

Phenological mismatch drives selection on elevation, but not on slope, of breeding time plasticity in a wild songbird

Jip J. C. Ramakers,^{1,2} Phillip Gienapp,¹ and Marcel E. Visser¹

¹Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW) 6700, AB Wageningen, The Netherlands

²E-mail: jip.ramakers@gmail.com

Received July 16, 2018

Accepted November 16, 2018

Phenotypic plasticity is an important mechanism for populations to respond to fluctuating environments, yet may be insufficient to adapt to a directionally changing environment. To study whether plasticity can evolve under current climate change, we quantified selection and genetic variation in both the elevation (RN_E) and slope (RN_S) of the breeding time reaction norm in a long-term (1973–2016) study population of great tits (*Parus major*). The optimal RN_E (the caterpillar biomass peak date regressed against the temperature used as cue by great tits) changed over time, whereas the optimal RN_S did not. Concordantly, we found strong directional selection on RN_E , but not RN_S , of egg-laying date in the second third of the study period; this selection subsequently waned, potentially due to increased between-year variability in optimal laying dates. We found individual and additive genetic variation in RN_E but, contrary to previous studies on our population, not in RN_S . The predicted and observed evolutionary change in RN_E was, however, marginal, due to low heritability and the sex limitation of laying date. We conclude that adaptation to climate change can only occur via micro-evolution of RN_E , but this will necessarily be slow and potentially hampered by increased variability in phenotypic optima.

KEY WORDS: Fitness, genetic variation, life-history evolution, phenotypic plasticity, quantitative genetics, natural selection.

Phenotypic plasticity is an important mechanism by which an individual can adapt its phenotype in response to fluctuating environmental conditions (Pigliucci 2001; Schlichting and Pigliucci 1998). For example, life-history traits such as phenology (e.g., timing of breeding or migration) or litter size are often phenotypically plastic, and this plasticity is adaptive as it tracks the environmental variability in the optimal phenotype (Scheiner 1993). Timing of avian seasonal reproduction (or laying date) is an illustrative example in this respect; since laying date is an important determinant of reproductive success and its optimal timing varies from year to year (Visser and Both 2005; Verhulst and Nilsson 2008), it is often phenotypically plastic to environmental conditions, usually spring temperatures (Brommer et al. 2005; Nussey et al. 2005b; Charmantier et al. 2008; Avilés et al. 2014). Similarly, avian clutch size can be phenotypically plastic with respect to population density, such that birds maximize the number of successfully raised, high-quality offspring under

varying levels of food availability and competition (Ricklefs 1980; Both et al. 2000; Sæther et al. 2016). More generally, many forms of animal behavior are highly context dependent in a wide range of taxa (Dingemanse et al. 2010), making phenotypic plasticity ubiquitous in nature (Pigliucci 2001).

Phenotypic plasticity can be described by a reaction norm (Scheiner 1993; Woltereck 1909); that is, the (often assumed to be linear) function of the phenotype against the environment, characterized by the intercept or elevation (i.e., the phenotype in the average environment) and slope (i.e., the sensitivity of the phenotype to the environment). The degree of plasticity may vary among individuals (individual-by-environment interaction or $I \times E$) and this variance may have a partly genetic basis (genotype-by-environment interaction or $G \times E$), making phenotypic plasticity itself an evolvable trait (Scheiner 1993; Van Tienderen and Koelewijn 1994; Via et al. 1995). In a directionally changing environment, evolution of the reaction norm may be necessary because

the environmental driver of the trait no longer accurately predicts future environmental conditions, rendering plasticity alone insufficient to respond to environmental change (Visser 2008). Quantifying variation in reaction norms is therefore imperative for understanding evolutionary processes because it can elucidate whether populations are capable of responding to such directional selection. Predicting such responses may be difficult when $G \times E$ leads to nonlinear changes in genetic variation across environments (Tomkins et al. 2004; Turelli and Barton 2004; Kokko and Heubel 2008), or when genetic variation and selection are negatively correlated with one another (Wood and Brodie III 2016; but see Ramakers et al. 2018). Ultimately, the extent to which phenotypic plasticity modulates evolutionary processes will be highly context dependent (Hoffman and Merilä 1999).

A largely unexplored aspect in the light of directional environmental change is how selection on consumer phenology translates to selection on the reaction norm (Visser 2008). If the optimal reaction norm—that is, the relationship between resource phenology and the environment driving phenology of the consumer (the “cue”)—does not change over time, selection on neither the elevation nor the slope of the consumer reaction norm should occur. If, on the other hand, the sensitivity of the optimal reaction norm to the environmental cues changes over time, this may lead to two scenarios: selection for increased phenotypic plasticity to track the plasticity of the resource (Nussey et al. 2007; Lande 2009; Gienapp et al. 2014), or selection for reduced plasticity when the cue environment changes to such an extent that the consumer can no longer accurately predict the resource phenology (De Jong 1999; Reed et al. 2010). Alternatively still, if the elevation of the optimal reaction norm changes over time, selection on the elevation, rather than the slope, of the consumer reaction norm should occur. Global climate warming has led to the disruption of phenological synchrony between trophic levels (Visser and Holleman 2001; Both et al. 2006; Durant et al. 2007; Both et al. 2009; Schultz et al. 2009; Thackeray et al. 2010; Thackeray et al. 2016; Visser and Gienapp, unpubl. ms.) and in some cases to directional selection on consumer phenology (Van Noordwijk et al. 1995; Visser et al. 1998). It is unclear whether selection on phenology reflects selection on the elevation, the slope or both components of the reaction norm. Most of what we know about the evolutionary dynamics of the reaction norms stems from theoretical work or laboratory experiments (see Scheiner 1993; De Jong 1999; Van Asch et al. 2007; Lande 2009) or from phylogenetic and population comparisons of reaction norms (Murren et al. 2014), with very few empirical examples of how selection as well as (additive) genetic variation led to an evolutionary change in the reaction norm in wild populations (cf. Carter et al. 2017; Van Asch et al. 2013).

Several long-term, vertebrate study populations have been shown to exhibit phenotypic plasticity in phenology (e.g., Réale et al. 2003; Brommer et al. 2005; Nussey et al. 2005a,b;

Charmantier et al. 2008). Evidence for $I \times E$ and $G \times E$, however, is overall mixed. For example, phenotypic plasticity of laying date against spring temperature in Dutch great tits (*Parus major*) was shown to vary between individuals ($I \times E$), with part of this variation being heritable ($G \times E$; Nussey et al. 2005b). In a UK great tit population, on the other hand, there was no $I \times E$ (Charmantier et al. 2008). In a reanalysis for both populations, phenotypic, but not genetic, variation was found to be present in the elevation as well as the slope of the reaction norm in both populations (Husby et al. 2010).

In general, heritable variation in phenological traits is widespread (e.g., van Noordwijk et al. 1981; Blondel et al. 1990; Réale et al. 2003; Sheldon et al. 2003), which in the absence of variation in plasticity slopes should reflect variation in the elevation of the reaction norm. Evidence for $G \times E$ in wild populations is, however, rare (Wood and Brodie III 2016; Hayward et al. 2018; but see Ramakers et al. 2018), and estimates of selection on reaction norms are even scarcer. This is because it requires estimating selection on both the elevation and the slope, which is a statistically challenging procedure (see discussion in Weis and Gorman (1990) and in Brommer et al. (2012)). Several studies used random regression techniques to get individual estimates (best linear unbiased predictions, or BLUPs) for reaction norm elevation and slope and performed a separate analysis on these components to quantify selection on them (e.g., Brommer et al. 2005; Nussey et al. 2005b). Such a two-step approach is now considered inappropriate (Hadfield et al. 2010; Morrissey and Liefing 2016) and an alternative method has been suggested to use random regression models to estimate variation in reaction norms as well as selection thereon in a single analysis (Brommer et al. 2012). Application of this method to estimate selection on reaction norms in general has been rare (Brommer et al. 2012; Hayward et al. 2014). Thus, quantitative estimates of both (genetic) variation in and selection on reaction norms in wild populations are rare or ambiguous, and there is a clear need to empirically quantify the evolutionary dynamics of reaction norms in the light of environmental change (Visser 2008).

