

Title	Testing for biases in selection on avian reproductive traits and partitioning direct and indirect selection using quantitative genetic models
Authors	Reed, Thomas E.;Gienapp, Phillip;Visser, Marcel E.
Publication date	2016-08-24
Original Citation	Reed, T.E., Gienapp, P. and Visser, M.E. (2016) 'Testing for biases in selection on avian reproductive traits and partitioning direct and indirect selection using quantitative genetic models', Evolution. doi:10.1111/evo.13017
Type of publication	Article (peer-reviewed)
Link to publisher's version	10.1111/evo.13017
Rights	© 2016, the Authors. This is the peer reviewed version of the following article: Reed, T.E., Gienapp, P. and Visser, M.E. (2016) 'Testing for biases in selection on avian reproductive traits and partitioning direct and indirect selection using quantitative genetic models', Evolution. doi:10.1111/evo.13017, which has been published in final form at http://dx.doi.org/10.1111/ evo.13017. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.
Download date	2025-08-23 20:24:56
Item downloaded from	https://hdl.handle.net/10468/3080



University College Cork, Ireland Coláiste na hOllscoile Corcaigh



TESTING FOR BIASES IN SELECTION ON AVIAN REPRODUCTIVE TRAITS AND PARTITIONING DIRECT AND INDIRECT SELECTION USING QUANTITATIVE GENETIC MODELS

Journal:	Evolution
Manuscript ID	16-0279.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Reed, Thomas; University College Cork, School of Biological Earth and Environmental Sciences Gienapp, Phillip; Netherlands Institute of Ecology, Visser, Marcel; Netherlands Institute of Ecology,
Keywords:	phenology, Fitness, climate change, microevolution, Heritability, genetic correlation



1	TESTING FOR BIASES IN SELECTION ON AVIAN REPRODUCTIVE
2	TRAITS AND PARTITIONING DIRECT AND INDIRECT SELECTION
3	USING QUANTITATIVE GENETIC MODELS
4	Short title: Selection in a wild population is real, not apparent
5	
6	Authors: Thomas E. Reed ^{1†} , Phillip Gienapp ^{2†} , Marcel E. Visser ²
7	Affiliations:
8	¹ School of Biological, Earth & Environmental Sciences, University College Cork, Cork,
9	Ireland, <u>treed@ucc.ie</u>
10	² Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), P.O.
11	Box 50, 6700 AB Wageningen, The Netherlands. p.gienapp@nioo.knaw.nl,
12	m.visser@nioo.knaw.nl
13	Correspondence to: Thomas E. Reed, School of Biological Earth & Environmental Sciences,
14	Distillery Fields, North Mall, University College Cork, Cork, Ireland, treed@ucc.ie
15	†Joint first authorship
16	Keywords: phenology, climate change, microevolution, heritability, genetic correlation,
17	fitness
18	Type of article: Original article
19	Word count: 7450 words; References: 54; 4 Figures; 2 Tables; 1 Appendix, 1 supplementary
20	figure and 4 supplementary tables.
21	Data will be archived upon manuscript acceptance in a data repository.

22 Abstract:

23 Key life history traits such as breeding time and clutch size are frequently both heritable and under directional selection, yet many studies fail to document micro-evolutionary responses. 24 25 One general explanation is that selection estimates are biased by the omission of correlated 26 traits that have causal effects on fitness, but few valid tests of this exist. Here we show, using 27 a quantitative genetic framework and six decades of life-history data on two free-living 28 populations of great tits *Parus major*, that selection estimates for egg-laying date and clutch 29 size are relatively unbiased. Predicted responses to selection based on the Robertson-Price 30 Identity were similar to those based on the multivariate breeder's equation, indicating that unmeasured covarying traits were not missing from the analysis. Changing patterns of 31 32 phenotypic selection on these traits (for laying date, linked to climate change) therefore reflect changing selection on breeding values, and genetic constraints appear not to limit their 33 34 independent evolution. Quantitative genetic analysis of correlational data from pedigreed populations can be a valuable complement to experimental approaches to help identify 35 36 whether apparent associations between traits and fitness are biased by missing traits, and to 37 parse the roles of direct versus indirect selection across a range of environments.

- 38
- 39
- 40

41 Introduction

Determining the potential for microevolution is fundamental to assessing how populations
may adapt to climate change (Holt 1990; Visser 2008) and the likelihood of evolutionary
rescue in altered environments (Gomulkiewicz and Holt 1995; Carlson et al. 2014). Adaptive

45 evolution requires heritable variation and while studies of natural populations typically find 46 substantial genetic variation in traits under directional selection, observations of 47 'evolutionary stasis', i.e. a lack of selection response in heritable traits, are common (Merilä et al. 2001; Estes and Arnold 2007; Walsh and Blows 2009). One prominent hypothesis to 48 explain such stasis, or to explain discrepancies between observed and expected evolutionary 49 50 responses in general, is that selection estimates may be biased by missing traits or variables 51 that are correlated with both focal traits and fitness (Schluter et al. 1991; Rausher 1992; 52 Kruuk et al. 2001, 2002, 2003; Hadfield 2008; Stinchombe et al., 2002; 2014; Morrissey et al. 53 2010). This can be the case, for example, when the relationship between fitness and traits is 54 environmentally-inflated and hence we would expect weaker (or no) response to selection (Fisher 1958; Price et al. 1988). 55

A classic example of evolutionary stasis (potentially underpinned by environmental 56 57 correlations between trait and fitness) is seasonal timing of breeding in temperate birds (Price et al. 1988): early breeders generally have higher reproductive success than late breeders 58 59 (Verhulst and Nilsson 2008) and egg-laying dates are typically heritable (Charmantier and 60 Gienapp 2014), implying that earlier egg-laying should evolve. Using a quantitative genetic 61 model, Price et al. (1988) showed how a lack of microevolution of heritable breeding time 62 can be compatible with selection for earlier breeding, if both breeding time and fitness are 63 influenced by a purely environmental variable, nutritional status in their example. Birds in 64 good nutritional condition may both breed earlier and produce more surviving offspring, but 65 earlier egg-laying will not evolve if fitness differences are entirely driven by nutritional 66 status. If traits or environments that are correlated with both focal traits and fitness are 67 missing from selection analyses, then the regression coefficients of relative fitness on trait at 68 the genetic and environmental levels will not be the same and hence phenotypic estimates of selection will be biased (Rausher 1992; Hadfield 2008; Morrissey et al. 2010). 69

70	Several studies of plant (e.g. Stinchcombe et al. 2002; Morrissey et al. 2010) and animal (e.g.
71	Kruuk et al. 2001, 2002; Gienapp et al. 2006) populations have sought to test whether
72	environmentally-induced covariances between traits and fitness bias selection estimates.
73	'Environmentally-induced covariance' here refers to situations where the focal trait is
74	correlated with another variable (e.g. a largely non-heritable trait such as nutritional status)
75	that has a causal effect on fitness, and should not be confused with the process of ecological
76	selection itself, whereby the selective environment causes a covariance between trait and
77	fitness (MacColl 2011; Bouwhuis et al. 2015). In these studies of potential environmental
78	biases to selection, fitness was regressed on predicted breeding values (PBVs, estimates of
79	the net effects of an individual's genes on its phenotype relative to the population mean) for
80	the trait of interest. Such a two-step approach is no longer considered appropriate on
81	statistical grounds, however, as PBVs remain confounded with environmental effects on the
82	phenotype (Postma 2006) and hypothesis tests based on PBVs can be highly anti-
83	conservative (Hadfield et al. 2010). Hence, we still have limited evidence whether selection
84	estimates in general in nature are biased by environmental covariances between trait and
85	fitness (or by unmeasured genetically correlated traits), particularly in free-living animal
86	populations (work on plants indicates such biases may be substantial, Scheiner et al. 2000;
87	Stinchcombe et al. 2002). This lack of evidence is particularly apparent for the case of
88	changing phenotypic selection: only one study of mammals (Robinson et al. 2008), to the best
89	of our knowledge, has tested whether changes in phenotypic selection are reflected by
90	changes in selection on underlying breeding values. This is particularly relevant in the
91	context of broad-scale environmental changes such as those wrought by global warming:
92	changing environmental covariances between traits and fitness could give the impression that
93	natural selection is intensifying, when in fact the genetic relationship between traits and
94	fitness may remain unchanged, leading to erroneous predictions of evolutionary responses.

