EVOLUTIONARY BIOLOGY

Genotypes selected for early and late avian lay date differ in their phenotype, but not fitness, in the wild

Melanie Lindner[1,2]*, Jip JC Ramakers[3], Irene Verhagen[4], Barbara M Tomotani[1,5], A Christa Mateman[1], Phillip Gienapp[6], Marcel E Visser[1,2]*

Global warming has shifted phenological traits in many species, but whether species are able to track further increasing temperatures depends on the fitness consequences of additional shifts in phenological traits. To test this, we measured phenology and fitness of great tits (*Parus major*) with genotypes for extremely early and late egg lay dates, obtained from a genomic selection experiment. Females with early genotypes advanced lay dates relative to females with late genotypes, but not relative to nonselected females. Females with early and late genotypes did not differ in the number of fledglings produced, in line with the weak effect of lay date on the number of fledglings produced by nonselected females in the years of the experiment. Our study is the first application of genomic selection in the wild and led to an asymmetric phenotypic response that indicates the presence of constraints toward early, but not late, lay dates.

INTRODUCTION

Anthropogenic global warming has led to drastic changes in the environment and species must cope with these changes. Many species use ambient temperature as a cue to match the timing of seasonally expressed life-history traits to the year-specific environment (*1*). For many seasonally reproducing consumer species, ambient temperature predicts when food resources are plentiful. By using temperature as a cue, consumer species can match the timing of offspring provisioning with the timing of maximum food resource abundance. In many systems, however, consumer and resource species differ in how strongly they respond to increasing temperatures (*2*, *3*). They might differ in temperature sensitivity per se (i.e., the number of days phenology shifts per 1°C in temperature change) and/or in the time period at which they are temperature sensitive, and the temperature in these time periods may change at different rates (*4*). This way, increasing temperatures can uncouple the timing of offspring provisioning from the timing of maximum food resource abundance, leading to a phenological mismatch between consumer and resource species with potentially severe consequences on reproductive success of the consumer species (*5*). The poster child example for this in evolutionary ecology is the phenological mismatch between great tits (*Parus major*) and their caterpillar prey (*6*–*8*). In this system, the match between the food demands of great tit young and caterpillar biomass abundance is a strong selection pressure on first egg laying date (hereafter lay date). Hence, an advancement in the phenology of caterpillar biomass without a concomitant advancement in lay dates and thus chick rearing period is expected to negatively affect the reproductive success of great tits (*6*) and population stability in the long run (*7*, *9*).

Quantitative genetic studies have long used data from individual-based long-term study populations to estimate the strength of directional selection and the magnitude of additive genetic variation underlying focal traits to predict the potential for a microevolutionary response to selection (*10*, *11*). Despite many examples of directional selection [e.g., (*11*–*15*)], microevolutionary responses to selection in the wild that match the expectations based on estimates of quantitative genetic studies are rare (*16*) and often populations are in so-called evolutionary stasis (*16*). There are potential statistical and biological explanations for why this might be the case (*17*), and one of them is that we might not know whether the predicted response would indeed lead to an increased fitness if predicted responses lie outside the currently observed distribution. In other words, we have no data to estimate the fitness consequences of these shifted trait values and cannot assess whether the predicted response to directional selection would have positive or negative consequences for fitness.

For great tit populations, which are phenologically mismatched with their main food resource for chick feeding, a directional mean shift in phenotypes does not necessarily result in higher fitness (*5*). In one scenario, early lay dates that reduce the phenological mismatch between offspring provisioning and maximum food abundance are expected to have higher reproductive success when the phenological mismatch is the main selection pressure, i.e., fitness is maximized with a reduced mismatch [figure 1 in (*5*)]. Then, a phenological mismatch is considered true mistiming between consumer and resource species. In an alternative second scenario, other ecological variables, in addition to the match between consumer and resource phenology, are important determinants of fitness (*5*). For example, poor environmental conditions in early spring, such as cold temperatures and limited food resources, might lead to high fitness costs (e.g., increased mortality) for females that attempt to breed early enough to reduce the phenological mismatch. With these additional fitness components in place, the phenotype that maximizes fitness might maintain (a certain degree of) the phenological mismatch [figure 2 in (*5*)]. Then, the phenological

[1]Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, Netherlands. [2]Chronobiology Unit, Groningen Institute for Evolutionary Life Sciences (GELIFES), University of Groningen, Groningen, Netherlands. [3]Mathematical and Statistical Methods–Biometris, Wageningen University & Research (WUR), Wageningen, Netherlands. [4]Wageningen University & Research (WUR) Library, Wageningen, Netherlands. [5]Department of Arctic and Marine Biology, UiT The Arctic University of Norway, Tromsø, Norway. [6]Michael-Otto-Institut, NABU, Bergenhusen, Germany.
*Corresponding author. Email: m.lindner@nioo.knaw.nl (M.L.); m.visser@nioo.knaw.nl (M.E.V.)

mismatch is considered an adaptive mismatch between consumer and resource species. Only when earlier phenotypes have increased fitness (i.e., the phenological mismatch constitutes true mistiming; first scenario), do we expect to see a shift in the distribution of trait values as a response to directional selection [figure 4 in (*17*)]. We, however, lack sufficient data at the very ends of the current lay date distribution to test whether a phenological mismatch constitutes mistiming (if earlier females with a reduced phenological mismatch had higher fitness) or an adaptive mismatch (if earlier females had lower fitness than females that retain a certain degree of the phenological mismatch).