To address this gap, we quantified selection on and predicted evolution of the reaction norm of timing of breeding (laying date) in response to temperature (an important environmental cue; Schaper et al. 2012; Visser et al. 2009) in a Dutch long-term (1973–2016) study population of the great tit (*Parus major*) at the Hoge Veluwe. Laying date is a labile trait that can be expressed several times by an individual; this means that each female has her own reaction norm. Timing of breeding is under increased directional selection (for earlier dates) in this population due to a climate change-driven mismatch with the caterpillar food peak (Visser et al. 1998; Reed et al. 2016). We tested (i) whether the optimal (linear) reaction norm—determined by regressing the caterpillar biomass peak date against the temperature used by great tits to time laying date—changed over time, both in its elevation and

slope; (ii) whether a change in the optimal reaction norm over time led to selection on the phenotypic reaction norm of the consumer (great tit); and (iii) whether there was genetic variation in great tit reaction norms. Based on the results, we used a quantitative genetic model to predict quantitatively the amount of expected change in the population reaction norm (elevation and slope) due to years of selection, and verified the outcome by comparing the predicted reaction norms with those observed in the wild. Combined, these results should elucidate whether the breeding-time reaction norm can evolve given sustained directional selection.

Methods

DATA COLLECTION AND PREPARATION

Data were collected from a long-term great tit (*Parus major*) population at the Hoge Veluwe National Park (HV; 52°23'N, 05°51'E, central Netherlands). The HV population is situated within a matrix of natural habitat, facilitating dispersal from and into the study area, and has been monitored continuously since 1955. Nest boxes are provided in excess (~400) in suitable habitat. Each breeding season from April to July, boxes are checked at least once a week to monitor the breeding activity of hole-breeding passerines. Clutch size is noted and laying date (i.e., the date when a female's first egg of her first clutch in that season is laid) is calculated based on the number of eggs in the nest, assuming that one egg is laid per day. During the chick-feeding phase, parents and chicks are captured at the nest box and ringed, allowing for establishing a "social" pedigree. Extra-pair paternity in the neighboring population of Westerheide ranges from 6.5% to 12.5% of all chicks (Van Oers et al. 2008), a common rate for tit species (Brommer et al. 2010) that has been found to only marginally affect the accuracy of heritability estimates when sample sizes are sufficiently large (Charmanier and Réale 2005; Firth et al. 2015).

Temperature data were retrieved as daily averages from a nearby weather station of the Royal Dutch Meteorological Institute (KNMI; Deelen station: 52°05'N, 5°87'E; <http://projects.knmi.nl/klimatologie/daggegevens/>). Temperature data were averaged over the period from March 11 to April 20, which is the time window yielding the strongest correlation between annual mean daily average temperature and annual mean laying date ($r^2 = 0.74$). This was done using a sliding-window approach in the "climwin" R package (Bailey and Van de Pol 2017); details of the analyses are given elsewhere (Bailey et al. in prep.). We used these mean daily average temperatures as proxies for the environmental cues great tits use to time their reproduction, which in reality is an intricate interplay between day length and changes in temperature over time (Gienapp et al. 2005; Schaper et al. 2012).

The peak date of food availability has been estimated since 1985, using frass samples (see Visser et al. 2006 for details). The most common caterpillars are of the winter moth (*Operophtera*

brumata) and the oak leaf roller (*Tortrix viridana*), although other species are present. Caterpillar peak date correlates very well with the mean temperature from March 22 to May 16 ($r^2 = 0.81$, resulting from a sliding window analysis); we used this relationship to hindcast the caterpillar peak date from 1973 to 1984 (all temperatures fell within the analyzed range).

Here, we consider brood data from 1973 to 2016, as the study area was reorganised following a major storm in 1972, and the latest data on recruitment of fledged offspring (our fitness measure) was available for 2017. During the study period, a number of broods were manipulated in brood size. None of the broods that we considered were affected in their laying date, but brood size manipulations could affect the reproductive success of a brood. We therefore included broods whose size was manipulated in the laying date analyses, but removed them from annual selection analysis (but not for analysis of selection on the reaction norm; see below). Sample sizes per analysis are given in Table 1.

CHANGE IN THE RELATIONSHIP BETWEEN OPTIMAL PHENOTYPE AND THE ENVIRONMENT

If the optimal laying date (LD_0 , i.e., the laying date at which fitness is maximized) advanced at a faster rate over time than the observed laying date, this could lead to selection on (1) the reaction norm elevation, if the sensitivity of the optimal reaction norm to temperature (i.e., plasticity) remained constant over time, or (2) the slope, if the sensitivity of LD_0 to temperatures changed over time. We therefore determined LD_0 in each year and regressed it against the temperature cues used by the birds (i.e., the mean temperature across March 11–April 20). LD_0 was defined as the peak date of caterpillar biomass minus 33 (see Chevin et al. 2015). The rationale behind this was that it takes approximately 33 days from the laying date to the moment when chicks' food requirements are highest, assuming a clutch size of 9–10, followed by 12–13 days of incubation, and a peak in food demands at chick age 9–11 d (e.g., Keller and Van Noordwijk 1994; Mols et al. 2005).

To test whether the optimal reaction norm changed over time, we divided the datasets into three equal-interval time periods (spanning 14–15 years: 1973–1987, 1988–2001, 2002–2016; Table 1). We then fitted a linear model with LD_0 as a function of temperature and time period. We assessed statistical significance of model terms by bootstrapping (1000 iterations, bias-corrected, and accelerated confidence intervals) estimates for each period, as well as the temperature slope. We then ran a model that contained an interaction between temperature and period and bootstrapped the change in the slope of the optimal reaction norm between periods to determine whether the slope changed over time. The year 1991 was excluded from this analysis as an exceptionally late frost spell in that year damaged all fresh oak leaves and hence the food peak was exceptionally late (see also Visser et al. 2002, 2006).

Table 1. Sample sizes used in all analyses.

Analysis	Year span	N_{years}	N_{females}	N_{broods}
1. Optimal laying date vs. environment	1973–1987	15	—	—
	1988–2001	14	—	—
	2002–2016	15	—	—
2. Selection on plasticity ^a	1969–1987	19	456	1272
	1988–2001	14	365	953
	2002–2012	11	280	691
3. Quantifying G -matrix	1973–2016	44	3028	4890
4. Selection (β_z) on laying date ^b	1973–2016	44	>2347	3662
5. Observed reaction norms	1973–1987	15	1026	1650
	1988–2001	14	993	1551
	2002–2016	15	1126	1689

^aYear span here indicates cohort span; birds with incomplete lifetime reproductive success (LRS) at the end of the dataset were omitted, whereas LRS of birds breeding in 1973 were complemented with brood data from previous years, hence making the cohort span 1969–2012 (see text). Only birds with ≥ 2 breeding events were included here.

^bReduced dataset without manipulated broods; exact N_{females} is unknown because this analysis includes broods whose mother could not be identified.

SELECTION ON THE SLOPE AND ELEVATION OF THE GREAT TIT REACTION NORM

To estimate period-specific (fecundity) selection on plasticity, we assigned each female with at least two breeding attempts to a breeding cohort (based on the first year a bird was observed breeding) and split the phenotypic dataset into three time periods as above. A female's lifetime reproductive success (LRS) was the sum of all recruits she produced over her lifetime. To avoid truncation of LRS of birds breeding in 1973, we added broods from earlier years (before 1973) to complete each individual's LRS. For the same reason, we removed all observations of birds from the 2013 breeding cohort or later, as some birds from these cohorts were still known to be breeding in 2017 and therefore had incomplete LRS. Hence, the periods in this analysis were 1969–1987, 1988–2001 and 2002–2012 (Table 1).

As pointed out above, some broods had been manipulated, likely affecting the female's fitness. Since we were interested in the lifetime fitness consequences of being more or less phenotypically plastic with respect to temperature, as well as having a higher or lower mean response (i.e., the elevation of the reaction norm), we opted to include manipulated broods in this analysis. The reason for this was that because the probability of having any of her broods manipulated increases with a female's age, discarding data from manipulated females would lead to a biased subset of shorter lived individuals. Consequently, the inclusion of manipulated broods in our analysis may create some noise in the selection estimates, but we nevertheless believe it is a superior approach to removing a nonrandom subset of individuals from the dataset.