95	Here we test the extent to which phenotypic selection estimates for two key avian life-history
96	traits (egg-laying date, LD and clutch size, CS) may be affected by such biases, using six
97	decades of data from two Dutch study populations of great tits (Parus major Linnaeus, 1758).
98	Our approach is based on the logic of the secondary theorem of natural selection (STS, also
99	known as the Robertson-Price Identity), which states that the expected per-generation
100	evolutionary response (or 'genetic selection differential') equals the covariance between
101	relative fitness and the breeding value for a trait, which under a simple quantitative genetic
102	model corresponds to the additive genetic covariance between relative fitness and trait
103	(Robertson 1966; Price 1970; Crow and Nagylaki 1976). The multivariate breeder's equation
104	(MVBE) can also be used to predict joint responses to selection on two or more correlated
105	traits (Lande and Arnold 1983) and has the advantage over the STS approach that direct and
106	indirect components of selection (and selection responses) can be distinguished using
107	selection gradients (Stinchcombe et al. 2014). However, selection gradients only partition
108	direct and indirect selection accurately when all correlated traits affecting fitness are included
109	in the analysis (Lande and Arnold 1983; Stinchcombe et al. 2014). Thus the MVBE can give
110	inaccurate predictions of microevolution when correlated traits are not measured and
111	Morrissey et al. (2010) have advocated using the STS to avoid this problem (see also
112	Morrissey et al. 2012). More recently, Stinchcombe et al. (2014) have championed a
113	combined approach that blends the merits of the STS and MVBE and allows evolutionary
114	responses to be estimated directly without bias, as well as direct and indirect components of
115	selection and selection responses to be partitioned. Implementation in a Bayesian-MCMC
116	framework also allows for statistically robust estimates of uncertainty on all parameters to be
117	made in a single model (Stinchcombe et al. 2014).
118	Following the approach recommended by Stinchcombe et al. (2014), we implement trivariate

119 Bayesian-MCMC animal models involving three traits: *LD*, *CS* and annual reproductive

Page 6 of 42

120	success (ARS, the number of recruiting offspring produced by an individual each breeding
121	season, a proxy for reproductive fitness). Posterior distributions of the (co)variance
122	components were then used to derive estimates of genetic selection gradients (β_G), i.e.
123	regression coefficients of breeding values for ARS on breeding values for each trait.
124	Similarly, we quantified the relationship between environmental effects on fitness and
125	environmental effects on traits, denoted β_E . The difference between β_G and β_E then provides a
126	measure of the extent of environmental bias to phenotypic selection (Rausher 1992; Hadfield
127	2008). We predicted that β_E should be negative for <i>LD</i> , as females experiencing favourable
128	environments (e.g. good nutrition) are likely to both initiate egg-laying earlier (i.e. more
129	negative LD) and raise more young, independent of their breeding values for LD. For CS we
130	predicted that β_E should be positive, given that females in good condition are likely to both
131	lay more eggs and recruit more offspring, regardless of their breeding values for CS. While
132	experimental manipulations of phenotypes provide the most robust tests for causal effects on
133	fitness, such experiments can be logistically challenging in the wild and are typically
134	attempted in only a limited number of years or environments (see Discussion for avian
135	examples involving laying date and clutch size and associated problems).
136	The trivariate animal models also allow us to assess the relative contributions of direct versus
137	indirect genotypic selection on each trait – the latter mediated via a potential genetic
138	correlation between LD and CS. Previous studies have provided mixed evidence for such a
139	correlation; for example, Sheldon et al. (2003) reported a negative genetic correlation for a
140	Swedish population of collared flycatchers (Ficedula albicollis), as did Garant et al. (2008)
141	for a UK population of great tits. Husby et al. (2010) reported a negative genetic correlation
142	in one great tit population, but a positive (albeit non-significant) genetic correlation in
143	another. To test whether changing patterns of phenotypic selection (in the case of LD, related
144	to climate change and phenological mismatch; Visser et al. 1998, 2006; Reed et al. 2013)

145 provide a reliable guide to changing selection on underlying breeding values, we split years

146 into groups based on variation in phenotypic selection and compared β_G against β_E in each

147 case. Finally, net responses to selection on each trait for both the full and sub-sampled

148 datasets were estimated using both the STS and MVBE approaches. By comparing them, one

149 can assess the extent to which missing correlated traits may bias predictions of

- 150 microevolution (Morrissey et al. 2012).
- 151

152 Materials and Methods

153 Data

154 The great tit populations in the HV (52°23'N, 05°51'E, central Netherlands) and on Vlieland 155 (53°10'N, 05°02'E, one of the West Frisian Islands in the Wadden Sea) have been 156 continuously monitored since 1955. Here we consider brood years from 1955 to 2013 157 inclusive, with recruit data from 2014 being used to estimate selection on traits expressed in 158 2013 (thus 60 years of data were used in total). Nest boxes are supplied in excess in all 159 suitable habitats in both study areas. The laying date of the first egg of a clutch (LD) was 160 calculated from the number of eggs found during weekly nest box checks, assuming that one 161 egg is laid per day. Clutch size (CS) was defined as the maximum number of eggs found 162 before or during incubation. Adults were caught during chick feeding and identified by their 163 aluminium and colour rings, or ringed if not previously caught. All nestlings were ringed with 164 aluminium rings before fledging. Annual reproductive success (ARS) was defined as the 165 number of recruits, i.e. the number of offspring that returned as adults to breed in our study 166 population, produced by that female in a given breeding season (including recruits from 167 potential second clutches, as decisions regarding the timing or size of first clutches will affect 168 the total number of recruits, not just those from first broods).

169 During the study period a number of broods was manipulated, e.g. by supplying food or 170 manipulating clutch or brood size. Since these manipulations could affect offspring survival, 171 manipulated broods were excluded from all analyses. From 1996 to 2003 a clutch size 172 selection experiment was carried out in the Vlieland study population (Postma et al. 2007). 173 During this experiment a large proportion of clutches was removed or swapped but because 174 all these clutches were excluded from our analyses, this experiment would not affect our 175 analyses here (in total, eight full years of data were excluded for VE due to manipulations: 176 1955-57, 1961-62, 1967-68, 2012). We restricted our analyses to the Eastern subpopulation 177 on Vlieland as the pedigree for the Western subpopulation is considerably shallower due to 178 higher immigration from the mainland (Postma & van Noordwijk 2005). Full details on sample sizes are provided in Table 1. 179

180

181 Statistical models

182 Our focal traits, LD and CS, are determined by the female and unaffected by properties of the 183 male in great tits (Caro et al. 2009). We consequently modelled these traits to be sex-limited, 184 i.e. not expressed by males, but not genetically sex-linked, which means that males were not 185 assigned any phenotypes but paternal links were included in the pedigree. We analysed 186 genetic (co)variances of LD, CS and ARS using the so-called 'animal model' (Henderson 187 1950; Kruuk 2004; Wilson et al. 2010) implemented in a Bayesian framework. Animal 188 models allow genetic and environmental sources of trait (co)variation to be disentangled, and 189 as such are well suited for quantitative genetic analyses in pedigreed natural populations as 190 they use all information about relatedness among individuals, and can handle unbalanced 191 datasets. Key advantages of the Bayesian approach, which utilises Markov Chain Monte 192 Carlo (MCMC) techniques (Hadfield 2010), are that (1) all sources of variability and 193 uncertainty are accounted for in the estimation procedure, which produces full posterior

195 parameters of interest and (2) non-Gaussian trait distributions can be modelled more easily 196 and reliably than in frequentist approaches (Morrissey et al. 2014). 197 Since many females bred in multiple years, we included a permanent environment random 198 effect in all models and also a maternal effect. All three traits vary considerably among years 199 due to phenotypically plastic responses, e.g. to temperature (LD), population density (CS) or 200 winter conditions (ARS). To account for these plastic year-to-year variations, LD and CS were 201 mean- and variance-standardised within years and we included a fixed effect of year for ARS 202 in all models (standardising ARS within years was avoided as it was more appropriate to treat 203 this as a Poisson variable in the models, which requires integer values). First-time breeders 204 generally have a later LD, lay smaller clutches and have reduced reproductive success and we 205 hence included age (factor with two levels: 'second calendar-year' (=first time breeder) and 206 'older') as a fixed effect in all models. 207 The R-package MCMCglmm (Hadfield 2010) was used to run all animal models. 208 Uninformative, proper priors were used with an 3×3 identity matrix for V and nu = 1.002. 209 The results were robust to alternative prior specifications (e.g. stronger priors, results not

probability distributions, rather than point estimates and approximate standard errors, of

shown). We used a burn-in period of 250,000 for all models and a thinning interval of 10,000

to ensure proper mixing of the chain and independent samples (the autocorrelation between

- samples was always <0.2). The number of effective samples was never substantially smaller
- than the number of samples drawn (200).

214

215 Decomposing selection into genetic versus environmental components.

216 Selection is technically measured as the relationship between trait and *relative* fitness

217 (individual fitness divided by mean individual fitness), which can be expressed as a

regression slope, as in selection gradients (Lande and Arnold 1983), or as a covariance, as in
selection differentials (Price 1970; Endler 1986). However, relative fitness does not conform
to any known parametric distribution and hence we instead modelled the (genetic and
environmental) relationships between trait and absolute *ARS* using a log-link generalised
linear model (Poisson distributed errors). The regression coefficients from this type of model
are then equivalent to the Lande-Arnold regression coefficients using relative fitness (Smouse
et al., 1999).