Understanding the evolutionary consequences of a phenological mismatch requires to move phenotypes toward earlier lay dates in the wild and assess the fitness consequences. To directionally push phenotypes toward earlier or later lay dates, some kind of experimental manipulation is needed. It has, however, proven challenging to perform "clean" manipulations of avian breeding time in the wild (*18*). Methods that previously attempted to manipulate lay dates constitute manipulations on the phenotypic level, such as the exposure to a single long day (*19*), manipulation of photoperiod perception (*20*), a leptin implant treatment (*21*), or the use of supplemental food in the time before laying (*22*). These manipulations either failed to advance lay dates or are known to induce a bias in reproductive success as a consequence of the manipulation. Hence, these methods are unsuitable for estimating the fitness components of advanced lay dates, but see (*23*) for a successful population-specific phenotypic manipulation of lay dates.

Genomic selection offers a previously unexplored approach to induce directional shifts in lay dates of seasonally breeding birds that have genomic breeding values (GEBVs) for extremely early and late lay dates (*24*). In contrast to traditionally applied procedures for artificial selection, where one selects on the expressed phenotype or breeding values (BVs) estimated using pedigree-based relatedness, genomic selection is based on GEBVs, the additive effect of an individual's genotype on the phenotype relative to the population mean phenotype estimated using genomic approaches (*25*). Calculating GEBVs for a wild bird species allowed us to follow females, which are expected to have extremely early or late lay dates based on their GEBVs, throughout their lifetime. This way, the fitness consequences of individuals that are expected to have extremely early and late lay dates can be assessed while differentiating between (i) the fitness costs in the time from fledging to first-time recruitment and first egg lay date and (ii), if a female succeeds to recruit, the fitness benefits in the form of lifetime reproductive success. While genomic selection via GEBVs is a powerful tool for artificial selection commonly applied in animal and plant breeding (*26*) and its potential for studies in the wild has been acknowledged (*25*), we are currently not aware of any study in which genomic selection has been used in this way. Note, however, that GEBVs have previously been estimated in a wild population (*27*). This lack of application is likely explained by limitations in sample size and statistical power, as a large number of individuals genotyped at a sufficient marker density is required. These limitations are especially relevant for phenological traits that generally have low heritability [~0.2 for lay dates in our study population (*24*)] and might require a training population with a sample size of thousands.

Here, a training population of >2000 great tit females from a long-term study population at the Hoge Veluwe National Park (The Netherlands) with known lay dates and genotyped at >500,000 single-nucleotide polymorphisms (SNPs) was used to estimate GEBVs using the "genomic best linear unbiased prediction" (GBLUP) approach (*24*, *28*). GEBVs were used as the criterion for the selection procedure in such a way that birds with extremely negative and positive GEBVs were selected for the early and late selection lines for lay date. A detailed description of the selection procedure can be found in (*29*), and a short description can be found in Materials and Methods. We bred F3 generation breeding pairs of the early and late selection line from 2017 to 2019 and moved the F4 generation eggs they laid into the Hoge Veluwe study population. In the following year(s), we identified and monitored any selection line females of the F4 generation that recruited into the study population as breeding bird to record the realized lay dates and assess the fitness consequences.

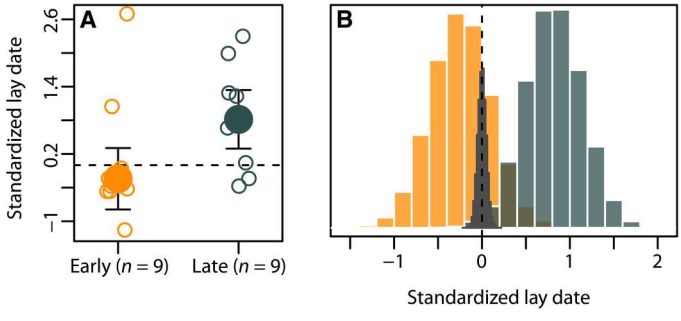

Fig. 1. Posterior predictions of standardized lay dates. (**A**) Posterior means of mean-standardized lay dates with 89% credible interval for female recruits form the early (yellow, $n = 8$) and late (blue, $n = 9$) selection line (posterior mean difference in standardized lay dates (late-early): 1.06; 89% credible interval: 0.30;1.81). The crossed-out data point indicates an outlier female that was removed from the data before analysis (see text S5 for a formal outlier analysis). (**B**) Posterior distribution of mean-standardized lay dates for female recruits from the early (yellow) and late (blue) selection line and for local female recruits (dark gray, $n = 433$) [posterior mean difference in standardized lay dates (local-early): 0.24, 89% credible interval: −0.31;0.80; posterior mean difference in standardized lay dates (local-late): −0.82, 89% credible interval: −1.34;−0.29]. The posterior mean of standardized lay dates for local female recruits is shown as vertical dashed line. Lay date observations were standardized as z scores using a year-specific SD of 3.26, 5.45, and 4.82 days for 2018, 2019, and 2020, respectively.