For each of the three periods, we fitted a bivariate random regression model (RRM) in “MCMCglmm” (Hadfield 2010;

Hadfield 2014) with laying date as a Gaussian and LRS as a (overdispersed) Poisson trait. Breaking down this bivariate model in scalar notation, the laying date (z) of the i^{th} individual in the j^{th} year in the k^{th} nest box within the l^{th} “environmental block” was modelled as

$$z_{ijkl} = \alpha_z + a_i + b_i (T_{ij} - \bar{T}_i) + b\bar{T}_i + age_{ij} + nb_k + yr_j + e_{z,ijl}, \quad (1)$$

where α_z is the overall mean laying date (intercept), a_i and b_i are the individual intercept and slope, respectively, related to temperature (T) and b is the population-level slope, age_{ij} is the female's age (first-year breeder, older, or unknown), nb_k and yr_j are the k^{th} nest box and j^{th} year, respectively (treated as random effects with estimated variance $nb_k \sim N(0, \sigma_{nb}^2)$ and $yr_j \sim N(0, \sigma_{yr}^2)$) and $e_{z,ijl}$ is the residual term. This residual term was estimated as $e_{z,ijl} \sim N(0, \sigma_{e,l}^2)$, where l is one of two equal-interval groups of years with similar temperatures, which was done to accommodate changes in residual variance along the temperature gradient (Lillehammer et al. 2009; Nicolaus et al. 2013). Temperature was divided into a within-individual ($b_i(T_{ij} - \bar{T}_i)$) and a between-individual ($b\bar{T}_i$) component, following Van de Pol and Wright (2009). This method disentangles any effect of temperature on laying date caused by having observed certain individuals only under certain temperatures and others under different temperatures (a between-subject effect of T on z) from a real, within-subject effect (but note that for the purpose of estimating variance in intercepts, centring may not be desired in all context; see Kreft et al. (1995)).

A female's LRS (W) is described as

$$\log(E[W]_i) = \alpha_W + c_i + e_{W,i}, \quad (2)$$

where α_W is the overall mean LRS, c_i is the individual intercept and $W_i \sim \text{Poisson}(E[W]_i)$. We set the residual term, $e_{W,i}$, to a fixed small variance (0.01) since LRS is a repeated measures trait and laying date is not; ideally, we would constrain variance to 0 (Brommer et al. 2012; Morrissey et al. 2012), but since the MCMC chain does not mix well under these conditions a minimum nonzero value needs to be assigned.

Variance in intercepts (a) and slopes (b) of the reaction norm (eq. (1)), as well as in intercepts of LRS (c ; eq. (2)) were jointly fitted as the random effect “female identity,” such that individual values were estimated as

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix}_i \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_a^2 & \sigma_{a,b} & \sigma_{a,c} \\ \sigma_{a,b} & \sigma_b^2 & \sigma_{b,c} \\ \sigma_{a,c} & \sigma_{b,c} & \sigma_c^2 \end{bmatrix}_i \right),$$

that is, drawn from a 3×3 unstructured, phenotypic covariance matrix with the variance components on the diagonals and the covariances between them on the off-diagonals.

To obtain independent samples from the MCMC sampling process, we used a thinning interval of 10^4 , a burn-in period of 10^5 and a total sample size of 10.1×10^6 (i.e., an effective sample size of 1000). We used default normal priors for the fixed effects and parameter-expanded priors for the random terms ($V = \text{diag}(n)$, $\text{nu} = n$, $\text{alpha.mu} = 0$, $\text{alpha.V} = \text{diag}(n) \times 25^2$, n being the dimension of the matrix), following Hadfield (2014). For the residual variance in equation (1), we used inverse-Wishart priors with $V = \text{diag}(2)$ and $\text{nu} = 0.002$.

PREDICTING CHANGES IN THE REACTION NORM

We used a multivariate version of the breeder’s equation to predict changes in the reaction norm as a result of selection across the three time periods (Table 1). The evolutionary response to selection from one year to the next is defined as the trait heritability times the selection differential ($R = h^2S$, i.e., the breeder’s equation; Falconer and Mackay 1996). Even in its multivariate form (Lande 1979), however, this would require estimating the heritability of both components of the reaction norm (elevation and slope), which is not straightforward (Hadfield et al. 2010). We therefore used a derivation of the breeder’s equation for reaction norms (Van Tienderen and Koelewijn 1994). In their Appendix 1, Van Tienderen and Koelewijn (1994) show that the selection gradients on reaction norm components, that is intercept and slope in our case, can be calculated from environment-specific—here hence annual—selection gradients:

$$\beta_g = \Phi^t \beta_z, \quad (3)$$

where β_g is the vector of selection gradients on intercept and slope, Φ^t is the transposed matrix of vectors consisting of the environmental values in environments $i \dots k$ with a leading 1, and

β_z the vector of selection gradients in environments $i \dots k$. The (expected) genetic change in the breeding values for intercept and slope is given by the multivariate breeders’ equation:

$$\Delta \bar{g} = G\beta_g, \quad (4)$$

where $\Delta \bar{g}$ is the change in intercept and slope breeding values, and G the additive genetic variance-covariance matrix for intercept and slope. Substituting equations (3) and (4) allows the change in reaction norm components ($\Delta \bar{g}$) as a result of selection in year j following

$$\Delta \bar{g} = G[1, x_j]^t \beta_{z_j}, \quad (5a)$$

where $[1, x_j]^t$ is a transposed vector characterising the environment in a given year (x_j), and β_{z_j} the selection gradient. We then only need to accommodate generation time by multiplying the response by the proportion of the breeding population in environment $j+1$ represented by recruits ($p_{recr_{j+1}}$) and further halve the response because laying date is only expressed by females (following Gienapp et al. 2006):

$$\Delta \bar{g} = G[1, x_j]^t \beta_{z_j} \times p_{recr_{j+1}} \times 0.5. \quad (5b)$$

To estimate G , we fitted a univariate RRM on the entire dataset (i.e., all years and all observations) that included an additive genetic term (“animal model,” RRAM; Henderson 1988; Kruuk 2004) in “MCMCglmm”. Random effects were as in equation (1), with the addition of an additive genetic term (via the pedigree) to estimate additive genetic variance in laying date. The difference was that we allowed both the permanent environment term (i.e., female identity) and the additive genetic term to vary with grand-mean-centred, as opposed to individual-mean-centred, temperature. Thus, the RRAM took the form:

$$z_{ijkl} = \alpha_z + a_i + A_i + (b + b_i + B_i)(T_{ij} - \bar{T}) + age_{ij} + nb_k + yr_j + e_{z,ijl}, \quad (6)$$

where definitions and indices are as for equation (1), but where the individual intercepts and slopes of the laying date–temperature reaction norm are now estimated on the grand-mean-centred temperature (\bar{T}) and partitioned into a permanent environment (a_i and b_i) and additive genetic component (A_i and B_i ; breeding values). Phenotypic and additive genetic variance in these components was estimated using two separate, 2×2 unstructured phenotypic and additive genetic (G) variance–covariance matrices containing σ_a^2 and σ_b^2 or σ_A^2 and σ_B^2 , respectively, as well as the covariance between each component. The residual term was estimated from a 4×4 matrix to allow for independent and heterogeneous variance as in equation (1), where 4 is the rounded number of years divided by 10, grouped based on temperature. We extracted posterior medians from G , along with 95% HPDIs (note that because variance

estimates are constrained to be positive, they have a skewed posterior sample distribution when close to 0, hence making the median a more appropriate point summary than the mean). We applied the same prior structure and sampling procedures as for equation (1). Estimates resulting from the RRAM were robust to excluding individuals with only one or only two breeding records.

To estimate the selection gradient, β_{z_j} , we ran annual Generalized Additive Models (GAMs; package “mgcv”; Wood 2017) on unmanipulated broods, with annual reproductive success (ARS; number of recruited offspring) as response variable with a negative binomial distribution and laying date as the predictor variable. We then used the “gam.gradients” function from the “gsg” package (Morrissey and Sakrejda 2015) to calculate β_{z_j} (see Lande and Arnold 1983) along with the standard error through parametric bootstrapping (1000 iterations). Females with unknown identity were included in these analyses as they comprise a potentially biased subset of individuals that laid their eggs too early or too late, in some cases leading to brood desertion before the nestling stage when adults could be identified.

With G and β_z in place, we could predict the evolution of the reaction norm throughout the study period. Since the reaction norm has to be estimated across years (because each year has only one temperature average and one breeding event), we could not predict change from the first year onward. Instead, we estimated the observed population reaction norm across years in period 1 (see Table 1) using a slightly modified version of equation (1) ($z_{ijkl} = \alpha_z + a_i + b(T_{ij} - \bar{T}_i) + b\bar{T}_i + age_{ij} + nb_k + yr_j + e_{z,ijl}$, i.e. a random-intercept model) and predicted the evolutionary outcome of selection in all subsequent years (i.e., excluding years in period 1) using equation (5b). Note that the predicted reaction norm after selection is then a property of a single year and will hence never accurately match the observed reaction norm, which is necessarily estimated across multiple years. We therefore view the predicted change as the upper limit of possible change within the studied period.