With two traits of interest (*Z1* and *Z2*) and a single fitness measure (*W*), one can fit a trivariate animal model that produces as output a 3×3 genetic covariance matrix, which

following Stinchcombe et al. (2014) we call G_{zw} :

$$\boldsymbol{G}_{zw} = \begin{bmatrix} Vg_{z1} & cov_{z1,z2} & cov_{z1,w} \\ Vg_{z2} & cov_{z2,w} \\ Vg_{W} \end{bmatrix}$$

228 Note that while use a g subscript here and throughout the paper when referring to genetic 229 parameters, these actually refer to the variance or covariance of additive genetic effects (i.e. 230 breeding values). Vg_W corresponds to the genetic variance in the fitness component (if total 231 relative fitness were used, this parameter would specify the upper limit on the rate of 232 evolution, according to Fisher's fundamental theorem, Fisher 1930). The off-diagonal matrix 233 elements of the column/row corresponding to fitness indicate genetic covariances between 234 traits and fitness; arranged as a vector these give s_{G} , the genetic selection differentials 235 (Stinchcombe et al. 2014). These correspond to the predicted evolutionary responses for each 236 trait, according to the STS (with the caveat that here we only consider a component of fitness, 237 ARS, as opposed to total fitness). Matrix elements not involving W represent the standard 238 genetic covariance matrix **G** for the traits (in this case a 2×2 matrix). Structurally identical 239 3×3 covariance matrices are produced for all random effects included in the animal models.

240	The vector of genetic selection gradients β_G can then be derived using $\beta_G = G^{-1}s_G$ (Lande and
241	Arnold 1983; Rausher 1992; Stinchcombe et al. 2014). Likewise, we estimated the overall
242	relationships between each trait and <i>ARS</i> at the environmental level as $\beta_E = E^{-1}s_E$, where β_E
243	was a vector of environmental "selection" gradients, E was an environmental covariance
244	matrix calculated by summing the posterior distributions of the covariance matrices for the
245	permanent environment effects (repeatable differences among individuals across years not
246	due to additive genetic effects), maternal effects, and residual deviations (within year
247	environmental effects on phenotype). s_E refers to the vector of environmental selection
248	differentials, calculated by summing the permanent environment, maternal and residual
249	covariances between trait and fitness. The estimates of β_E were very similar when maternal
250	effects (which could themselves contain a maternal genetic component) were excluded from
251	the calculations. We also re-ran the trivariate animal models (all years considered together
252	only) using unstandardised trait values and including year effects in the calculation of β_E in
253	order to explore whether the main conclusions were affected by our procedure of
254	standardising traits within years (see Appendix 1).
255	The environmental bias to phenotypic selection on each trait was then quantified as $\beta_E - \beta_G$
256	(bold symbols are used to denote the 2×1 vector of biases, with the first element
257	corresponding to the bias for LD and the second the bias for CS; when referring to the bias for
258	each trait separately we simply use $\beta_E - \beta_G$; see Fig.1 for a graphical representation of
259	environmental biases to selection). Statistical support for an environmental bias to selection
260	on either trait was then assessed by simply checking whether the posterior distributions of
261	this metric overlapped zero. If the 95% HPD (highest posterior density) interval included
262	zero, then the null hypothesis of no environmental bias was accepted.
263	

264 Changes in selection through time: real or apparent?

Page 12 of 42

265 To test whether changes in the magnitude of phenotypic selection were underpinned by 266 similar changes in selection on breeding values, we split the Hoge Veluwe (HV) and Vlieland 267 East (VE) datasets into years with 'strong', 'medium' and 'weak' phenotypic selection on 268 LD, and also (separately) based on 'strong', 'medium' and 'weak' phenotypic selection on 269 CS. Annual standardised phenotypic selection differentials (denoted s_P) were calculated by 270 dividing individual fitness by annual mean fitness and regressing this relative individual 271 fitness against (mean and variance) standardised egg-laying date or clutch size (Lande and 272 Arnold 1983). We then split years into three groups based on thirds on the distribution of $s_{P_{2}}$ 273 for each trait (see Table 1 for full details). The 'weak' and 'strong' categories not only 274 differed in strength but also partly in the direction of selection. The 'weak' category 275 contained years with weakly positive (LD) or negative to no selection (CS), the 'medium' 276 category years with no or weakly negative (LD) or positive (CS) selection, while the 'strong' 277 category contained years with strongly negative (LD) or strongly positive (CS) selection 278 (Table 1).

Trivariate animal models were fitted for each group of years and β_G and β_E were calculated as 279 280 before. The statistical significance of directional selection on underlying genotypes was 281 determined qualitatively by assessing whether the HPD interval of β_G overlapped zero for 282 each trait/phenotypic selection strength combination. Similarly, support for environmental 283 biases to phenotypic selection in each category was determined by assessing whether the HPD interval of $\beta_E - \beta_G$ did not overlap zero (see Fig.1 for a hypothetical example). Due to 284 285 the large data sets necessary to reliably estimate genetic covariances it was simply impossible 286 to conduct this analysis at a finer temporal scale, let alone at an annual basis. 287

288 Assessing the power to detect environmental biases

289 Even if the HPD interval of our metric of bias $(\beta_E - \beta_G)$ includes zero, the possibility remains 290 that insufficient statistical power was available to detect true biases (e.g. relatively small 291 biases). To get a better sense for this we undertook a power analysis, whereby two traits were 292 simulated (assuming multivariate normality for simplicity) to be uncorrelated at the genetic 293 level, but correlated at the environmental level. One of the 'traits' was assumed to be fitness 294 and the other either LD or CS; thus the simulations modelled a real (and complete) 295 environmental bias to phenotypic selection. Uncorrelated breeding values for each trait were 296 simulated using the *rbv* function in MCMCglmm (Hadfield 2010) across both the HV and VE 297 pedigrees for the same number of individuals for which actual phenotypic information was 298 available. Correlated environmental deviations were then simulated from a multivariate 299 normal distribution and added to the uncorrelated breeding values to generate simulated 300 phenotypes. Permanent environment and maternal effects were ignored for simplicity. 301 Bivariate animal models were then run on these simulated phenotypes to generate estimates of β_G and β_E as above. Six different strengths of environmental bias (i.e. six different β_E 302 values, and therefore also $\beta_E - \beta_G$ values, given that β_G was simulated to be zero) ranging 303 304 from 0 to 0.50 were simulated. For each, 500 replicate simulations were run for both HV and 305 VE and power was calculated as the proportion of simulations where the HPD interval of the 306 resulting posterior estimates of $\beta_E - \beta_G \operatorname{did}$ not include 0.

307

308 Comparing evolutionary predictions of the STS and MVBE

Estimates of the response to selection on *LD* and *CS* were obtained from the trivariate animal models by extracting the posterior distributions of $\mathbf{s}_{\mathbf{G}}$ (i.e. the additive genetic covariances between each trait and *ARS*), which corresponded to the evolutionary predictions based on the STS approach. Estimates based on the MVBE approach were obtained using $\Delta \bar{z} = G\beta$, where $\Delta \bar{z}$ indicates the change in the mean of each trait, *G* is the genetic covariance matrix as

estimated from the trivariate model (the upper 2×2 quadrant of G_{zw} , see above) and β is the 314 315 vector of phenotypic selection gradients, as estimated from the posterior distributions of the trivariate animal model (using $\beta = P^{-1}s_P$, where P = G + E and $s_P = s_G + s_E$). The goal 316 317 of this exercise was to compare predictions from the STS and MVBE relative to each other, 318 rather than to generate quantitatively accurate predictions of selection responses per se – the 319 latter would not be completely reliable in any case, given that assumptions of both the STS 320 and MVBE such as constant demography and non-overlapping generations are not met. The 321 predicted responses to selection based on both approaches were in phenotypic standard 322 deviation (PSD) units for each trait, because standardised trait values were used in both cases.

323

324 **Results**

325 Phenotypic patterns

326 Estimates of directional selection at the phenotypic level varied substantially among years in 327 strength and sign for both LD and CS in each population (Supplementary Fig.1). For the HV 328 population, earlier layers had higher fitness on average across all years (mean s_P : = -0.14, 329 range among years = -1.06 to 0.98), with 45 out of 59 years (76%) exhibiting negative 330 selection differentials. Phenotypic selection for earlier laying was on average weaker across 331 all years for the VE population (mean s_P : = -0.014, range among years = -0.86 to 0.64), with 332 28 of 52 years (54%) exhibiting negative selection differentials. In the HV population, females laying larger clutches had higher fitness on average across all years (mean $s_P = 0.14$, 333 range among years = -0.68 to 0.69), with 40 of 59 years (68%) exhibiting positive selection 334 335 differentials, whereas in the VE population phenotypic selection on CS was on average weaker (mean $s_P = 0.02$, range among years = -0.72 to 0.48), with 32 of 44 years (73%) 336 337 exhibiting positive selection differentials.