RESULTS

Phenotypic response to genomic selection in the wild

Lay date

We recorded the lay dates of female selection line recruits at the local study population in the years 2018 to 2020 to test whether genomic selection for early and late lay dates translated into a phenotypic response in the wild. In total, 936 F4 selection line eggs produced by F3 selection line breeding pairs in aviaries were introduced into the wild study population at the Hoge Veluwe of which <20 females locally recruited as breeding birds. Female recruits from the early selection line ($n = 8$) had earlier lay dates than female recruits from the late selection line ($n = 9$) (posterior mean difference in standardized lay dates (late-early): 1.06, 89% credible interval: 0.30;1.81, Fig. 1A and tables S1 and S2) when a potential outlier observation was removed (see text S5 for a formal outlier analysis and fig. S11). The finding did not change when the outlier remained

included (fig. S12). The posterior mean difference in standardized lay dates translated into a mean difference in lay dates of 3.45 (2018), 5.77 (2019), and 5.10 (2020) days. As lay dates were standardized, the posterior distribution of standardized lay dates for nonselection line female recruits (hereafter local female recruits; n = 433) was centered at zero (posterior mean of standardized lay dates: 0.00, 89% credible interval: −0.08;0.08; Fig. 1B and tables S1 and S2). While we did not succeed to advance the lay dates of female recruits from the early selection line relative to the lay dates of local female recruits [posterior mean difference in standardized lay dates (local-early): 0.24, 89% credible interval: −0.31;0.80], we succeeded to delay the lay dates of female recruits from the late selection line relative to lay dates of local female recruits [posterior mean difference in standardized lay dates (local-late): −0.82, 89% credible interval: −1.34;-0.29, Fig. 1B and tables S1 and S2]. This interpretation is also supported by the posterior distribution of standardized lay dates for female recruits from the early and late selection line (posterior mean of standardized lay dates for female recruits from the early selection line: −0.24, 89% credible interval: −0.79;0.31; posterior mean of standardized lay dates for female recruits from the late selection line: 0.82, 89% credible interval: 0.30;1.34; Fig. 1A and tables S1 and S2). Hence, the genomic selection experiment for early and late lay dates led to a somewhat asymmetric phenotypic response in the wild (Fig. 1B). However, there is no difference between the selection lines in their phenotypic response relative to local female recruits (posterior mean of difference in standardized lay dates relative to local females between selection lines: 0.57, 89% credible interval: −0.20;1.35; tables S1 and S2).

Phenological mismatch

We assessed whether the difference in lay date between female recruits from the early and late selection line (Fig. 1) translated into a difference in phenological mismatch, i.e., the difference in days between the expected chick rearing period and the period of maximum caterpillar biomass availability. Female recruits from the early selection line (n = 8) showed a reduced phenological mismatch relative to female recruits from the late selection line (n = 9) [posterior mean difference in standardized phenological mismatch (late-early): 1.90, 89% credible interval: 0.44;3.38; Fig. 2 and tables S3 and S4]. The phenological mismatch of female recruits from either selection line did not differ from the phenological mismatch of local female recruits, with a posterior mean difference in standardized phenological mismatch of 0.93 (local-early) (89% credible interval: −0.17;2.02) and −0.97 (local-late) (89% credible interval: −2.02;0.07; Fig. 2 and tables S3 and S4) for female recruits from the early and late selection line, respectively. For local female recruits (n = 433), the number of days of the phenological mismatch differed between years (fig. S1 and tables S5 and S6). While females were on average mismatched by 13.38 days in 2018 (89% credible interval: 12.70;14.05) and 9.74 days in 2020 (89% credible interval: 9.17;10.31), they were, on average, better matched in 2019 with a negative mismatch of 2.62 days (89% credible interval: −3.22;-2.02; tables S5 and S6), i.e., they laid, on average, 2.62 days too early rather than too late.

Consequences of genomic selection for fitness in the wild

To test for fitness differences between lines we analyzed the first-year survival of selection line fledglings [in the form of (i) mortality before laying and (ii) probability to locally recruit as a breeding bird and (iii) lifetime number of fledglings produced when recruited]. We then estimated the (iv) total fitness of selection line fledglings as the product of their local recruitment probability and the lifetime number of fledglings produced.

Consequences of genomic selection on mortality before laying

One proxy for survival is the potentially selective disappearance of early selection line females before their first breeding event. This selective disappearance might arise as a consequence of higher mortality risk for females that attempt to breed early, when the environment makes it challenging to produce eggs at that time [i.e., second scenario (*5*)]. For this, we tested whether early selection line females were less likely than late selection line females to recruit into the local study population as a breeding bird when females were identified in late winter [n = 7 (early) and n = 15 (late)] or during nest building in early spring [n = 13 (early) and n = 16 (late)]. We did not find any difference between early and late selection lines females in local recruitment probability irrespectively of whether females were identified in January or early spring (figs. S2 and S3 and tables S7 to S10). This analysis, however, is limited by low statistical power.