Since both G and β_z came with estimation errors, we accommodated uncertainty in predicted cumulative change in the following way. We used the upper and lower 95% HPDI of G to calculate the variance (V) for $G[1, x_j]^t$ in each year by squaring its standard error (half the 95% HPD range divided by 1.96); similarly, we squared the standard error of β_{z_j} to get its variance. The standard error of the full product was then calculated following simple error-propagation rules (Taylor 1997):

$$SE_{\Delta \bar{g}} = \sqrt{(G[1, x_j]^t)^2 V_{\beta_{z_j}} + \beta_{z_j}^2 V_{G[1, x_j]^t}}, \quad (7)$$

which, in the case of equation (5b), was in turn multiplied by $p_{recr_{j+1}} \times 0.5$. We calculated the cumulative difference across years ($\Delta \bar{g}_{cum.}$) by summing $\Delta \bar{g}$ of each year; the 95% CI was calculated as $\Delta \bar{g}_{cum.} \pm 1.96 \times (SE_{\Delta \bar{g}_1}^2 + \dots + SE_{\Delta \bar{g}_n}^2)^{0.5}$, where n

is the number of years in period 2 and 3 combined. We need to make the cautionary note that although this way of calculating of errors should work satisfactorily when posterior distributions approximate normality, it may cause upward bias (i.e., be conservative) for variance estimates that are skewed because they are bounded at 0.

We visually compared the predicted reaction norm with the observed reaction norms in each of the three periods, which we estimated in “MCMCglmm” using a trimmed version of equation (6) as described above and using the same priors and MCMC sampling procedures as described for equation (1). Although we cannot compare the predicted and observed reaction norms quantitatively for the reasons given above, we do present them alongside one another as the latter serves as an empirical validation of the former.

Results

CHANGE IN OPTIMAL REACTION NORM OVER TIME

We found that the optimal reaction norm for laying date (estimated as the date of the caterpillar biomass peak versus the temperatures in the time window as used by great tits to time their reproduction) changed to earlier dates over three distinct periods (Fig. 1A); the optimal phenotype (LD_0) advanced by 7.72 days (95% CI: [-10.89, -4.79]) from the first to the second period, and by 8.45 days (95% CI: [-12.75, -5.13]) from the first to the third period. There was, however, no interaction between temperature and time period (change in slope coefficient from first to second period: -1.54 [-4.19, 2.15]; from the first to the third period: 0.58 [-2.10, 3.31]), indicating that the response of LD_0 to temperature did not change over the three periods (slope coefficient across all years: -3.15 [-4.36, -1.76] days/°C). Besides the mean, inter-annual variance in optimal laying date changed across the three periods (24.31 [10.82, 50.36], 29.98 [15.09, 64.27] and 63.70 [30.56, 149.37], respectively).

SELECTION ON THE GREAT TIT REACTION NORM

Along with a change to earlier dates in the elevation of the optimal reaction norm, we found statistical evidence for directional selection on the elevation of the great tit reaction norm in period 2; that is, lifetime reproductive success (LRS) covaried negatively with the elevation, indicating selection for a lower reaction norm (Table 2). Despite overwhelmingly negative annual selection gradients (Table S1), this selection on the intercept weakened in period 3, potentially related to the high variability in phenotypic optima in that period. We found no evidence for directional selection on slopes, and estimates of slope variation were overall small and zero-bound (Table 2). This low slope variance was also found when we used the entire 1973–2016 dataset (Table 3; see Discussion).

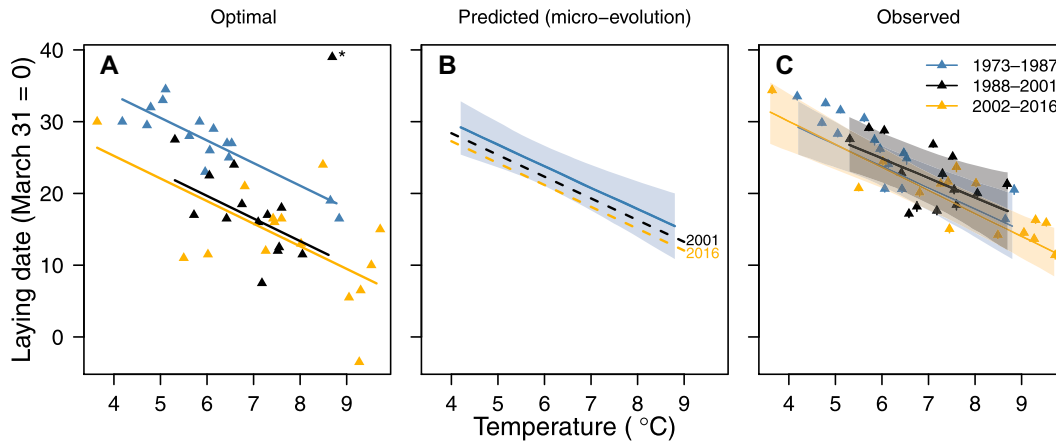


Figure 1. Optimal (A), predicted (B), and observed (C) reaction norms of laying date against spring temperature in three consecutive periods (blue: 1973–1987; black: 1988–2001; orange: 2002–2016) in the HV great tit population. (A) The optimal laying date for each year was the caterpillar peak date minus 33 days. Lines are estimates from linear regressions, excluding 1991 (marked by an asterisk) because late frost damaged the oak leaves in that year. (B) The black and orange dashed lines are the predicted evolutionary deviation from the observed reaction norm in period 1 (solid blue line) by the end of the second (2001) and third (2016) time period, respectively, based on cumulative change due to annual selection (eq. (5b); see Methods for interpretation). (C) Observed laying dates are yearly averages (± 1 SEM); solid lines and shadings are the regression lines and 95% HPDI regions from a univariate mixed-effects model of laying date against temperature (i.e., the mean individual-level slope).

Table 2. Posterior medians (and 95% HPDIs) of the phenotypic (co)variance matrix resulting from an analysis of selection on the reaction norm of laying date (LD) against temperature (via lifetime reproductive success (LRS); eqs. (1) and (2)) for three distinct cohort periods in the HV great tit population, excluding females with only one observation (see Table 1 for sample sizes).

	LD _{intercept}	LD _{slope}	LRS
<i>Period 1</i>			
LD _{intercept}	6.48 (4.78, 8.09)		
LD _{slope}	0.29 (−0.26, 1.11)	0.15 (0.00, 0.77)	
LRS	−0.32 (−0.74, 0.11)	−0.02 (−0.25, 0.20)	1.12 (0.89, 1.40)
<i>Period 2</i>			
LD _{intercept}	9.88 (7.67, 12.62)		
LD _{slope}	1.00 (−0.32, 2.80)	0.43 (0.00, 1.61)	
LRS	−0.98 (−1.58, −0.32)	−0.24 (−0.76, 0.15)	1.51 (1.14, 1.97)
<i>Period 3</i>			
LD _{intercept}	7.92 (5.52, 10.70)		
LD _{slope}	−0.24 (−1.04, 0.55)	0.21 (0.00, 0.69)	
LRS	−0.53 (−1.15, 0.13)	−0.22 (−0.46, 0.05)	1.23 (0.85, 1.62)

Variance estimates are given on the diagonals, whereas covariance estimates are on the off-diagonals. Covariance estimates whose HPDI did not include zero are marked in bold.

PREDICTED AND OBSERVED GREAT TIT REACTION NORMS

There was ample additive genetic variation in the elevation, but not the slope, of the great tit reaction norm (Table 3). Based on selection patterns in Table 2, we may expect the reaction norm to evolve toward an earlier average over time, but its slope to change only marginally because of the weak selection on it. Using the quantitative genetic model (eq. (5b)), we found that both the elevation and the slope of the reaction norm were expected to evolve only marginally over time (Fig. 1B; cumulative change in

elevation across years in period 2 and 3: -2.34 [$-4.20, -0.48$] days; in slope: -0.04 [$-0.09, 0.02$]; see Table S1 for annual predicted responses). The small predicted response in the elevation—compared to the change in the optimal reaction norm—despite strong selection in period 2 was due to the generation time of about two years and the sex limitation of the trait; when we disregarded these factors (eq. (5a)), the predicted response in elevation was substantially stronger (cumulative change in elevation: -10.30 [$-18.63, -1.98$] days; in slope: -0.16 [$-0.40, 0.08$]; Table S1).

Table 3. Model estimates resulting from the RRAM (eq. (6)) quantifying the G matrix for great tit laying date in the HV population between 1973 and 2016 (see Table 1 for sample sizes).