LD and CS were negatively phenotypically correlated (HV population all years: standardised 338 339 trait values: $r_p = -0.23$; unstandardised trait values: $r_p = -0.21$; VE population all years: 340 standardised trait values: $r_p = -0.12$; unstandardised trait values: $r_p = -0.06$; all P<0.05). These 341 reflected within-year associations between LD and CS, as the annual means were not 342 significantly correlated for either population (HV population: r = -0.10, P = 0.45; VE population: r = 0.10, P = 0.46). For the HV population, there was a trend towards earlier egg-343 laying (across all years) of 0.1 days per year ($b = -0.10 \pm 0.04$, $t_{1.57} = -2.47$, P = 0.016) and 344 also a trend towards smaller first clutches ($b = -0.02 \pm 0.006$, $t_{1.57} = -2.44$, P = 0.018) 345 346 (Supplementary Fig.1). For the VE population, there were no significant temporal trends in either LD ($b = -0.08 \pm 0.04$, $t_{1,52} = -1.81$, P = 0.076) or CS ($b = -0.01 \pm 0.01$, $t_{1,52} = -1.21$, P = -1.21, P = -347 0.27) (Supplementary Fig.1). 348

349

350 Trivariate animal models: all years considered together

351 Additive genetic variance was found to be non-zero for all three traits in both populations 352 (see Supplementary Tables for full results of trivariate models). For the HV population, the heritability (h^2) of standardised LD was estimated at 0.16 (HPD interval: 0.09 - 0.20; the 353 point estimate here and for all subsequently reported parameters refers to the posterior mode, 354 and the range to the HPD interval), h^2 of standardised CS was estimated at 0.21 (0.14 – 0.30) 355 and the h^2 of (unstandardised) ARS was estimated at 0.29 (0.15 – 0.38). For the VE 356 population, the heritability (h^2) of standardised LD was estimated at 0.17 (HPD interval: 0.12) 357 - 0.29), h^2 of standardised CS was estimated at 0.18 (0.13 – 0.27) and the h^2 of 358 (unstandardised) ARS was estimated at 0.24 (0.16 - 0.36). LD and CS were standardised 359 within years, while a fixed effect of year was included for ARS, and thus the h^2 estimates here 360 correspond to the fraction of within-year variation (additive genetic + permanent environment 361 + maternal + residual) explained by additive genetic effects. For ARS, the h^2 estimate is at the 362

363	scale of the linear predictor. For purposes of comparison with other traits, we back-
364	transformed this estimate to the observed scale (h^2_{obs}) using the following equation (Foulley
365	1993): $h_{obs}^2 = \frac{\mu^2 \sigma_a^2}{\mu + \mu^2 [exp(\sigma_a^2) - 1]}$, where μ was the mean on the observed scale and σ_a^2 was the
366	additive genetic variance estimated by the model. This gave an estimate of h_{obs}^2 for ARS of
367	0.05 (0.03 - 0.07) for the HV population and $0.05 (0.03 - 0.07)$ for the VE population.
368	Considering all years together, β_G for <i>LD</i> was estimated as -0.08 (-0.31 – 0.26) for the HV
369	population, while β_E was estimated at -0.20 (-0.24 – -0.06; Table 2). The negative
370	relationship between LD and ARS at the environmental level was driven predominantly by a
371	statistically significant (HPD interval not overlapping zero) negative residual covariance
372	(Supplementary Table 4), as the permanent environment (Supplementary Table 2) and
373	maternal covariances (Supplementary Table 3) were overlapping zero. The bias statistic (β_E -
374	β_G) for <i>LD</i> was estimated as -0.03 (-0.46 – 0.21, Table 2, Fig. 2); note that the posterior mode
375	of the derived statistic $\beta_E - \beta_G$ can deviate from the difference in the posterior modes of β_E
376	and β_G due to posterior distributions not being perfectly symmetrical.
377	Considering all years together, β_G for <i>CS</i> was estimated as 0.06 (-0.15 – 0.33) for the HV
378	population, while β_E was estimated at 0.09 (0.04 – 0.21; Table 2). The positive relationship
379	between CS and ARS at the environmental level was driven predominantly by a statistically
380	significant positive residual covariance (Supplementary Tables). The bias statistic ($\beta_E - \beta_G$)
381	for <i>CS</i> was estimated as 0.01 (-0.24 – 0.32, Table 2, Fig. 2).
382	For the VE population, β_G for <i>LD</i> was estimated as -0.06 (CI: -0.26 – 0.12) considering all
383	years together, while β_E was estimated at 0.01 (-0.07 – 0.05; Table 2). Surprisingly, a positive
384	permanent environment covariance between LD and ARS was evident across all years for the
385	VE population (Supplementary Table 2), whereas a negative residual covariance was found
386	(Supplementary Table 4). These counteracting covariances explain why the overall β_E was

close to zero. The bias statistic ($\beta_E - \beta_G$) for *LD* was estimated as -0.01 (-0.20 – 0.24, Table 2, Fig. 2).

389	For the VE population, β_G for CS was estimated as -0.02 (-0.17 – 0.19) considering all years
390	together, while β_E was estimated at 0.08 (0.01 – 0.12; Table 2). The positive relationship
391	between CS and ARS at the environmental level was driven predominantly by a positive
392	residual covariance (Supplementary Tables). The bias statistic ($\beta_E - \beta_G$) for <i>CS</i> was estimated
393	as 0.11 (-0.14 – 0.27, Table 2, Fig. 2). The trivariate animals based on unstandardised trait
394	values produced very similar results to those based on standardised <i>LD</i> and <i>CS</i> (Appendix 1).
395	
396	Trivariate animal models: splitting years by selection strength categories
397	For both populations, changes in phenotypic selection strength for both traits were generally
398	paralleled by similar changes in selection on the additive genetic component of trait variation
399	(Fig. 2, Table 2). The 'strong' phenotypic selection category for CS in the HV population was
400	the only one where the HPD intervals for β_G were completely non-overlapping zero (Fig.2,
401	Table 2), indicating that selection on CS breeding values was consistently positive in these
402	years. In general, however, the model estimates for β_G became larger in absolute terms (more
403	positive for CS and more negative for LD) as phenotypic selection became stronger. Although
404	the posterior modes for β_G deviated somewhat from those for β_E (Fig. 1), the full posterior
405	distributions overlapped considerably and the HPD intervals for $\beta_E - \beta_G$ overlapped zero in all
406	cases (Table 2). Full details on the additive genetic, permanent environment, maternal and
407	residual covariance matrices for each population/trait/ selection strength category
408	combination are given in Supplementary Tables 1-4.

409

410 Power to detect environmental biases

411	The power analyses showed that there was >80% power to detect true environmental biases
412	to selection of approximately 0.40 ($\beta_E - \beta_G$) or higher for both populations, but only
413	approximately 25-50% power to detect environmental biases of 0.20 to 0.30 (Fig. 3). Power
414	declined approximately sigmoidally as simulated $\beta_E - \beta_G$ decreased. Power to detect biases
415	was slightly higher for the VE population, likely reflecting the better pedigree (more
416	relatedness links) compared to the HV pedigree.

418 Comparing evolutionary predictions of the STS and MVBE

419 For LD, the MVBE predicted a very small response to selection ($\Delta z = -0.02$ PSD or 0.09 420 days, per generation) overall across the whole time period in the HV population, whereas the 421 STS predicted a smaller response to selection (s_g =-0.003 PSD, HPD interval: -0.053 – 0.031). 422 Similarly, for the VE population, a very weak response to selection was predicted (error bars 423 overlapping zero for both methods) by both the MVBE and the STS (Fig.4). These responses 424 refer to the expected net rate of microevolution per generation, assuming constant directional 425 selection. For CS, the MVBE predicted a very small positive response to selection ($\Delta z =$ 426 0.025 PSD, or 0.05 eggs, per generation) overall across the whole time period in the HV 427 population. The modal estimate of the response to selection according to the STS was similar 428 $(s_g = 0.035 \text{ PSD})$ but with a broader HPD interval that overlapped zero (-0.020 - 0.078 PSD). 429 For the VE population, a slightly positive response to selection was predicted by the MVBE 430 $(\Delta z = 0.011 \text{ PSD}, \text{ or } 0.020 \text{ eggs per generation})$ across all years, while the STS predicted a 431 slightly negative response (-0.003 PD) but with a HPD interval (-0.034 - 0.035 PSD) that overlapped zero (Fig.4). 432

Predicted responses to selection were on average larger for both methods in years where
phenotypic selection was stronger, and the MVBE and STS gave qualitatively and

435	quantitatively similar predictions when years were grouped according to phenotypic selection
436	strength (Fig.4). The uncertainty associated with the STS predictions was considerably larger
437	than that associated with the MVBE predictions (Fig.4). The general concordance between
438	the MVBE and STS predictions reflected the fact that the genetic covariance between LD and
439	CS overlapped zero in all trivariate animal models (Supplementary Table 1) and that no
440	strong environmental biases to selection were found (which could have biased the MVBE,
441	but not the STS, predictions). Thus indirect selection responses appeared not to play any role,
442	at least with respect to the two traits considered in the analysis, as there was no evidence for
443	statistically significant genetic covariance between them.