Consequences of genomic selection on local recruitment probability

Another proxy for survival is the local recruitment probability [i.e., the probability that a fledgling survives to locally breed the following year(s)], for which we did not find a difference between early (n = 318) and late (n = 331) selection line fledglings [posterior mean difference (late-early): −0.005, 89% credible interval: −0.035;0.026; Fig. 3A and tables S10 to S12] or between selection line fledglings and nonselection line fledglings (hereafter local fledglings; n = 1675) [posterior mean difference (local-early): −0.018, 89% credible interval: −0.044;0.005; posterior mean difference (local late): −0.014, 89% credible interval: −0.038;0.009; tables S10 to S12]. Overall, local recruitment probability of local fledglings was low at the study site and showed a decrease throughout spring and an increase with fledgling weight (Fig. 3B and table S10), in line with the expectation that birds that fledge early and are heavier at the time of fledging are more likely to recruit (*30*).

We moved the F4 selection line eggs in mixed-selection line broods from the aviaries into the Hoge Veluwe study site. With this, we aimed to rear selection line individuals in a common environment in which early and late selection line individuals did not differ in their realized hatch date and subsequent fledging date. We formally tested this assumption and, indeed, did not find a difference in fledging date between early and late selection line fledglings [posterior mean difference in standardized fledging date (late-early): 0.074; 89% credible interval: −0.010;0.158; fig. S4 and tables S13 and S14]. Hence, our experimental design removed any differences between selection lines that, e.g., might arise as a consequence of the maternal lay date (i.e., the lay date of F3 selection line females in the aviaries) and, in turn, might induce a bias in local recruitment probability.

We, furthermore, tested for a difference in the fledgling weight of early and late selection line fledglings, which might be a correlated response to genomic selection for lay date. We found that early selection line fledglings were lighter than late selection line fledglings [posterior mean difference in standardized fledgling weight (late-early): 0.10, 89% credible interval: 0.01;0.18; fig. S5 and tables S15 and S16]. Moreover, early selection line fledglings were lighter than local fledglings [posterior mean difference in standardized fledgling weight (local-early): 0.16; 89% credible interval: 0.09;0.23; tables S15

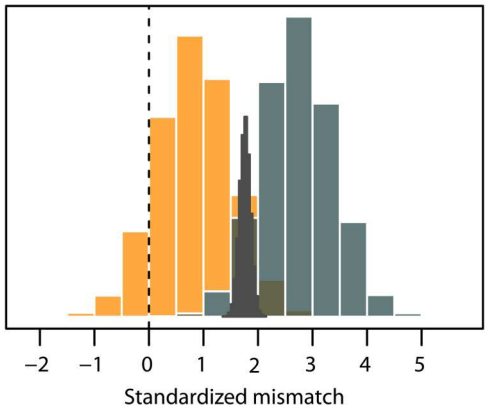

Fig. 2. Posterior predictions of standardized mismatch. Posterior distribution of mean-standardized mismatch for female recruits from the early (yellow, $n = 8$) and late (blue, $n = 9$) selection line and for local female recruits (dark gray, $n = 433$) [posterior mean difference in standardized phenological mismatch (late-early): 1.90, 89% credible interval: 0.44;3.38; posterior mean difference in standardized phenological mismatch (local-early): 0.93, 89% credible interval: −0.17;2.02; posterior mean difference in standardized phenological mismatch (local-late): −0.97, 89% credible interval: −2.02;0.07]. The vertical dashed line indicates the lay date that corresponds to the lay date that results in a perfectly matched phenology. Mismatch was standardized as *z* scores using a year-specific SD of 3.26, 5.45, and 4.82 days for 2018, 2019, and 2020, respectively.

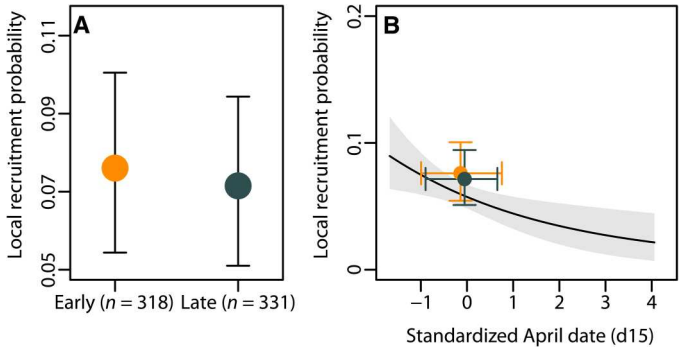

Fig. 3. Posterior predictions of local recruitment probability. (**A**) Posterior means of mean local recruitment probability with 89% credible interval for early (yellow, $n = 318$) and late (blue, $n = 331$) selection line fledglings [posterior mean difference (late-early): −0.005, 89% credible interval: −0.035;0.026]. (**B**) Mean local recruitment probability of early (yellow) and late (blue) selection line fledglings in comparison to local fledglings (black/gray, $n = 1675$). For selection line fledglings, vertical locations of filled circles and error bars correspond to the posterior means of local recruitment probability with 89% credible interval, and horizontal location of filled circles and error bars correspond to mean-standardized April date at d15 with 89% credible interval. For local fledglings, the black line and gray shaded area represent the posterior mean of local recruitment probability over standardized April dates at d15 with 89% credible interval. April dates were standardized as *z* scores using a SD of 10.29 days.

and S16], while there was no difference in fledgling weight between late selection line fledglings and local fledglings [posterior mean difference in standardized fledgling weight (local-late): 0.06, 89% credible interval: −0.00;0.13; tables S15 and S16].