Parameter	Posterior median	95% HPDI	
<i>Fixed effects</i>			
Intercept	21.84	20.82	22.77
Age (old)	—	—	—
Age (unkn.)	1.27	0.42	2.13
Age (young)	1.78	1.51	2.05
Temperature (T_c)	-3.28	-3.92	-2.68
<i>Random effects</i>			
V_{year}	9.65	5.92	14.45
V_{PE}	3.24	0.88	5.68
$Cov(\beta_{0PE}, \beta_{1T_c})$	0.06	-0.16	0.47
$V_{PE \times T_c}$	0.02	0.00	0.15
V_A	4.38	1.98	6.70
$Cov(\beta_{0A}, \beta_{1T_c})$	0.07	-0.20	0.43
$V_{A \times T_c}$	0.02	0.00	0.15
V_{NB}	1.22	0.85	1.66
$V_{R_{3.64-5.16C}}$	11.09	9.40	13.05
$V_{R_{5.16-6.68C}}$	13.63	12.20	14.99
$V_{R_{6.68-8.20C}}$	19.21	17.35	21.17
$V_{R_{8.20-9.72C}}$	19.27	16.32	21.90

T_c = mean-centred temperature; V_x = variance component associated with each random effect (PE = permanent environment, A = additive genetic, NB = nest box; R = residual); $Cov(\beta_0, \beta_1)$ = intercept-slope covariance.

The permanent-environment and additive genetic (co)variance components are marked in bold.

In close agreement with the predicted reaction norms, the observed reaction norm showed no distinct advancement over the three periods (Fig. 1C), with largely overlapping 95% HPDIs for the intercepts (posterior medians and 95% HPDIs for period 1–3: 41.80 [33.40, 49.29], 41.27 [30.72, 53.50], and 42.73 [35.85, 50.81]).

Discussion

We studied whether the optimal reaction norm of great tit laying date against temperature changed over three consecutive time periods in a long-term study population and whether this resulted in intensified directional selection on—and an evolutionary response in—its two components (elevation and slope, i.e., the sensitivity to the environmental variable). We found that, whereas the optimal laying date (as predicted by spring temperatures and determined by the timing of the caterpillar peak) across the temperature range (the elevation of the optimal reaction norm) advanced over the past decades, the sensitivity of the optimal laying date to temperature (the slope of the optimal reaction norm) did not change. In agreement with this, there was selection on the great tit reaction

norm elevation, but not on the slope, in the second third of our time series, but this selection waned in the third period, potentially due to the increased variability in phenotypic optima (Fig. 1A). Despite directional selection, we predicted quantitatively that the elevation of the reaction norm would shift only marginally over time (a maximum of a few days), simply because of the low heritability, the generation time of about two years (Garant et al. 2004; Kvist et al. 2007) and the sex limitation of laying date (Caro et al. 2009). Indeed, the actual (observed) reaction norm did not change over time, neither in slope nor elevation. Our results suggest that, despite the apparent lack of an evolutionary response in our population, adaptation of timing of breeding to climate change can only occur through evolution of the elevation of the reaction norm, as evolution of increased (or decreased) phenotypic plasticity is not possible.

Selection on plasticity in our study population was limited because the slope of the optimal reaction norm did not change over time. However, even if there had been substantial directional selection on the slope, the response to selection would have been weak as there was very little variation in individual reaction norm slopes, both phenotypically and additive genetically (Tables 2 and 3). This qualitative statement can be quantified by comparing the variances in environment-specific expected values of laying date attributable separately to variation in reaction norm elevations (i.e., the estimates from Table 3) and slopes (i.e., estimated variance multiplied by the variance in centred temperatures; see Appendix 7 in Morrissey and Liefting 2016). This was 3.24 [0.88, 5.68] and 4.38 [1.98, 6.70] for the phenotypic and additive genetic intercepts, respectively, whereas it was only 0.05 [0.00, 0.34] and 0.04 [0.00, 0.34] for the respective slopes. Thus, the contribution of variance in slopes to variance in expected laying dates in a given environment was negligible.

The lack of variation in (phenotypic) plasticity slopes ($I \times E$) contradicts previous work on our study population (Nussey et al. 2005b; Husby et al. 2010, 2011). The source of this discrepancy lies in the residual variance structure of the random regression models. Besides the problems inherent to using BLUPs in this context, Nussey et al. (2005b) fitted homogeneous residual variance, thereby forcing any heterogeneity in residual variance caused by the environment (i.e., temperature) to be estimated by the phenotypic covariance matrix and hence inflating estimates of $I \times E$. Husby et al. (2010, 2011), on the other hand, fitted a heterogeneous residual structure but grouped years by decade, assuming that variance in laying date increased over the years because temperature increased over time. This assumption does not hold up, however, as we found that the variance in laying date in HV correlated very weakly with year (slope = 0.001 [-0.22, 0.22]; r^2 = 0.00 [0.00, 0.001]), whereas it correlated reasonably well with temperature (2.40 [0.70, 4.49]; r^2 = 0.13 [0.01, 0.36]). In general, it would make more sense to partition

residual variance based on years with similar environments rather than based on decades. Indeed, when we fitted the model of equation (6) (but dropping the $G \times E$ term for efficiency) on the same data subset (1973–2006) as in Husby et al. (2010, 2011) and partitioned residual variance into three groups by decade, we found substantial individual variation in slopes (posterior median slope variance and 95% HPDI: 0.91 [0.36, 1.53]). When we refitted the model on the same data subset (1973–2006) with residual variance partitioned into three groups based on temperature, however, the slope variance decreased substantially (0.25 [0.00, 0.87]). These results were confirmed by simulations; when we simulated a population with a small slope variance, specifying the incorrect residual structure in the random regression model led to upward bias in the variance estimate and a high false-positive rate (Supplementary Info 2, Fig. S1). Thus, we conclude that variation in plasticity in HV really is limited, reinforcing the notion that, despite previous studies (Nussey et al. 2005b), there can be only limited response to selection on plasticity in this population (cf. e.g., Brommer et al. 2012; Hayward et al. 2014).

Although we can only speculate about the reasons for an apparent lack of $I \times E$ and $G \times E$ in our population, previous simulation studies suggest it is unlikely that we had limited statistical power (or precision in our case) to detect either (Martin et al. 2011; Van de Pol 2012). Illustrative in this respect are the narrow 95% HPDIs for slope variances (Table 3), which seem to indicate that the RRAM was able to estimate $I \times E$ (strictly speaking $PE \times E$ and $G \times E$) with a fair amount of precision. To put this in perspective, when we applied a version of the model of equation (6) (that included individual-mean-centred temperatures) to another of our long-term great tit study populations on the Dutch island of Vlieland (VL), we found substantial evidence for $I \times E$ in laying date (posterior median slope variance and 95% HPDI: 1.32 [0.53, 1.89]), but not $G \times E$ (0.10 [0.00, 0.67]; unpublished results), both estimated with substantially lower precision (the zero constraint of the estimates notwithstanding). The HV and VL datasets are quite similar in terms of the number of individuals and observations and we should therefore have been able to detect $I \times E$ in HV if it was really there. Since the pedigree in VL is more informative (“deeper”) than the one in HV because of limited immigration from the mainland (Gienapp et al. 2013b; Postma and Van Noordwijk 2005), the VL data may be better suited to test for $G \times E$, but even there $G \times E$ appeared to be limited. (No caterpillar biomass data are available for that population, making it an unsuitable dataset for the purpose of this article.) This apparent absence of $G \times E$ is consistent with the notion that $G \times E$ in general is hard to detect in wild populations when not in an experimental setting (Gienapp and Brommer 2014; Wood and Brodie III 2016; Hayward et al. 2018; Ramakers et al. 2018). Nevertheless, since $I \times E$ is the upper limit of $G \times E$ (Gienapp and Brommer 2014), the absence of $I \times E$ in our study

system in HV suggests that $G \times E$ is absent. The reason for this absence may be that early-spring temperatures have historically been highly predictive of the food peak and selection for being well matched with this peak has been strong (Reed et al. 2013b; Visser et al. 2006), possibly eroding (genetic) variation in plasticity in laying date over time (cf. Tomkins et al. 2004; Turelli and Barton 2004).

Despite selection on, and ample additive genetic variation in, the elevation of the reaction norm (Tables 1 and S1), the observed, period-specific reaction norms changed only marginally over the course of time, as was also predicted from the quantitative genetic model (eq. (5b)). A close look at Figure 1C reveals that the population is “sliding” up and down the same reaction norm; that is, the observed temperature range has become wider and individuals are still using these E temperatures to time the onset of reproduction (see also Visser et al. 2006). Clearly, given the increase in mismatch with the caterpillar peak in this population, this is not an adaptive strategy (Reed et al. 2013b; Thomas et al. 2001); the population needs to evolve a lower elevation to become better matched with the caterpillar peak. The apparent lack of such a shift is most likely explained by the low heritability of elevation, the generation time and the sex limitation of laying date, as shown by our quantitative genetic model (eq. (5b)). Additionally, increased variability in phenotypic optima in the last third of the study period may further hamper adaptation as early-breeding genotypes no longer consistently have a reproductive advantage.