445 Discussion

446 Using six decades of individual-based life history data and advanced, powerful statistical 447 techniques we have shown that (1) heritable variation in a key component of fitness (the 448 annual number of recruits) exists and thus microevolution is possible in our study populations, (2) heritable variation exists for two key reproductive traits (LD and CS) known 449 450 to affect fitness, and (3) selection estimates are relatively unbiased by missing traits or 451 variables that may be correlated with these traits and fitness. This latter result is our most 452 important finding and can be interpreted as a "quantitative genetic signature" (c.f. Morrissey 453 and Ferguson 2011) of changing patterns of natural selection (see also Robinson et al. 2008). 454 Phenotypic selection estimates in our great tit study populations are therefore reliable and not 455 entirely driven by changes in environmental correlations between traits and fitness. This does 456 not imply that the latter do not exist (β_E for each trait was typically non-zero in the datasets 457 analysed here, Fig.2, Table 2), nor that environmental relationships between trait and fitness 458 are not also changing (β_E was different for different phenotypic selection strength categories

459	in line with our predictions, i.e. it was more negative in years where s_P for LD was more
460	negative, and more positive in years where s_P for CS was more positive, Fig.2, Table 2).
461	Rather, changes to β_E were paralleled by similar changes to β_G (Fig.2), which implies that our
462	phenotypic selection estimates were not unduly biased. Directional environmental change, for
463	example associated with regional warming (Gienapp et al. 2013), should therefore induce
464	evolutionarily-relevant selection. We note, however, that while equality of β_G and β_E for each
465	trait is consistent with these traits causally affecting fitness, it is not sufficient: proportionality
466	of the phenotypic and genetic covariance matrices for the focal and selected traits also gives
467	rise to equality of β_G and β_E even when the regression coefficients do not represent the causal
468	effect of the focal trait on fitness (see Section 2 in Appendix A of Hadfield 2008). Should
469	covariance in year-effects on each trait should be included in the calculation of β_E ? The
470	answer is not immediately obvious and depends on the extent to which (interannual)
471	genotype-by-environment interactions contribute to overall trait variation and whether one
472	conceives of selection as operating within years, or also across years. In our case, including
473	year effects in the calculation of β_E tended to make the latter deviate slightly more from β_G
474	(i.e. more bias) compared to when year-effects were excluded, but the differences were
475	relatively minor, being somewhat more pronounced for CS because the year covariance was
476	positive for that trait (Appendix 1).
477	If we had found a significant deviation of β_E from β_G in our datasets (i.e. if the posterior
478	distributions of $\beta_E - \beta_G$ had not overlapped zero), this would have indicated that the null
479	hypothesis of no bias to selection should have been rejected, which was not the case for any

- 480 of the datasets we analysed. However, absence of evidence is not necessarily evidence of
- 481 absence: a lack of significant bias could simply be explained by a lack of statistical power to
- detect true bias. Our power analyses indicated that we only had sufficient power to detect
- 483 large biases (Fig.3), although what constitutes 'large bias' is somewhat subjective and

484	difficult to define. According to our power analysis, we had >80% power to detect biases in
485	excess of approximately 0.4, but only 25-50% power to detect 'moderate' biases in the region
486	of 0.2 to 0.3 (which encompassed many of the actual estimates of $\beta_E - \beta_{G_s}$ see Table 2) with
487	the units here corresponding to those for standardised selection gradients, i.e. proportional
488	change in relative fitness per phenotypic standard deviation. Stinchcombe et al. (2002)
489	provided an analysis of environmentally-induced biases in phenotypic selection estimates
490	based on field experiments with three species of annual plants and reported standardised
491	selection gradients at both the phenotypic (β_P) and additive genetic (β_G) levels. The mean
492	absolute bias based on their data (calculated as $ \beta_P - \beta_G $, extracting the β_P and β_G values from
493	their Tables 2, 3 and 4) was 0.28 (note that with no bias, $\beta_P = \beta_G = \beta_E$) and ranged from 0.02 to
494	0.77. Using this as a yardstick suggests that we had sufficient power in the current study to
495	detect only relatively large biases, but Stinchcombe et al. (2002) noted that their estimated
496	biases were likely conservative in that they were based on data from spatially replicated field
497	experiments; i.e. most studies of selection in the wild are based on correlational data collected
498	under uncontrolled environmental conditions, where environmental biases may be
499	considerably larger. In the current study, the standard deviation in s_P (s_P is equivalent to
500	univariate β_P for our great tit populations was 0.34 for <i>LD</i> and 0.28 for <i>CS</i> (pooling annual
501	s_P estimates from both populations). Denoting this standard deviation as $\sigma(s_P)$, as a rule of
502	thumb one might consider biases between $\sigma(s_P)$ and $2\sigma(s_P)$ as 'moderate' and biases in excess
503	of $2\sigma(s_P)$ as 'large'. Thus while we lacked sufficient statistical power to detect 'small' biases
504	(e.g. $< \sigma(s_P)$), such minor biases would be less of a concern in the sense that inferences
505	regarding evolutionarily relevant selection would be unlikely be too 'far off the mark' if only
506	phenotypic-level information were available. Likewise, predictions of the response to
507	selection based on the MVBE should not be too inaccurate (predictions based on the STS

would not suffer from the same problem, as they are unbiased by potential environmentalcovariances or missing traits).

510 The STS and MVBE approaches yielded similar predicted responses to selection on each trait 511 in each population (Fig.4). While the STS has the advantages over the MVBE that responses 512 to selection can be estimated in a single model and are unbiased, one cannot disentangle 513 direct from indirect components of selection/selection responses (Stinchcombe et al. 2014). 514 The MVBE approach on the other hand suffers from the major disadvantage that one can only be sure that the predictions are accurate when all correlated traits under selection are 515 516 included in the analysis (Stinchcombe et al. 2014). The broad concordance we found between the STS and MVBE predictions implies that missing correlated traits were not a major issue 517 518 in our case. However, the uncertainty associated with both sets of predictions was substantial 519 and thus we cannot rule out the existence of missing correlated traits completely, we can just 520 infer that their potential absence did not unduly bias the MVBE estimates. The quantitative 521 predictions themselves (under both approaches) must be treated with caution to some extent, 522 however, because both the STS and MVBE make assumptions that are not entirely met by 523 our data, such as constant demography and non-overlapping generations. Our primary goal in 524 comparing the predictions of both approaches, however, was to assess the extent to which 525 missing correlated traits may have been an issue, rather than to generate quantitatively accurate predictions of selection responses per se. 526 527 By applying the analytical framework recommended by Stinchcombe et al. (2014), we were

able to estimate partial genetic selection gradients for each trait and therefore to separate the

effects of direct versus indirect selection. The results indicated that indirect components of

selection were relatively unimportant, given that the estimates for β_G were very similar to the

estimates for s_G . The phenotypic correlations between *LD* and *CS* were also relatively weak

in both populations and the genetic correlations were not significantly different from zero

533 (Supplementary Table 1), implying that selection on one trait would not cause a correlated 534 response in the other. A positive genetic correlation in this case would imply a genetically-535 based trade-off, in that the traits are typically selected in opposite directions. Studies of other 536 songbird populations have previously reported a negative genetic correlation between these traits (Sheldon et al. 2003, Garant et al. 2008) or no genetic correlation/a positive correlation 537 538 (Husby et al. 2010), suggesting that genetic trade-offs between these avian reproductive traits 539 are not inevitable and may even be population- or environment-specific. Estimates of genetic 540 covariances/correlations are typically associated with large uncertainties (Lynch and Walsh 541 1998) however, and comparisons of their strength across contexts must therefore be treated 542 with caution.

543 Patterns of phenotypic selection on LD and CS differed somewhat between the HV and VE 544 study populations, with s_P deviating more from zero in particular for LD in the HV 545 population (Table 1, Supplementary Fig.1). In the early part of the study (1950s to early 546 1980s) the breeding time of great tits in the HV study area was relatively well-matched, on 547 average, with the caterpillar food peak and hence no net directional selection for earlier egg-548 laying was expected or observed (Visser et al. 1998; Reed et al. 2013). An increasing 549 phenological mismatch between great tits and their food then developed from the 1980s 550 onwards as climate change unfolded (Visser et al. 1998, 2006, Chevin et al. 2015) and as a 551 result, phenotypic selection for earlier laying intensified (Reed et al. 2013, Supplementary 552 Fig.1). The strong selection category for LD therefore consisted of (largely, but not 553 exclusively, more recent) years where phenological mismatch was high and this explains why 554 β_G was more negative in these (Fig. 2) and why a stronger response to selection was predicted 555 (Fig. 4). The fact that the HPD intervals associated with β_G and s_G for LD overlap zero in all 556 selection strength categories indicates that years with varying selection pressures (not only in 557 terms of magnitude, but potentially also sign) are still pooled in these analyses, and also that

Page 24 of 42

558 genetic signatures of directional selection are more difficult to pick out from the 'noise' when 559 sample sizes are reduced like this. The importance of phenological matching with a shifting 560 food peak has been less well-studied in the VE population, but it is likely that timing relative 561 to seasonal peaks in caterpillar biomass plays a similar role in driving selection on LD in that area. Fluctuations in population density appear to drive selection on CS (Both et al. 2000; 562 563 Saether et al. In Press): under high population densities with increased competition for 564 resources or territories, individuals in good 'condition' would have a selective advantage, 565 which means that under high densities breeders should trade-off a larger clutch size for an 566 increased investment in offspring, leading to selection for smaller clutch size under high 567 densities. Climate change may also select indirectly on CS via a genetic correlation with LD, 568 but as we have shown, evidence for genetic linkages between these traits was lacking in this 569 study.