As we can only calculate the consequences of genomic selection on total fitness for females and not for males (and hence need to propagate sex-specific posterior distributions for local recruitment probability; see the "Consequences of genomic selection on reproductive success" section), we estimated the sex-specific local recruitment probability for selection line fledglings using 159 early and 167 late female selection line fledglings as well as 158 early and 164 late male selection line fledglings. There was no difference in the local recruitment probability between selection lines, sexes, or their interactions (tables S17 and S18).

Consequences of genomic selection on reproductive success

To assess reproductive success, we used the lifetime number of fledglings produced by female selection line recruits, i.e., female selection line fledglings that recruited into the local study population and were identified as a breeding bird, as a proxy for reproductive success. The reported difference in lay date (Fig. 1) and phenological mismatch (Fig. 2) between female selection line recruits from the early ($n = 8$) and late ($n = 9$) line did not translate into a difference in reproductive success [posterior mean difference in the standardized lifetime number of fledglings produced (late-early): 0.31; 89% credible interval: −0.53;1.15; Fig. 4A and tables S19 and S20]. This, however, is in line with the overall weak effect of lay dates on the lifetime number of fledglings produced by local female recruits ($n = 254$) at the study population in the years of the experiment (posterior mean: −0.13, 89% credible interval: −0.25;0.00; Fig. 4B and tables 21 and S22). Moreover, the quality of fledglings produced did not differ between female recruits from the early ($n = 61$) and late ($n = 68$) selection line in terms of fledgling weight, tarsus length, and third primary (P3) length (figs. S6 to S8 and tables S23 to S28).

In addition to assessing differences in reproductive success, we tested whether female recruits from the early and late selection line differed in other aspects that have potential effects on reproductive success. However, female recruits from the early and late selection line did not differ in their daily energy expenditure during chick feeding or daily chick feeding frequency (figs. S9 and S10 and tables S29 to S32). Overall, female recruits from the early and late selection line did not differ in reproductive success when the lifetime number of fledglings produced was used as a proxy for reproductive success.

Consequences of genomic selection on total fitness

Females from the selection lines for early and late avian lay date did not differ in their total fitness, when the total fitness was defined as the product of the local recruitment probability of female selection line fledglings and the lifetime number of fledglings produced by female selection line recruits (table S32). Overall, genomic selection for avian lay date did not lead to any fitness consequences in the wild in the years we performed the experiment.

DISCUSSION

Whether species will be able to cope with global warming-induced changes to their environment depends on how well they can adapt to these new environmental conditions, for instance, by shifting their phenology (*2*, *3*). We, however, lack sufficient data at the very end of the current lay date distribution, making it impossible to estimate whether a further advancement of lay dates would indeed lead to an increase in fitness, i.e., is selected for. We used the Dutch great tit long-term study population at the Hoge Veluwe National Park with the aim to induce directional shifts in the currently observed distribution of lay dates, without inducing a bias in the fitness measurement, as is often the case for phenotypic manipulations (see Introduction). For this, we aimed to create birds

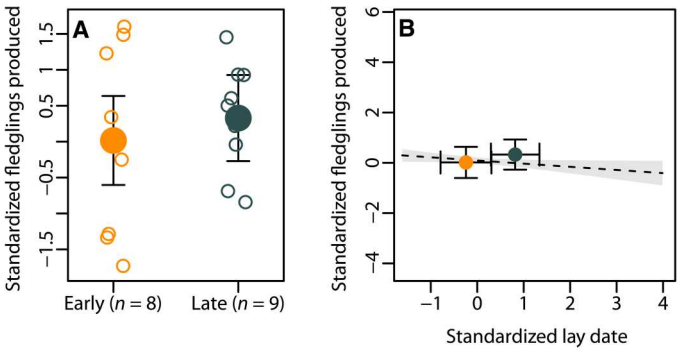

Fig. 4. Posterior predictions of the lifetime number of fledglings produced. (**A**) Posterior means of the mean-standardized lifetime number of fledglings produced with 89% credible interval for female recruits from the early (yellow, $n = 8$) and late (blue, $n = 9$) selection line [posterior mean difference in the standardized lifetime number of fledglings produced (late-early): 0.31; 89% credible interval: −0.53;1.15]. (**B**) Mean lifetime number of fledglings produced by female recruits from the early (yellow) and late (blue) selection line in comparison to the standardized lifetime number of fledglings produced by local female recruits (black/gray, $n = 254$). For female selection line recruits, vertical locations of filled circles and error bars correspond to the posterior mean-standardized lifetime number of fledglings produced with 89% credible interval, and horizontal location of filled circles and error bars correspond to posterior mean-standardized lay dates with 89% credible interval. For local female recruits, black dashed line and gray shaded area show posterior means of the standardized lifetime number of fledglings produced with 89% credible intervals over standardized lay dates. The lifetime number of fledglings produced was standardized as *z* scores using a year-specific SD of 3.13, 3.53, and 3.14 fledglings for 2018, 2019, and 2020, respectively. Lay date observations were standardized as *z* scores using a year-specific SD of 3.26, 5.45, and 4.82 days for 2018, 2019, and 2020, respectively.