In the quantitative genetic model, we used annual phenotypic selection gradients to predict the change in the reaction norm from one year to the next. One concern with using phenotypic (annual) selection gradients is that estimates of selection in reality reflect an environmental covariance between the trait and fitness or selection on a correlated trait (Lande and Arnold 1983; Price et al. 1988; Hadfield 2008). In our population, however, we know that such environmental bias in selection is limited and that estimates of selection at the phenotypic and genetic level are very similar (Gienapp et al. 2006; Reed et al. 2016). Thus, the necessary ingredients for genetic adaptation—genetic variation and selection—are real and present in HV, but evolutionary change is simply too small to be detected due to the large environmental variation in laying date among years (see also discussion in Gienapp et al. 2006). Such small rates of adaptation can put a strain on population persistence in the longer run as adaptation continues to be outpaced by climate change (Gienapp et al. 2013a; Carlson et al. 2014; Visser and Gienapp, unpubl. ms.; but see Reed et al. 2013a,b).

Many populations have thus far responded to the increased mismatch with the phenology of their food through phenotypic plasticity (Charmantier and Gienapp 2014; Merilä and Hendry 2014), but it has been predicted that this cannot be sufficient in the long run as climate change continues to disrupt synchrony

between trophic levels (Visser et al. 2004; Visser 2008; Thackeray et al. 2010; Carlson et al. 2014; Thackeray et al. 2016; Visser and Gienapp, unpubl. ms.). Consequently, we need to know whether an absence of $G \times E$ in phenology as reported here is the general case or an exception. As pointed out above, such studies in natural populations are rare (but see Ramakers et al. 2018), which is likely due to the logistical challenges of obtaining the necessary pedigree, so far constraining these studies to mainly birds and mammals. Replacing the observational, social pedigree by relatedness estimated from genetic markers has earlier been suggested (Moore and Kukul 2002; Ritland 1996), but applications of this approach remained unsuccessful (Coltman 2005; Garant and Kruuk 2005; Csilléry et al. 2006). However, with the advent of high-throughput, high-density genotyping to “nonmodel” species, “genomic” relatedness estimates should have become sufficiently reliable to replace pedigrees (Gienapp et al. 2017) and thereby allow us to study genetic variation in and selection on phenotypic plasticity both in labile and fixed traits in a broader range of taxa.

To date, evidence for successful evolutionary rescue through a genetic shift in the reaction norm remains rare (Merilä and Hendry 2014). One textbook example of successful evolutionary rescue is that of the great tit’s most important food source, caterpillars of the winter moth (*O. brumata*); three Dutch populations have now restored the match of their hatching date with oak (*Quercus robur*) bud burst through a genetic shift in the elevation of the reaction norm (Van Asch et al. 2013). It remains unclear whether such a change will be also observable in vertebrate populations, as the methods thus far deployed have been largely insufficient to infer evolutionary change (Merilä and Hendry 2014). We provide an important step to this discussion by using rigorous statistical tools to reveal the evolutionary potential of a key life-history trait in a reaction norm context.

AUTHOR CONTRIBUTIONS

J.J.C.R., P.G. and M.E.V. designed the study. J.J.C.R. and P.G. designed the analysis protocol. J.J.C.R. performed the analysis and drafted the manuscript. P.G. and M.E.V. commented on the manuscript.

ACKNOWLEDGMENTS

We are grateful to the board of the Hoge Veluwe National Park for allowing us to do fieldwork all these years, to the countless number of students, assistants, and volunteers who collected the data, and to Jan Visser and Louis Vernooij for carefully managing the database. We are grateful to Thomas E. Reed for the inspiring discussions that eventually led to this manuscript. Sam Scheiner, Michael Morrissey, and an anonymous referee provided useful comments on a previous version of the manuscript. We thank Jarrod Hadfield for his help with the model specifications. Funding for this project was provided by an ERC advanced grant (339092-E-Response) to M.E.V.

DATA ARCHIVING

The doi for our data is <https://doi.org/10.5061/dryad.35k1n3m>.

LITERATURE CITED

- Avilés, J. M., M. Molina-Morales, and J. Gabriel Martínez. 2014. Climatic effects and phenological mismatch in cuckoo-host interactions: a role for host phenotypic plasticity in laying date? *Oikos* 123:993–1002.
- Bailey, L. D., and M. Van de Pol. 2017. Package “climwin”: climate window analysis. <https://cran.r-project.org/web/packages/climwin/index.html>.
- Blondel, J., P. Perret, and M. Maistre. 1990. On the genetical basis of the laying-date in an island population of blue tits. *J. Evol. Biol.* 3:469–475.
- Both, C., S. Bouwhuis, C. M. Lessells, and M. E. Visser. 2006. Climate change and population declines in a long-distance migratory bird. *Nature* 441:81–83.
- Both, C., J. M. Tinbergen, and M. E. Visser. 2000. Adaptive density dependence of avian clutch size. *Ecology* 81:3391–3403.
- Both, C., M. van Asch, R. G. Bijlsma, A. B. van den Burg, and M. E. Visser. 2009. Climate change and unequal phenological changes across four trophic levels: constraints or adaptations? *J. Anim. Ecol.* 78:73–83.
- Brommer, J. E., J. S. Alho, C. Biard, J. R. Chapman, A. Charmantier, A. Dreiss, I. R. Hartley, M. B. Hjernerquist, B. Kempenaers, et al. 2010. Passerine extrapair mating dynamics: a Bayesian modeling approach comparing four species. *Am. Nat.* 176:178–187.
- Brommer, J. E., P. Kontiainen, and H. Pietiäinen. 2012. Selection on plasticity of seasonal life-history traits using random regression mixed model analysis. *Ecol. Evol.* 2:695–704.
- Brommer, J. E., J. Merilä, B. C. Sheldon, and L. Gustafsson. 2005. Natural selection and genetic variation for reproductive reaction norms in a wild bird population. *Evolution* 59:1362–1371.
- Carlson, S. M., C. J. Cunningham, and P. A. H. Westley. 2014. Evolutionary rescue in a changing world. *Trends Ecol. Evol.* 29:521–530.
- Caro, S. P., A. Charmantier, M. M. Lambrechts, J. Blondel, J. Balthazart, and T. D. Williams. 2009. Local adaptation of timing of reproduction: females are in the driver’s seat. *Funct. Ecol.* 23:172–179.
- Carter, M. J., M. I. Lind, S. R. Dennis, W. Hentley, and A. P. Beckerman. 2017. Evolution of a predator-induced, nonlinear reaction norm. *Proc. R Soc. B* 284:20170859.
- Charmantier, A., and P. Gienapp. 2014. Climate change and timing of avian breeding and migration: evolutionary versus plastic changes. *Evol. Appl.* 7:15–28.
- Charmantier, A., R. H. McCleery, L. R. Cole, C. M. Perrins, L. E. B. Kruuk, and B. C. Sheldon. 2008. Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science* 320:800–803.
- Charmantier, A., and D. Réale. 2005. How do misassigned paternities affect the estimation of heritability in the wild? *Mol. Ecol.* 14:2839–2850.
- Chevin, L. M., M. E. Visser, and J. Tufto. 2015. Estimating the variation, autocorrelation, and environmental sensitivity of phenotypic selection. *Evolution* 69:2319–2332.
- Coltman, D. W. 2005. Testing marker-based estimates of heritability in the wild. *Mol. Ecol.* 14:2593–2599.
- Csilléry, K., T. Johnson, D. Beraldi, T. Clutton-Brock, D. Coltman, B. Hansson, G. Spong, and J. M. Pemberton. 2006. Performance of marker-based relatedness estimators in natural populations of outbred vertebrates. *Genetics* 173:2091–2101.
- De Jong, G. 1999. Unpredictable selection in a structured population leads to local genetic differentiation in evolved reaction norms. *J. Evol. Biol.* 12:839–851.
- Dingemanse, N. J., A. J. Kazem, D. Réale, and J. Wright. 2010. Behavioural reaction norms: animal personality meets individual plasticity. *Trends Ecol. Evol.* 25:81–89.
- Durant, J. M., D. Ø. Hjernmann, G. Ottersen, and N. C. Stenseth. 2007. Climate and the match or mismatch between predator requirements and resource availability. *Climate Res.* 33:271–283.