570 On average over the six decades considered, selection appeared to favour earlier egg-laving 571 and larger clutches in both populations and in the HV population mean LD advanced 572 significantly over time, yet mean CS also exhibited an overall negative temporal trend (Supplementary Fig.1). For the VE population, both mean LD ($b = -0.16 \pm 0.06$, P=0.01) and 573 574 mean CS ($b = -0.03 \pm 0.01$, P<0.001) exhibited significant negative temporal trends when the data were restricted to 1970 onwards (sample sizes were much smaller in the earlier years). 575 576 Both patterns are likely mostly explained by phenotypic plasticity rather than microevolution. 577 For LD it is well established that earlier egg-laying occurs as a plastic response to higher 578 spring temperatures, with springs getting progressively warmer in recent decades in the 579 Netherlands (Visser et al. 1998; 2006; Nussey et al. 2005, Husby et al. 2010). An increase in 580 population density may drive a decrease in mean CS as a plastic response, yet population size 581 has not exhibited a directional trend in the HV over time, although it has increased

583	responsible for the observed trends in CS in both populations.
584	
585	Testing for biases to selection using observational versus experimental approaches
586	Here we tested for potential biases to selection using very long-term datasets and an animal
587	model approach, which had the advantage of generality in the sense that the analyses
588	integrated across many different types of years and hence variable selective pressures,
589	whereas experimental approaches to the same question (e.g. Stinchombe et al. 2002) typically
590	can only be carried out in one or a few years. Nonetheless, we acknowledge that correlational
591	data have their limits and that experimental manipulations of putative targets of selections
592	(i.e. phenotypes of interest) provide the most robust tests of whether traits truly causally
593	affect fitness. Such experiments are logistically challenging, however.
594	Several studies with birds have manipulated LD and CS (or brood size), and found that these
595	manipulations affected reproductive success (e.g. Dijkstra et al. 1990; Daan et al. 1990;
596	Verhulst and Tinbergen 1991; Brinkhof et al. 1993; Svensson 1997; Pettifor et al. 1998;
597	Visser and Lessells 2001). Delaying breeding time by removing eggs, which were then
598	replaced by the breeding female, reduced the reproductive success of the manipulated broods
599	as expected (reviewed by Verhulst and Nilsson 2008). One problem with these egg-removal
600	experiments, however, is that the manipulated females paid the cost of producing additional
601	eggs (Visser & Lessels 2001), which could have impaired their subsequent parental effort and
602	thereby also their reproductive success. Other experiments advanced LD by supplementary
603	feeding (e.g. Nager et al. 1997). This manipulation, however, also affects the females'
604	condition (and thus potentially their fitness, independently of changes in LD) and it would be
605	difficult to conclude that LD causally affects fitness from these experiments. Under the

significantly in VE. Other factors such as changes in food supply or habitat may also be

606 'individual optimisation hypothesis', both reducing and enlarging CS should lead to a fitness 607 decline (Nur 1997). Experiments manipulating CS generally found this (Pettifor et al. 1998; 608 Tinbergen and Both 1999) but the fitness decline of enlarged broods (in the case of brood size 609 manipulations) was often smaller than expected, which can be explained by the fact that 610 these females did not incur a cost for egg-production and incubation (Visser and Lessells 611 2001, Monaghan and Nager 1997). 612 While experiments therefore hint at causal relationships with fitness for both LD and CS, the extent of potential biases to selection estimates are more difficult to predict a priori and 613 614 quantitative genetic analysis of correlational data, as we performed here, can help to clarify 615 this. Such approaches applied to mammals (Kruuk et al. 2002; Robinson et al. 2008; 616 Morrissey et al. 2012) indicate that environmental biases to selection can be substantial. 617 Previous quantitative genetic tests in birds have been more equivocal (Sheldon et al. 2003; 618 Gienapp et al. 2006) but based on two-step analyses of PBVs, which are known to be statistically unreliable (Postma 2006; Hadfield et al. 2010). Our analyses were based on a 619 620 statistically robust, one-step animal model approach, as recommended by Hadfield 2008 (see 621 also Morrissey et al. 2010) and recently applied by Robinson et al. (2008), by Morrissey and 622 Ferguson (2011), by Morrissey et al. (2012) and by Tarka et al. (2015). 623

624 Conclusions

625 Our data show that potential for microevolution exists in this population and, crucially, that

626 changing relationships between phenotypes and fitness are underpinned by changing

627 selection on breeding values, which are both essential requirements for adaptive evolution in

628 changing environments (Endler 1986). Future climate change is likely to lead to further

directional selection on *LD* in particular (Gienapp et al. 2014). While phenotypic plasticity

630 will allow for adaptive tracking of environmental change to some extent (Charmantier et al. 631 2008; Vedder et al. 2013), microevolution will be crucial for long-term adaptation and 632 population persistence (Visser 2008; Gienapp et al. 2013). The fact that selection acts on the 633 genetic component of breeding time implies that evolution of *LD* can track climate change, 634 provided the pace of climate change remains within demographically tolerable limits 635 (Gienapp et al. 2013). We cannot however rule out the possibility of small to moderate 636 magnitude biases to selection estimates, and thus environmental change may lead to weaker 637 (or stronger) selection on underlying breeding values than might be predicted based on 638 phenotypic relationships alone. Missing traits were not a major problem in our selection 639 analyses, as indicated by the concordance between predictions based on the STS and MVBE 640 approaches, but it is worth noting that unmeasured phenotypes may themselves have a 641 genetic basis and be targets of section in a changing environment. Combining inferences from 642 quantitative genetic analyses with experimental tests of causality will allow for better 643 forecasting of potential responses to environmental change. Finally, we note that feedbacks 644 between ecology and evolution, or so-called 'eco-evolutionary dynamics', require that 645 ecologically-induced phenotypic selection actually results in microevolutionary responses, 646 which in turn requires that selection acts on the genotypic component of trait variation (as we 647 have shown here) rather than simply on the environmental component. This reinforces the 648 need to better understand how different types of ecological change alter the relationship 649 between breeding values for key traits and fitness.

650

651 Acknowledgements

We thank Arild Husby and Michael Morrissey for valuable discussions. Jarrod Hadfield and
three anonymous reviewers provided very useful criticisms of a previous draft. TER was

- supported by the Beaufort Marine Research Award in Fish Population Genetics funded by the
- 655 Irish Government under the Sea Change Programme.
- 656

657 References

- Both, C., J. M. Tinbergen, and M. E. Visser. 2000. Adaptive density dependence of avian
- 659 clutch size. Ecology 81:3391–3403.
- 660 Bouwhuis, S., O. Vedder, C. J. Garroway, and B. C. Sheldon. 2015. Ecological causes of
- 661 multilevel covariance between size and first-year survival in a wild bird population. J. Anim.
- 662 Ecol. 84:208–218.
- Brinkhof, M. W., A. J. Cavé, F. J. Hage, and S. Verhulst. 1993. Timing of reproduction and
- fledging success in the coot Fulica atra: evidence for a causal relationship. J. Anim. Ecol.
- 665 577–587.
- Burt, A. 1995. Perspective: the evolution of fitness. Evolution 49:1–8.
- 667 Carlson, S. M., C. J. Cunningham, and P. A. Westley. 2014. Evolutionary rescue in a
- 668 changing world. Trends Ecol. Evol. 29:521–530.
- 669 Caro, S. P., A. Charmantier, M. M. Lambrechts, J. Blondel, J. Balthazart, and T. D. Williams.
- 670 2009. Local adaptation of timing of reproduction: females are in the driver's seat. Funct.
- 671 Ecol. 23:172–179.
- 672 Charmantier, A., and P. Gienapp. 2014. Climate change and timing of avian breeding and
- 673 migration: evolutionary versus plastic changes. Evol. Appl. 7:15–28.
- 674 Charmantier, A., R. H. McCleery, L. R. Cole, C. Perrins, L. E. Kruuk, and B. C. Sheldon.
- 675 2008. Adaptive phenotypic plasticity in response to climate change in a wild bird population.
- 676 Science 320:800–803.
- 677 Chevin, L.-M., M. E. Visser, and J. Tufto. 2015. Estimating the variation, autocorrelation,
- and environmental sensitivity of phenotypic selection. Evolution 69:2319–2332.

679	Crow, J. F., and T. Nagylaki. 1976. The rate of change of a character correlated with fitness.
680	Am. Nat. 110: 207–213.