with extremely early and late expected lay dates based on their GEBVs, which would allow us to test for the fitness consequences of extremely early and late lay dates, including breeding attempt failure or mortality of females that attempt to lay eggs very early and under harsh conditions. We obtained birds with extreme phenotypes from a genomic selection experiment in which the breeding pairs were selected on the basis of their GEBVs for lay dates (rather than lay dates per se) (*24*, *29*). We moved F4 selection line eggs that were laid in aviaries to the local study population to hatch, fledge, and recruit there as a breeding female the following year(s). We found differentiation in lay dates between female recruits from the early and late selection line, indicating that genomic selection for lay dates was indeed successful. However, while genomic selection delayed lay dates relative to the lay dates of local female recruits, we failed to advance lay dates relative to lay dates of local female recruits. Although the difference in lay dates between female recruits from the early and late selection line translated into a reduced phenological mismatch for female recruits from the early selection line, there was no difference in the lifetime number of fledglings produced. This was in line with the overall weak effect of lay date on the lifetime number of fledglings produced by local female recruits during the same time period.

Genomic selection led to a difference in lay dates in the wild, showing that the difference in GEBVs for lay dates of female recruits from the early and late selection line indeed translated into a phenotypic response under wild conditions. While we refer to this response as asymmetric, because genomic selection delayed lay dates relative to the lay dates of local female recruits while it failed to advance lay dates relative to lay dates of local female recruits, it is important to emphasize that the two selection lines did not statistically differ in their response to selection. Nevertheless, the phenotypic response observed in the wild is in line with the phenotypic response observed in half-open aviaries, where F3 generation females from the early and late selection line of the genomic selection experiment also showed a difference in lay dates (*29*). Moreover, early selection line females showed a smaller shift in lay dates over generations (mean lay date in days from 1 April; F1: 13.9 days ±2.5 and F3: 12.6 days ±2.4) than late selection line females [mean lay date in days from 1 April; F1: 15.9 days ±3.6 and F3: 22.2 days ±3.4; (*29*)]. However, it is difficult to directly compare lay dates in the wild with lay dates in aviaries because of the complex and environment-dependent nature of the trait (*31*, *32*). Nevertheless, the asymmetric shift in lay dates in aviaries, where ad libitum food is supplied, indicates that food availability early in spring per se is unlikely to explain the asymmetric response in lay dates.

The asymmetric response to genomic selection can potentially be explained by a methodological limitation with regard to the efficiency of genomic selection. Especially in wild study systems, we are limited by small sample sizes and statistical power as a large number of individuals genotyped at a sufficient marker density is required. These limitations are especially relevant for phenological traits that generally have low heritability [~0.2 for lay dates in the local study population; (*24*)] and might require a sample size of thousands. Here, a rather large training population of >2000 wild great tit females with known lay dates and genotyped at >500,000 SNPs was used for estimating GEBVs. However, estimated accuracy of GEBVs for lay date was moderate [~0.2; (*24*)] and rather low for what is normally reported in domesticated species [e.g., dairy cattle (*33*) and crop species (*34*)]. Comparably, low accuracy of GEBVs for lay dates might be explained by lower sample size than typically available for domesticated species [>20,000 genotyped individuals; (*34*, *35*)] and higher number of independently segregating genome segments (*36*). Moreover, the experimental design did not include any replicated lines, which makes it difficult to differentiate any direct response to genomic selection from genetic drift. On the basis of a fixation index (F_{st}) outlier analysis using a method that supposedly can distinguish drift from selection, we previously established a strong response to genomic selection at the genetic level (*29*). This indicates that the phenotypic differentiation observed in aviaries and the wild is at least partly a response to the genomic selection experiment rather than genetic drift alone.

An alternative explanation for the asymmetric phenotypic response to genomic selection is the presence of environmental constraints early in the breeding season that limit a translation of GEBVs for early lay dates into early phenotypes. (*37*). Harsh environmental conditions during the energy-intense period of egg production can directly affect female condition by reducing female survival probability (*38*), decreasing foraging efficiency (*39*) and increasing energetic costs for the production of eggs (*40*, *41*) or have carry-over effects on brood success, for example, in the form of reduced caring capacity during chick feeding (*42*). Especially, the direct effects on female condition before egg laying might cause females, which are genetically primed to have early lay dates, to fail their breeding attempt and are consequently not recorded as a local recruit. However, we did not find any indication that early and

late selection line females differed in their survival in late winter or early spring. This indicates that the lack of advancement is not simply explained by females that attempted to lay their eggs early but died in the attempt to do so, as a consequence of environmental constraints in early spring. Moreover, although early selection line females were selected for GEBVs for extremely early lay dates, it is possible that environmental variables (e.g., ambient temperature) set some kind of a hard threshold for the earliest lay dates (*43*). [However, see (*44*) for details on the concept of lay date as threshold trait.] Albeit we do not find any clear indication for environmental constraints on early lay dates, we cannot exclude their presence.