- Falconer, D. S., and T. F. C. Mackay. 1996. Introduction to quantitative genetics. Longmans Green, Harlow, Essex, UK.
- Firth, J. A., J. D. Hadfield, A. W. Santure, J. Slate, and B. C. Sheldon. 2015. The influence of nonrandom extra-pair paternity on heritability estimates derived from wild pedigrees. *Evolution* 69:1336–1344.
- Garant, D., and L. E. Kruuk. 2005. How to use molecular marker data to measure evolutionary parameters in wild populations. *Mol. Ecol.* 14:1843–1859.
- Garant, D., L. E. Kruuk, R. H. McCleery, and B. C. Sheldon. 2004. Evolution in a changing environment: a case study with great tit fledging mass. *Am. Nat.* 164:E115–E129.
- Gienapp, P., and J. E. Brommer. 2014. Evolutionary dynamics in response to climate change. in A. Charmantier, D. Garant, and L. E. B. Kruuk, eds. Quantitative genetics in the wild. Oxford Univ. Press, Oxford, UK.
- Gienapp, P., S. Fior, F. Guillaume, J. R. Lasky, V. L. Sork, and K. Csilléry. 2017. Genomic quantitative genetics to study evolution in the wild. *Trends Ecol. Evol.* 32:897–908.
- Gienapp, P., L. Hemerik, and M. E. Visser. 2005. A new statistical tool to predict phenology under climate change scenarios. *Global Change Biol.* 11:600–606.
- Gienapp, P., M. Lof, T. E. Reed, J. McNamara, S. Verhulst, and M. E. Visser. 2013a. Predicting demographically sustainable rates of adaptation: can great tit breeding time keep pace with climate change? *Philos. Trans. R Soc. B* 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.
- 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.
- Gienapp, P., A. J. van Noordwijk, and M. E. Visser. 2013b. Genetic background, and not ontogenetic effects, affects avian seasonal timing of reproduction. *J. Evol. Biol.* 26:2147–2153.
- Hadfield, J. 2008. Estimating evolutionary parameters when viability selection is operating. *Proc. R Soc. B Biol. Sci.* 275:723–734.
- Hadfield, J. 2014. MCMCglimm Course Notes. <https://cran.r-project.org/web/packages/MCMCglimm/. . . /CourseNotes.pdf>.
- Hadfield, J. D. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglimm R package. *J. Stat. Softw.* 33:1–22.
- Hadfield, J. D., A. J. Wilson, D. Garant, B. C. Sheldon, and L. E. B. Kruuk. 2010. The misuse of BLUP in ecology and evolution. *Am. Nat.* 175:116–125.
- Hayward, A. D., D. H. Nussey, A. J. Wilson, C. Berenos, J. G. Pilkington, K. A. Watt, J. M. Pemberton, and A. L. Graham. 2014. Natural selection on individual variation in tolerance of gastrointestinal nematode infection. *PLoS Biol.* 12:e1001917.
- Hayward, A. D., J. M. Pemberton, C. Berenos, A. J. Wilson, J. G. Pilkington, and L. E. Kruuk. 2018. Evidence for selection-by-environment but not genotype-by-environment interactions for fitness-related traits in a wild mammal population. *Genetics* 208:349–364.
- Henderson, C. R. 1988. Theoretical basis and computational methods for a number of different animal models. *J. Dairy Sci.* 71:1–16.
- Hoffman, A. A., and J. Merilä. 1999. Heritable variation and evolution under favourable and unfavourable conditions. *Trends Ecol. Evol.* 14:96–101.
- Husby, A., D. H. Nussey, M. E. Visser, A. J. Wilson, B. C. Sheldon, and L. E. B. Kruuk. 2010. Contrasting patterns of phenotypic plasticity in reproductive traits in two great tit (*Parus major*) populations. *Evolution* 64:2221–2237.
- Husby, A., M. E. Visser, and L. E. B. Kruuk. 2011. Speeding up microevolution: the effects of increasing temperature on selection and genetic variance in a wild bird population. *PLoS Biol.* 9:e1000585.
- Keller, L. F., and A. J. Van Noordwijk. 1994. Effects of local environmental conditions on nestling growth in the great tit (*Parus major* L.). *Ardea* 82:349–362.
- Kokko, H., and K. Heubel. 2008. Condition-dependence, genotype-by-environment interactions and the lek paradox. *Genetica* 134:55–62.
- Kreft, I. G., J. De Leeuw, and L. S. Aiken. 1995. The effect of different forms of centering in hierarchical linear models. *Multivariate Behav. Res.* 30:1–21.
- Kruuk, L. E. B. 2004. Estimating genetic parameters in natural populations using the “animal model”. *Philos. Trans. R Soc. B* 359:873–890.
- Kvist, L., T. Arbabi, M. Paeckert, M. Orell, and J. Martens. 2007. Population differentiation in the marginal populations of the great tit (*Paridae: Parus major*). *Biol. J. Linn. Soc.* 90:201–210.
- Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution* 33:402–416.
- . 2009. Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. *J. Evol. Biol.* 22:1435–1446.
- Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. *Evolution* 37:1210–1226.
- Lillehammer, M., J. Ødegård, and T. H. E. Meuwissen. 2009. Reducing the bias of estimates of genotype by environment interactions in random regression sire models. *Genet. Selection Evol.* 41:30.
- Martin, J. G., D. H. Nussey, A. J. Wilson, and D. Réale. 2011. Measuring individual differences in reaction norms in field and experimental studies: a power analysis of random regression models. *Methods Ecol. Evol.* 2:362–374.
- Merilä, J., and A. P. Hendry. 2014. Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. *Evol. Appl.* 7:1–14.
- Mols, C. M. M., A. J. Van Noordwijk, and M. E. Visser. 2005. Assessing the reduction of caterpillar numbers by great tits *Parus major* breeding in apple orchards. *Ardea* 93:259–269.
- Moore, A. J., and P. F. Kukuk. 2002. Quantitative genetic analysis of natural populations. *Nat. Rev. Genet.* 3:971.
- Morrissey, M. B., and M. Liefiting. 2016. Variation in reaction norms: statistical considerations and biological interpretation. *Evolution* 70:1944–1959.
- Morrissey, M. B., D. J. Parker, P. Korsten, J. M. Pemberton, L. E. B. Kruuk, and A. J. Wilson. 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.
- Morrissey, M. B., and K. Sakrejsda. 2015. Package “gsg”: calculation of selection coefficients. <https://cran.r-project.org/web/packages/gsg/gsg.pdf>.
- Murren, C. J., H. J. Maclean, S. E. Diamond, U. K. Steiner, M. A. Heskell, C. A. Handelsman, C. K. Ghalambor, J. R. Auld, H. S. Callahan, and D. W. Pfennig. 2014. Evolutionary change in continuous reaction norms. *Am. Nat.* 183:453–467.
- Nicolaus, M., J. E. Brommer, R. Ubels, J. M. Tinbergen, and N. J. Dingemanse. 2013. Exploring patterns of variation in clutch size–density reaction norms in a wild passerine bird. *J. Evol. Biol.* 26:2031–2043.
- Nussey, D. H., T. H. Clutton-Brock, D. A. Elston, S. D. Albon, and L. E. Kruuk. 2005a. Phenotypic plasticity in a maternal trait in red deer. *J. Anim. Ecol.* 74:387–396.
- Nussey, D. H., E. Postma, P. Gienapp, and M. E. Visser. 2005b. Selection on heritable phenotypic plasticity in a wild bird population. *Science* 310:304–306.
- Nussey, D. H., A. J. Wilson, and J. E. Brommer. 2007. The evolutionary ecology of individual phenotypic plasticity in wild populations. *J. Evol. Biol.* 20:831–844.

