016).

74 Caye, K., Jumentier, B., Lepeule, J. & François, O. LFMM 2: Fast and Accurate Inference of Gene-Environment Associations in Genome-Wide Studies. *Mol Biol Evol* **36**, 852-860, doi:10.1093/molbev/msz008 (2019).

75 Capblancq, T. & Forester, B. R. Redundancy analysis: A Swiss Army Knife for landscape genomics. *Methods in Ecology and Evolution* **12**, 2298-2309 (2021).

76 de Villemereuil, P. & Gaggiotti, O. E. A new FST-based method to uncover local adaptation using environmental variables. *Methods in Ecology and Evolution* **6**, 1248-1258 (2015).

77 Frichot, E. & François, O. LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution* **6**, 925-929 (2015).

78 Babin, C., Gagnaire, P-A., Pavey, S. A. & Bernatchez, L. RAD-Seq Reveals Patterns of Additive Polygenic Variation Caused by Spatially-Varying Selection in the American Eel (*Anguilla rostrata*). *Genome Biology and Evolution* **9**, 2974-2986, doi:10.1093/gbe/evx226 (2017).

79 Xuereb, A., Kimber, C. M., Curtis, J. M. R., Bernatchez, L. & Fortin, M-J. Putatively adaptive genetic variation in the giant California sea cucumber (*Parastichopus californicus*) as revealed by environmental association analysis of restriction-site associated DNA sequencing data. *Molecular Ecology* **27**, 5035-5048 (2018).

80 Stokke, B. G. *et al.* Predictors of resistance to brood parasitism within and among reed warbler populations. *Behavioral Ecology* **19**, 612-620, doi:10.1093/beheco/arn007 (2008).

81 Busby, J. R. BIOCLIM - a bioclimate analysis and prediction system. (1991).

82 Higgisson, W., Cobb, A., Tschierschke, A. & Dyer, F. The Role of Environmental Water and Reedbed Condition on the Response of *Phragmites australis* Reedbeds to Flooding. *Remote Sensing* **14**, 1868 (2022).

83 Caye, K., Jumentier, B., Lepeule, J. & François, O. LFMM 2: Fast and Accurate Inference of Gene-Environment Associations in Genome-Wide Studies. *Molecular Biology and Evolution* **36**, 852-860, doi:10.1093/molbev/msz008 (2019).

84 Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *European Journal of Human Genetics* **19**, 807-812 (2011)

85 Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004, doi:10.1111/j.0006-341x.1999.00997.x (1999).

86 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300 (1995).

87 Forester, B. R., Lasky, J. R., Wagner, H. H. & Urban, D. L. Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Molecular Ecology* **27**, 2215-2233 (2018).

88 Knaus, B. J. & Grünwald, N. J. vcfr: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources* **17**, 44-53 (2017).

89 Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403-1405, doi:10.1093/bioinformatics/btn129 (2008).

90 Goudet, J. hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes* **5**, 184-186 (2005).

91 Storey, J. D. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 479-498 (2002).

92 Käll, L., Storey, J. D., MacCoss, M. J. & Noble, W. S. Posterior Error Probabilities and False Discovery Rates: Two Sides of the Same Coin. *Journal of Proteome Research* **7**, 40-44, doi:10.1021/pr700739d (2008).

93 Roselaar, C. Songbirds of Turkey. An atlas of biodiversity of Turkish passerine birds. A & C Black (1995).

94 Procházka, P. *et al.* Low genetic differentiation among Reed Warbler *Acrocephalus scirpaceus* populations across Europe. *Journal of Avian Biology* **42**, 103-113, doi:10.2307/23018103 (2011).

95 Procházka, P., Van Wilgenburg, S. L., Neto, J. M., Yosef, R. & Hobson, K. A. Using stable hydrogen isotopes (δ2H) and ring recoveries to trace natal origins in a Eurasian passerine with a migratory divide. *Journal of Avian Biology* **44**, 541-550 (2013).

96 Arbabi, T., Gonzalez, J., Witt, H.-H., Klein, R. & Wink, M. Mitochondrial phylogeography of the Eurasian Reed Warbler *Acrocephalus scirpaceus* and the first genetic record of *A. s. fuscus* in Central Europe. *Ibis* **156**, 799-811 (2014).

97 Pârâu, L. G. & Wink, M. Common patterns in the molecular phylogeography of western palearctic birds: a comprehensive review. *Journal of Ornithology* **162**, 937-959, doi:10.1007/s10336-021-01893-x (2021).