In addition to the phenotypic response to genomic selection, we assessed the fitness consequences of genomic selection for lay dates in the form of two fitness components: the fitness costs of advanced lay dates before egg laying in terms of apparent survival until local recruitment and the fitness benefits of a reduced phenological mismatch in terms of lifetime reproductive success. While we discussed apparent survival probability in the context of environmental constraints in early spring (see above), it is important to clarify that our experimental design allowed us to test for a difference in local recruitment probability as a direct consequence of expected early and late lay dates based on GEBVs for lay date. Lay date and subsequent fledging date are known to affect the local recruitment probability of fledglings (*30*, *45*). When moving the F4 selection line eggs from the aviaries into the Hoge Veluwe study population, we prepared the clutches such, that both early and late selection line eggs were included in each clutch and, this way, ensured that early and late selection line individuals were reared in a common environment that did not propagate the differences in lay dates between early and late F3 generation selection line females in the aviaries (i.e., the lay date of genetic mothers of early and late selection line individuals). Hence, our experimental design allowed us to test for a difference in local recruitment probability as a direct consequence of expected early and late lay dates based on GEBVs for lay date.

In line with our expectations, the difference in lay date between female recruits from the early and late selection line indeed translated into a reduced phenological mismatch for female recruits from the early selection line (relative to female recruits from the late selection line). In contrast to our expectation based on previous findings (*6*), the reduced phenological mismatch, however, did not result in a difference in the lifetime number or quality of fledglings produced between female recruits from the early and late selection line. Although there was strong directional selection for advanced lay dates in past decades as a consequence of the phenological mismatch at the local study population (*6*, *46*, *47*), we recently showed that natural variability in temperature has reduced the phenological mismatch in recent years and decreased selection pressures on lay dates (*4*). The small number of selection line recruits led to low statistical power for detecting a difference in fitness between female recruits from the early and late selection line. However, even when using all local female recruits of the study population, we did not detect any effect of lay date on the lifetime number of fledglings produced, such that a larger sample size would likely not have change our findings. Using projected temperatures from a large ensemble of climate simulations, we showed that global warming will again lead to an intensified phenological mismatch between great tits and caterpillar biomass peak date in the long run (*4*).

Global warming has shifted phenological traits in many species, but whether species are able to further track increasing temperatures depends on the fitness consequences of a shifted phenology. To test this, *we* here applied genomic selection for lay dates in a long-term study population of great tits and assessed the consequences of expected early and late lay dates based on GEBVs for lay dates in the wild. While genomic selection led to a differentiation in lay dates under wild conditions, we did not advance lay dates relative to lay dates of local females recruits from the wild study population. Moreover, the differentiation in lay dates between female recruits from the early and late selection line did not translate into a difference in reproductive success, which is in line with reduced selection on lay dates in the years of the experiment in the study population. Such a reduction of selection despite the presence of a phenological mismatch in most years indicates that the consequences of lay dates on reproductive success are multifaceted and may not be explained by the phenological (mis)match during chick provisioning alone. In the light of future global warming, it is important to identify the constraints that led to an asymmetric phenotypic response to genomic selection, as climate projections hint toward anew increased selection pressures toward early lay dates (*4*).

MATERIALS AND METHODS

Genomic selection lines for early and late lay dates

Genomic selection is commonly applied in animal and plant breeding and has proven a powerful tool for artificial selection (*26*), but we are not aware of a study where genomic selection has been applied in a wild population. Here, we applied genomic selection for early and late lay date (i.e., the date a female initiates egg laying in a respective year) in a wild long-term study population of great tits at the Hoge Veluwe National Park (The Netherlands). A wild training population of >2000 great tit females from the study population with known lay dates and genotyped at >500,000 SNPs was used to estimate GEBVs using the GBLUP approach (*24*, *28*). In this approach, the pedigree-based relatedness matrix within an animal model is replaced by a SNP-based relatedness matrix. The animal model constitutes a specific form of a mixed-effect model frequently used in quantitative genetic studies (*48*) and the BLUPs for the additive genetic effect constitute the (genomic) BVs. Using GEBVs rather than lay dates as entity of selection had two advantages: (i) GEBVs of males can be estimated and hence males can be used for the selection procedure, and (ii) GEBVs can be estimated in the nestlings and supersede the need to wait for females to express the phenotype during their first year of breeding before being selected into breeding pairs. A detailed description on how the GEBVs for lay dates were estimated is provided in (*24*).

To initiate the selection lines for early and late lay dates, 28 breeding pairs from the Hoge Veluwe study population were selected in 2014 as "parental" generation based on their BVs for lay date (*29*). Rather than the parental generation itself, all nestlings of the parental generation were taken to the aviary facilities at the Netherlands Institute of Ecology (NIOO-KNAW) on the 10th day after hatching (d10) where they were hand raised until independence [see (*49*) for details] and constituted the F1 generation. The nestlings were genotyped to estimate their GEBVs for lay date, and individuals with the most extreme GEBVs were selected into breeding pairs (n = 20 breeding pairs per selection line) to breed in aviaries while maintaining as much of the genetic variation within line as possible [see (*29*) for details]. Eggs laid during the breeding season of 2015 (that constituted the F2 generation) were moved

into the nests of wild foster parents that undertook the brood care. At d10, the nestlings were again brought to the aviary facilities at the NIOO-KNAW for further hand raising and continuation of the selection lines (via GEBVs). The procedure was repeated until the F3 generation was produced. A detailed description of the selection procedure is provided in (*29*).