- 681 Daan, S., C. Dijkstra, and J. M. Tinbergen. 1990. Family planning in the kestrel (Falco
- tinnunculus): the ultimate control of covariation of laying date and clutch size. Behaviour
- **683** 114:83–116.
- 684 Dijkstra, C., A. Bult, S. Bijlsma, S. Daan, T. Meijer, and M. Zijlstra. 1990. Brood size
- 685 manipulations in the kestrel (Falco tinnunculus): effects on offspring and parent survival. J.
- 686 Anim. Ecol. 269–285.
- Endler, J. A. 1986. Natural selection in the wild. Princeton University Press, Princeton.
- Estes, S., and S. J. Arnold. 2007. Resolving the paradox of stasis: models with stabilizing
- selection explain evolutionary divergence on all timescales. Am. Nat. 169:227–244.
- Falconer, D. S., and T. F. C. Mackay. 1996. Introduction to Quantitative Genetics. Longman,
- 691 London.
- Fisher, R. A. 1958. The genetical theory of natural selection. 2nd ed. Dover, New York.
- Foulley, J. L. and Im, S. 1993. A marginal quasi-likelihood approach to the analysis of
- Poisson variables with generalized linear mixed models. Genet. Sel. Evol. 23: 101-107.
- Garant, D., J. D. Hadfield, L. E. Kruuk, and B. C. Sheldon. 2008. Stability of genetic
- variance and covariance for reproductive characters in the face of climate change in a wild
- 697 bird population. Mol. Ecol. 17:179–188.
- Gienapp, P., M. Lof, T. E. Reed, J. McNamara, S. Verhulst, and M. E. Visser. 2013.
- 699 Predicting demographically sustainable rates of adaptation: can great tit breeding time keep
- pace with climate change? Philos. Trans. R. Soc. B Biol. Sci. 368:20120289.
- Gienapp, P., E. Postma, and M. E. Visser. 2006. Why breeding time has not responded to
- selection for earlier breeding in a songbird population. Evolution 60:2381–2388.

- 703 Gienapp, P., T. E. Reed, and M. E. Visser. 2014. Why climate change will invariably alter
- selection pressures on phenology. Proc. R. Soc. B Biol. Sci. 281:20141611.
- Gomulkiewicz, R., and R. D. Holt. 1995. When does evolution by natural selection prevent
- 706 extinction? Evolution 49: 201–207.
- Gonzalez, A., O. Ronce, R. Ferriere, and M. E. Hochberg. 2013. Evolutionary rescue: an
- ros emerging focus at the intersection between ecology and evolution. Philos. Trans. R. Soc. B
- 709 Biol. Sci. 368:20120404.
- 710 Hadfield, J. D. 2010. MCMC methods for multi-response generalized linear mixed models:
- the MCMCglmm R package. J. Stat. Softw. 33:1–22.
- Hadfield, J. D., A. J. Wilson, D. Garant, B. C. Sheldon, and L. E. Kruuk. 2010. The misuse of
- 713 BLUP in ecology and evolution. Am. Nat. 175:116–125.
- Holt, R. D. 1990. The microevolutionary consequences of climate change. Trends Ecol. Evol.
- **5:311–315**.
- Husby, A., D. H. Nussey, M. E. Visser, A. J. Wilson, B. C. Sheldon, and L. E. Kruuk. 2010.
- 717 Contrasting patterns of phenotypic plasticity in reproductive traits in two great tit (Parus
- major) populations. Evolution 64:2221–2237.
- 719 Kruuk, L. E. 2004. Estimating genetic parameters in natural populations using the "animal
- model." Philos. Trans. R. Soc. Lond. B. Biol. Sci. 359:873–890.
- 721 Kruuk, L. E., J. Merilä, and B. C. Sheldon. 2001. Phenotypic selection on a heritable size trait
- revisited. Am. Nat. 158:557–571.
- 723 Kruuk, L. E., J. Merilä, and B. C. Sheldon. 2003. When environmental variation short-
- reircuits natural selection. Trends Ecol. Evol. 18:207–209.
- 725 Kruuk, L. E., J. Slate, J. M. Pemberton, S. Brotherstone, F. Guinness, and T. Clutton-Brock.
- 726 2002. Antler size in red deer: heritability and selection but no evolution. Evolution 56:1683–
- 727 1695.

- Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution.
- 729 Evolution 30: 314–334.
- T30 Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters.
- 731 Evolution 37: 1210–1226.
- 732 Lynch, M., and B. Walsh. 1998. Genetics and Analysis of Quantitative Traits. 1st ed. Sinauer
- 733 Associates, Incorporated, Sunderland.
- MacColl, A. D. 2011. The ecological causes of evolution. Trends Ecol. Evol. 26:514–522.
- 735 Merilä, J., B. C. Sheldon, and L. E. B. Kruuk. 2001. Explaining stasis: microevolutionary
- studies in natural populations. Genetica 112:199–222
- 737 Monaghan, P., and R. G. Nager. 1997. Why don't birds lay more eggs? Trends Ecol. Evol.
- 738 12:270–274.
- 739 Morrissey, M. B., P. de Villemereuil, B. Doligez, and O. Gimenez. 2014. Bayesian
- approaches to the quantitative genetic analysis of natural populations. In: Charmantier A,
- Garant D and Kruuk LEB, editors. Quantitative Genetics in the Wild. Oxford University
- 742 Press, Oxford, UK. pp. 228–253.
- 743 Morrissey, M. B., and M. M. Ferguson. 2011. A test for the genetic basis of natural selection:
- an individual-based longitudinal study in a stream-dwelling fish. Evolution 65:1037–1047.
- 745 Morrissey, M. B., L. E. B. Kruuk, and A. J. Wilson. 2010. The danger of applying the
- breeder's equation in observational studies of natural populations. J. Evol. Biol. 23:2277–
- 747 2288.
- 748 Morrissey, M. B., D. J. Parker, P. Korsten, J. M. Pemberton, L. E. Kruuk, and A. J. Wilson.
- 749 2012. The prediction of adaptive evolution: empirical application of the secondary theorem of
- selection and comparison to the breeder's equation. Evolution 66:2399–2410.
- 751 Nager, R. G., C. Ruegger, and A. J. Van Noordwijk. 1997. Nutrient or energy limitation on
- egg formation: a feeding experiment in great tits. J. Anim. Ecol. 495–507.

- Nur, N. 1987. Alternative reproductive tactics in birds: individual variation in clutch size. Pp.
- 754 49–77 in Perspectives in ethology. Springer.
- 755 Pettifor, R. A., C. M. Perrins, and R. H. McCleery. 1988. Individual optimization of clutch
- size in great tits. Nature 336:160–162.
- 757 Postma, E., and A. J. van Noordwijk. 2005. Gene flow maintains a large genetic difference in
- rss clutch size at a small spatial scale. Nature 433:65–68.
- 759 Postma, E. 2006. Implications of the difference between true and predicted breeding values
- for the study of natural selection and micro-evolution. J. Evol. Biol. 19:309–320.
- 761 Postma, E., J. Visser, and A. J. Van Noordwijk. 2007. Strong artificial selection in the wild
- results in predicted small evolutionary change. J. Evol. Biol. 20:1823–1832.
- 763 Price, G. R. 1970. Selection and covariance. Nature 227:520–21.
- 764 Price, T., M. Kirkpatrick, and S. J. Arnold. 1988. Directional selection and the evolution of
- ⁷⁶⁵ breeding date in birds. Science(Washington) 240:798–799.
- Rausher, M. D. 1992. The measurement of selection on quantitative traits: biases due to
- renvironmental covariances between traits and fitness. Evolution 46:616–626.
- 768 Reed, T. E., S. Jenouvrier, and M. E. Visser. 2013. Phenological mismatch strongly affects
- individual fitness but not population demography in a woodland passerine. J. Anim. Ecol.
- 770 82:131–144.
- Robertson, A. 1966. A mathematical model of the culling process in dairy cattle. Anim. Prod.
 8:95–108.
- 773 Robinson, M. R., J. G. Pilkington, T. H. Clutton-Brock, J. M. Pemberton, and L. E. Kruuk.
- 2008. Environmental heterogeneity generates fluctuating selection on a secondary sexual
- 775 trait. Curr. Biol. 18:751–757.

776	Sæther, BE.,	Visser, M.A.,	Grøtan, V	., and Engen.	S. In Press.	Evidence for r	- and K-
	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~			.,			

- selection in a wild bird population: a reciprocal link between ecology and evolution. Proc. R.
- 778 Soc. Lond. B Biol. Sci.
- Scheiner, S. M., K. Donohue, L. A. Dorn, S. J. Mazer, and L. M. Wolfe. 2002. Reducing
- environmental bias when measuring natural selection. Evolution 56:2156–2167.
- Schluter, D. 1996. Adaptive radiation along genetic lines of least resistance. Evolution 1766–
  1774.
- 783 Sheldon, B. C., L. E. B. Kruuk, and J. Merila. 2003. Natural selection and inheritance of
- breeding time and clutch size in the collared flycatcher. Evolution 57:406–420.
- Smouse, P. E., T. R. Meagher, and C. J. Kobak. 1999. Parentage analysis in *Chamaelirium*
- *luteum* (L.) Gray (Liliaceae): why do some males have higher reproductive contributions? J.
- 787 Evol. Biol. 12:1069–1077.
- 788 Stinchcombe, J. R., M. T. Rutter, D. S. Burdick, P. Tiffin, M. D. Rausher, and R. Mauricio.
- 789 2002. Testing for environmentally induced bias in phenotypic estimates of natural selection:
- theory and practice. Am. Nat. 160:511–523.
- 791 Stinchcombe, J. R., A. K. Simonsen, and M. Blows. 2014. Estimating uncertainty in
- multivariate responses to selection. Evolution 68:1188–1196.
- 793 Svensson, E. 1997. Natural selection on avian breeding time: causality, fecundity-dependent,
- and fecundity-independent selection. Evolution 1276–1283.
- 795 Tarka, M., B. Hansson, and D. Hasselquist. 2015. Selection and evolutionary potential of
- spring arrival phenology in males and females of a migratory songbird. J. Evol. Biol. 5:
- 797 1024–1038.
- Tinbergen, J. M., and C. Both. 1999. Is clutch size individually optimized? Behav. Ecol.
- 799 10:504–509.