98 Cahill, J. A. *et al.* Positive selection in noncoding genomic regions of vocal learning birds is associated with genes implicated in vocal learning and speech functions in humans. *Genome Res* **31**, 2035-2049, doi:10.1101/gr.275989.121 (2021).

99 Shiflett, M. W., Gavin, M. & Tran, T. S. Altered hippocampal-dependent memory and motor function in neuropilin 2-deficient mice. *Transl Psychiatry* **5**, e521, doi:10.1038/tp.2015.17 (2015).

100 Smith, J., Lipkin, E., Soller, M., Fulton, J. E. & Burt, D. W. Mapping QTL Associated with Resistance to Avian Oncogenic Marek's Disease Virus (MDV) Reveals Major Candidate Genes and Variants. *Genes (Basel)* **11**, doi:10.3390/genes11091019 (2020).

101 Charton, C. *et al.* The transcriptomic blueprint of molt in rooster using various tissues from Ginkkoridak (Korean long-tailed chicken). *BMC Genomics* **22**, 594, doi:10.1186/s12864-021-07903-9 (2021).

102 Røskaft, E., Moksnes, A., Stokke, B. G., Moskát, C. & Honza, M. The spatial habitat structure of host populations explains the pattern of rejection behavior in hosts and parasitic adaptations in cuckoos. *Behavioral Ecology* **13**, 163-168 (2002).

103 Røskaft, E., Takasu, F., Moksnes, A. & Stokke, B. G. Importance of spatial habitat structure on establishment of host defenses against brood parasitism. *Behavioral Ecology* **17**, 700-708, doi:10.1093/beheco/ark019 (2006).

104 Nogueira, A., Rey, P. J., Alcántara, J. M., Feitosa, R. M. & Lohmann, L. G. Geographic mosaic of plant evolution: extrafloral nectary variation mediated by ant and herbivore assemblages. *PLoS One* **10**, e0123806, doi:10.1371/journal.pone.0123806 (2015).

105 Craig, T. P. & Itami, J. K. A geographic mosaic of coevolution between Eurosta solidaginis (Fitch) and its host plant tall goldenrod *Solidago altissima* (L.). *Evolution* **75**, 3056-3070 (2021).

106 Cubaynes, S. *et al.* Disease outbreaks select for mate choice and coat color in wolves. *Science* **378**, 300-303, doi:10.1126/science.abi8745 (2022).

107 Davies, N. B. & Brooke, M. d. L. Cuckoos versus reed warblers: Adaptations and counteradaptations. *Animal Behaviour* **36**, 262-284 (1988).

108 Lowry, D. B. *et al.* Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Mol Ecol Resour* **17**, 142-152, doi:10.1111/1755-0998.12635 (2017).

109 Palmero-Iniesta, M., Espelta, J. M., Gordillo, J. & Pino, J. Changes in forest landscape patterns resulting from recent afforestation in Europe (1990–2012): defragmentation of pre-existing forest versus new patch proliferation. *Annals of Forest Science* **77**, 43, doi:10.1007/s13595-020-00946-0 (2020).

110 Mitchell, F. J. G. How open were European primeval forests? Hypothesis testing using palaeoecological data. *Journal of Ecology* **93**, 168-177 (2005).

111 Hoy, S. R. *et al.* The far-reaching effects of genetic process in a keystone predator species, grey wolves. *Sci Adv* **9**, eadc8724, doi:10.1126/sciadv.adc8724 (2023).

112 Bassitta, M. *et al.* Genomic signatures of drift and selection driven by predation and human pressure in an insular lizard. *Sci Rep* **11**, 6136, doi:10.1038/s41598-021-85591-x (2021).

113 Tams, V., Nickel, J. H., Ehring, A. & Cordellier, M. Insights into the genetic basis of predator-induced response in *Daphnia galeata*. *Ecol Evol* **10**, 13095-13108, doi:10.1002/ece3.6899 (2020).

114 Stewart, B. *et al.* The genetic architecture underlying prey-dependent performance in a microbial predator. *Nature Communications* **13**, 319, doi:10.1038/s41467-021-27844-x (2022).

115 Abdellaoui, A. & Verweij, K. J. H. Dissecting polygenic signals from genome-wide association studies on human behaviour. *Nature Human Behaviour* **5**, 686-694, doi:10.1038/s41562-021-01110-y (2021).

116 Bendesky, A. *et al.* The genetic basis of parental care evolution in monogamous mice. *Nature* **544**, 434-439, doi:10.1038/nature22074 (2017).