- Pigliucci, M. 2001. Phenotypic plasticity; beyond nature and nurture. Synthesis in ecology and evolution. John Hopkins Univ. Press, Baltimore, MD.
- Postma, E., and A. J. Van Noordwijk. 2005. Gene flow maintains a large genetic difference in clutch size at a small spatial scale. *Nature* 433:65.
- Price, T., M. Kirkpatrick, and S. J. Arnold. 1988. Directional selection and the evolution of breeding date in birds. *Science* 240:798–799.
- Ramakers, J. J. C., A. Culina, M. E. Visser, and P. Gienapp. 2018. Environmental coupling of heritability and selection is rare and of minor evolutionary significance in wild populations. *Nat. Ecol. Evol.* 2:1093–1103.
- Réale, D., A. G. McAdam, S. Boutin, and D. Berteaux. 2003. Genetic and plastic responses of a northern mammal to climate change. *Proc. R Soc. Lond. B Biol. Sci.* 270:591–596.
- Reed, T. E., P. Gienapp, and M. E. Visser. 2016. Testing for biases in selection on avian reproductive traits and partitioning direct and indirect selection using quantitative genetic models. *Evolution* 70:2211–2225.
- Reed, T. E., V. Grotan, S. Jenouvrier, B.-E. Saether, and M. E. Visser. 2013a. Population growth in a wild bird is buffered against phenological mismatch. *Science* 340:488–491.
- Reed, T. E., S. Jenouvrier, and M. E. Visser. 2013b. Phenological mismatch strongly affects individual fitness but not population demography in a woodland passerine. *J. Anim. Ecol.* 82:131–144.
- Reed, T. E., R. S. Waples, D. E. Schindler, J. J. Hard, and M. T. Kinnison. 2010. Phenotypic plasticity and population viability: the importance of environmental predictability. *Proc. R. Soc. B* 277:3391–3400.
- Ricklefs, R. E. 1980. Geographical variation in clutch size among passerine birds: Ashmole's hypothesis. *The Auk* 97:38–49.
- Ritland, K. 1996. A marker-based method for inferences about quantitative inheritance in natural populations. *Evolution* 50:1062–1073.
- Sæther, B.-E., M. E. Visser, V. Grøtan, and S. Engen. 2016. Demonstrating r- and K-selection in a wild bird population. *Proc. R Soc. B Biol. Sci.* 283:20152411.
- Schaper, S. V., A. Dawson, P. J. Sharp, P. Gienapp, S. P. Caro, and M. E. Visser. 2012. Increasing temperature, not mean temperature, is a cue for avian timing of reproduction. *Am. Nat.* 179:E55–E69.
- Scheiner, S. M. 1993. Genetics and evolution of phenotypic plasticity. *Ann. Rev. Ecol. Syst.* 24:35–68.
- Schlichting, C. D., and M. Pigliucci. 1998. Phenotypic evolution: a reaction norm perspective. Sinauer Associates, Sunderland, MA.
- Schultz, M. T., J. F. Piatt, A. M. A. Harding, A. B. Kettle, and T. I. Van Pelt. 2009. Timing of breeding and reproductive performance in mirres and kittiwakes reflect mismatched seasonal prey dynamics. *Mar. Ecol. Prog. Ser.* 393:247–258.
- Sheldon, B. C., L. E. B. Kruuk, and J. Merilä. 2003. Natural selection and inheritance of breeding time and clutch size in the collared flycatchers. *Evolution* 57:406–420.
- Taylor, J. 1997. Introduction to error analysis, the study of uncertainties in physical measurements. University Science Books, New York.
- Thackeray, S. J., P. A. Henrys, D. Hemming, J. R. Bell, M. S. Botham, S. Burthe, P. Helaouet, D. G. Johns, I. D. Jones, and D. I. Leech. 2016. Phenological sensitivity to climate across taxa and trophic levels. *Nature* 535:241.
- Thackeray, S. J., T. H. Sparks, M. Frederiksen, S. Burthe, P. J. Bacon, J. R. Bell, M. S. Botham, T. M. Brereton, P. W. Bright, L. Carvalho et al. 2010. Trophic level asynchrony in rates of phenological change for marine, freshwater and terrestrial environments. *Global Change Biol.* 16:3304–3313.
- Thomas, D. W., J. Blondel, P. Perret, M. M. Lambrechts, and J. R. Speakman. 2001. Energetic and fitness costs of mismatching resource supply and demand in seasonally breeding birds. *Science* 291:2598–2600.
- Tomkins, J. L., J. Radwan, J. S. Kotiaho, and T. Tregenza. 2004. Genic capture and resolving the lek paradox. *Trends Ecol. Evol.* 19:323–328.
- Turelli, M., and N. H. Barton. 2004. Polygenic variation maintained by balancing selection: pleiotropy, sex-dependent allelic effects and g x e interactions. *Genetics* 166:1053–1079.
- Van Asch, M., L. Salis, L. J. M. Holleman, B. van Lith, and M. E. Visser. 2013. Evolutionary response of the egg hatching date of a herbivorous insect under climate change. *Nat. Clim. Change* 3:244–248.
- Van Asch, M., P. H. Van Tienderen, L. J. M. Holleman, and M. E. Visser. 2007. Predicting adaptation of phenology in response to climate change, an insect herbivore example. *Global Change Biol.* 13:1596–1604.
- Van de Pol, M. 2012. Quantifying individual variation in reaction norms: how study design affects the accuracy, precision and power of random regression models. *Methods Ecol. Evol.* 3:268–280.
- Van de Pol, M., and J. Wright. 2009. A simple method for distinguishing within-versus between-subject effects using mixed models. *Anim. Behav.* 77:753–758.
- Van Noordwijk, A. J., R. H. McCleery, and C. M. Perrins. 1995. Selection for the timing of great tit breeding in relation to caterpillar growth and temperature. *J. Anim. Ecol.* 64:451–458.
- Van Noordwijk, A. J., J. H. van Balen, and W. Scharloo. 1981. Genetic variation in the timing of reproduction in the great tit. *Oecologia* 49:158–166.
- Van Oers, K., P. J. Drent, N. J. Dingemans, and B. Kempenaers. 2008. Personality is associated with extrapair paternity in great tits, *Parus major*. *Anim. Behav.* 76:555–563.
- Van Tienderen, P. H., and H. P. Koelewijn. 1994. Selection on reaction norms, genetic correlations and constraints. *Genet. Res.* 64:115–125.
- Verhulst, S., and J.-Å. Nilsson. 2008. The timing of birds' breeding seasons: a review of experiments that manipulated timing of breeding. *Philos. Trans. R Soc. B* 363:399–410.
- Via, S., R. Gomulkiewicz, G. De Jong, S. M. Scheiner, C. D. Schlichting, and P. H. Van Tienderen. 1995. Adaptive phenotypic plasticity: consensus and controversy. *Trends Ecol. Evol.* 10:212–217.
- 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.
- Visser, M. E., and C. Both. 2005. Shifts in phenology due to global climate change: the need for a yardstick. *Proc. R Soc. B Biol. Sci.* 272:2561–2569.
- Visser, M. E., C. Both, and M. M. Lambrechts. 2004. Global climate change leads to mistimed avian reproduction. *Adv. Ecol. Res.* 35:89–110.
- Visser, M. E. and P. Gienapp. Submitted. Seasonal timing in a warming world: a review of the phenological mismatch hypothesis.
- Visser, M. E., and L. J. M. Holleman. 2001. Warmer springs disrupt the synchrony of oak and winter moth phenology. *Proc. R Soc. Lond. B* 268:289–294.
- Visser, M. E., L. J. M. Holleman, and S. P. Caro. 2009. Temperature has a causal effect on avian timing of reproduction. *Proc. R Soc. Lond. B* 276:2323–2331.
- Visser, M. E., L. J. M. Holleman, and P. Gienapp. 2006. Shifts in caterpillar biomass phenology due to climate change and its impact on the breeding biology of an insectivorous bird. *Oecologia* 147:164–172.
- Visser, M. E., B. Silverin, M. M. Lambrechts, and J. M. Tinbergen. 2002. No evidence for tree phenology as a cue for the timing of reproduction in tits *Parus* spp. *Avian Sci.* 2:77–86.
- 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. B Biol. Sci.* 265:1867–1870.
- Weis, A. E., and W. L. Gorman. 1990. Measuring selection on reaction norms: an exploration of the Eurosta-Solidago system. *Evolution* 44:820–831.

- Woltereck, R. 1909. Weitere experimentelle Untersuchungen über Artveränderung, speziell über das Wesen quantitativer Artunterschiede bei Daphniden. Verhandlungen der Deutschen Zoologischen Gesellschaft 19:110–173.
- Wood, C. W., and E. D. Brodie III. 2016. Evolutionary response when selection and genetic variation covary across environments. *Ecol. Lett.* 19:1189–1200.
- Wood, S. 2017. Package “mgcv”: Mixed GAM Computation Vehicle with Automatic Smoothness Automation. <https://cran.r-project.org/web/packages/mgcv/mgcv.pdf>.

Associate Editor: M. Morrissey
Handling Editor: P. Tiffin

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Annual selection gradients (β_Z), absolute (T) and mean-centred (T_c) mean spring temperature, and the predicted annual evolutionary changes (Δ) in the reaction norm elevation (intercept, Int) and slope (Slp), along with standard errors (SE), as a result of annual selection on great tit laying date from the second time period onward (i.e. excluding 1973–1987 because the ‘initial’ reaction norm was estimated over these first 15 years).

Figure S1. Results of the simulation testing the effect of the residual structure of the random regression model on the estimate of variance in slopes (a, c, e) and the respective statistical power to detect this variance (b, d, f).

Supporting Information