- 800 Vedder, O., S. Bouwhuis, and B. C. Sheldon. 2013. Quantitative assessment of the
- 801 importance of phenotypic plasticity in adaptation to climate change in wild bird populations.
- 802 PLoS Biol. 11:e1001605.
- 803 Verhulst, S., and J. M. Tinbergen. 1991. Experimental evidence for a causal relationship
- between timing and success of reproduction in the great tit Parus m. major. J. Anim. Ecol.

805 269–282.

- Verhulst, S., and J.-Å. Nilsson. 2008. The timing of birds' breeding seasons: a review of
- 807 experiments that manipulated timing of breeding. Philos. Trans. R. Soc. B Biol. Sci.
- 808 363:399-410.
- 809 Visser, M. E., and C. M. Lessells. 2001. The costs of egg production and incubation in great
- tits (Parus major). Proc. R. Soc. Lond. B Biol. Sci. 268:1271–1277.
- Visser, M. E. 2008. Keeping up with a warming world; assessing the rate of adaptation to
- climate change. Proc. R. Soc. B Biol. Sci. 275:649–659.
- 813 Visser, M. E., L. J. Holleman, and P. Gienapp. 2006. Shifts in caterpillar biomass phenology
- 814 due to climate change and its impact on the breeding biology of an insectivorous bird.
- 815 Oecologia 147:164–172.
- 816 Visser, M. E., A. J. Van Noordwijk, J. M. Tinbergen, and C. M. Lessells. 1998. Warmer
- springs lead to mistimed reproduction in great tits (Parus major). Proc. R. Soc. Lond. B Biol.
- 818 Sci. 265:1867–1870.
- 819 Walsh, B., and M. W. Blows. 2009. Abundant genetic variation+ strong selection=
- 820 multivariate genetic constraints: a geometric view of adaptation. Annu. Rev. Ecol. Evol. Syst.
- **821** 40:41–59.
- Wilson, A. J. 2008. Why h2 does not always equal VA/VP? J. Evol. Biol. 21:647–650.

823	Wilson, A. J., D. Réale, M. N. Clements, M. M. Morrissey, E. Postma, C. A. Walling, L. E.
824	B. Kruuk, and D. H. Nussey. 2010. An ecologist's guide to the animal model. J. Anim. Ecol.
825	79:13–26.
826	

## 828 Figure legends:

Fig.1 Schematic of hypothetical relationships between trait and fitness at the genetic (filled circles, solid lines in insets) and environmental levels (open circles, dashed lines in insets). Each panel corresponds to a different scenario of environmental bias (quantified as  $\beta_E - \beta_G$ ), with three different strengths of phenotypic selection (overall relationship between trait and fitness at phenotypic level) shown in each.

834

835	Fig.2 Relationships between trait and fitness, measured as standardised selection gradients, at
836	the genetic ( $\beta_G$ , filled circles) versus environmental level ( $\beta_E$ , open circles) for years with
837	weak, medium and strong phenotypic selection on each trait in each study population. Shown
838	are posterior modes $\pm$ highest posterior density intervals.
839	

Fig. 3 Power analysis results. Filled circles and solid line: HV population. Open circles anddashed line: VE population.

842

843 Fig.4 Comparing predictions of responses to selection based on the secondary theorem of

selection (STS, grey bars) and multivariate breeder's equation (MVBE, black bars)

845 approaches. Units are phenotypic standard deviations. See main text for explanation of error 846 bars.

- 848 Supplementary Fig. 1: Phenotypic patterns. Top row:  $s_p$  for LD as a function of year for
- 849 each population. Second row:  $s_p$  for CS as a function of year for each population. Third row:
- .1) as a 1 850 mean LD (± standard deviation) as a function of year for each population. Bottom row: mean
- 851 CS (± standard deviation) as a function of year for each population.
- 852
- 853



Strength of relationship (regression slope) between trait and fitness at phenotypic level



Page 38 of 42





**Vlieland East CS** 

Vlieland East LD



- 1 Table 1: Datasets analysed and associated sample sizes. For both study areas, years were split into groups according to variation in standardised
- 2 phenotypic selection differentials (*s_P*) for laying date and clutch size. 'N records' refers to the number of first clutches monitored. 'N females'
- 3 refers to the number of uniquely marked individual females producing those clutches (some females breed in multiple years).

			Laying c	late (LD)		Clutch size (CS)						
	Weak Mediur		Medium	Strong		Weak	Medium	Strong				
		All years	phenotypic	phenotypic	phenotypic	All years	phenotypic	phenotypic	phenotypic			
		LD	selection on	selection on	selection on	CS	selection on	selection on	selection on			
			LD	LD	LD		CS	CS	CS			
Hoge	N years	59	20	19	20	59	20	19	20			
Veluwe	N records	4062	1333	1642	1087	4062	1271	1500	1291			
	N females	2871	1186	1414	960	2871	1154	1238	1093			
	Mean $s_P$	-0.14	0.17	-0.22	-0.55	0.16	-0.16	0.12	0.46			
	Min $s_P$	-1.06	-0.05	-0.34	-1.06	-0.68	-0.68	0.03	0.25			
	Max $s_P$	0.98	0.98	-0.07	-0.36	0.69	0.01	0.24	0.69			
Vlieland	N years	51	17	17	17	51	17	17	17			
East	N records	2714	504	1373	837	2714	863	977	874			
	N females	1663	439	1030	763	1663	747	807	729			
	Mean $s_P$	-0.004	0.29	-0.02	-0.28	0.02	-0.24	0.05	0.26			
	$\operatorname{Min} s_P$	-0.86	0.08	-0.11	-0.86	-0.72	-0.72	-0.01	0.14			
	Max $s_P$	0.64	0.64	0.06	-0.13	0.48	-0.02	0.14	0.48			
							1	>				

5

6

- 1 Table 2: Estimates of the extent of environmental bias to selection ( $\beta_{E} \beta_{G}$ ) based on the trivariate animal models. PS = phenotypic selection.  $\beta_{G}$
- 2 = genetic selection gradient.  $\beta_E$  = environmental selection gradient. Mode = mode of posterior distribution. LCI/UCI = lower/upper highest
- 3 posterior density intervals.

Study area	Trait	Dataset	$\beta_E - \beta_G$ (mode)	$\beta_E - \beta_G$ (LCI)	$\beta_E - \beta_G$ (UCI)	β _G (mode)	β _G (LCI)	β _G (UCI)	β _E (mode)	$\beta_E$ (LCI)	$\beta_E$ (UCI)
Hoge Veluwe	Clutch size	All years	0.01	-0.24	0.32	0.06	-0.15	0.33	0.09	0.04	0.21
		Weak PS	-0.13	-0.42	0.74	-0.31	-0.78	0.23	-0.05	-0.35	0.09
		Medium PS	-0.01	-0.23	0.46	0.10	-0.26	0.29	0.11	-0.01	0.25
		Strong PS	-0.21	-0.59	0.26	0.51	0.14	0.77	0.34	0.10	0.50
	Laying date	All years	-0.03	-0.46	0.21	-0.08	-0.31	0.26	-0.20	-0.24	-0.06
		Weak PS	-0.09	-0.59	0.42	0.29	-0.18	0.63	0.05	-0.09	0.21
		Medium PS	0.02	-0.59	0.43	-0.20	-0.43	0.47	-0.14	-0.32	-0.04
		Strong PS	-0.26	-0.92	0.49	-0.27	-0.92	0.26	-0.56	-0.73	-0.31
Vlieland East	Clutch size	All years	0.11	-0.14	0.27	-0.02	-0.17	0.19	0.08	0.01	0.12
		Weak PS	-0.22	-0.57	0.23	-0.15	-0.27	0.41	-0.15	-0.25	-0.05
		Medium PS	0.10	-0.42	0.32	-0.06	-0.26	0.42	0.02	-0.05	0.13
		Strong PS	0.07	-0.24	0.46	0.25	-0.12	0.47	0.22	0.10	0.32
	Laying date	All years	-0.01	-0.20	0.24	-0.06	-0.26	0.12	0.01	-0.07	0.05
		Weak PS	0.17	-0.33	0.44	0.19	-0.21	0.45	0.21	0.08	0.36
		Medium PS	-0.02	-0.38	0.21	0.03	-0.20	0.28	0.00	-0.12	0.06
		Strong PS	 -0.05	-0.29	0.45	 -0.05	-0.54	0.12	-0.11	-0.25	-0.05