Selection line eggs moved into the wild

To introduce selection line females with expected early and late lay date based on their GEBVs into the wild, we moved F4 generation selection line eggs (i.e., eggs laid in the aviaries by F3 selection line females) into the local study population. During the breeding seasons of 2017–2019, we housed same-selection line breeding pairs of the F3 generation in half-open aviaries with nest boxes and nest building material [under housing conditions as described in (*29*)]. Every morning shortly after sunrise (earliest at 6 a.m.), we checked all nest boxes for newly laid eggs, which we replaced by artificial eggs. We marked all collected eggs with a unique identification code and stored them for up to 14 days at ambient temperature (or 10°C when ambient temperature exceeded 10°C) on an egg turner (Automatic Egg Turner, GQF, Georgia, US). To rear F4 selection line individuals in a common environment that does not propagate any difference in maternal lay date between the selection lines, we prepared mixed-selection line clutches with (up to 12) eggs from both selection lines. With the mixed-selection line clutches, we aimed to rear selection line individuals in a common environment in which early and late selection line individuals did not differ in their realized hatch date and subsequent fledging date. In a few cases where there were no eggs from one of the selection lines available, we prepared same-selection line clutches (11 of 72 clutches in 2017 and 3 of 94 clutches in 2020; table S33). To optimized the use of limited nests available in the local study population at the Hoge Veluwe, we moved the selection line clutches in two steps. First, we moved clutches to nests of incubating females at one of two intermediate locations [the great tit populations in Bennekomse Bos (n = 2549) and Heikamp (n = 400), The Netherlands], such that the foster females started to incubate the selection line clutches in the wild. After 5 days of incubation, we selected the eggs that showed embryonic development that were, in the second step, moved to nest boxes at the local study population at the Hoge Veluwe, the final location (n = 936). This way, we discarded any eggs that were unfertilized or got damaged during handling of the eggs before the move to the final location. Depending on the number of eggs within a clutch that showed embryonic development, we merged eggs from different clutches to increase clutch size. At the local study population, we selected nests where females had initiated incubation approximately 5 days before the final egg movement to match the developmental time of the pre-incubated selection line eggs with the development time of the discarded eggs from the wild foster parents.

We monitored the foster nests with selection line eggs at the local study population in line with the standard protocol at the local study population (see text S1) with minor modifications and additions. The standard protocol includes a "capture" of nestlings on d15 during which we equipped selection line nestlings with passive integrated transponder (PIT) tags (2.6-mm EM4102 PIT bird tag, Eccel Technology Ltd., Leicester, UK) in addition to the aluminum rings that have unique identifier codes. The PIT tags have unique PIT-tag IDs and were used to identify nestlings that fledged and recruited into the local study population the following year(s) (see below). During the d15 capture, we also measured the weight, tarsus length, and length of the P3 feather and took a 10 μl of blood sample stored in 1 ml of Queen's buffer. We used the blood samples to assign the selection line nestlings to their genetic parents (i.e., F3 generation selection line breeding pairs) via molecular markers following a previously established protocol [see e.g. (*50*)].

We moved a total of 936 F4 generation selection line eggs from the NIOO-KNAW to the local study population at the Hoge Veluwe of which 475 and 461 eggs were derived from the selection lines for early and late lay dates, respectively. The number of eggs moved to the study population differed between years, with 358, 456, and 122 eggs in 2017, 2018, and 2019, respectively (table S34).

Selection line birds in the wild

In the years following the introduction of selection line eggs to the local study population (i.e., in 2018–2020), we monitored female fledglings from the selection lines for early and late lay dates that recruited into the local study population. We monitored nest boxes of female selection line recruits in accordance with the standard protocol at the study site (see text S1) with minor modifications and additions. Instead of identifying selection line females during the capture of adults at d7, we identified and localized selection line females during the start of their nest building activities in late March (see text S2), which, in combination with sightings from roosting inspections in January (see text S2), also provided an estimate of apparent winter survival of selection line fledglings before their first breeding attempt. We measured the daily energy expenditure of female selection line recruits in the 24-hour period between the d10 capture and a subsequent capture at d11 (see text S3). Last, we measured the daily feeding frequency of female selection line recruits on d12 and d13 using PIT-tags and transponder readers (see text S4).

Statistical analysis

For statistical interference, we applied generalized linear multilevel models using Stan's Hamiltonian Monte Carlo algorithm to estimate the respective posterior probability distributions. We implemented the models in R v4.0.3 (*51*) using the R package rethinking v2.13 (*52*) with R package cmdstanr v0.3.0.9000 (*53*) as an interface to Stan [CmdStan v2.28.2; (*54*)]. Following McElreath (*55*), we report the 89% credible interval (instead of the commonly used 95% interval) with the posterior means to discourage readers from conducting unconscious hypothesis testing. For all models, we set the number of iterations for sampling to 10,000 for each of four independently sampled chains that were distributed over four processor cores. Half of the iterations for sampling were used as warmup (and do not contribute to the predicted posterior distributions). Trace and trank plots of the Markov chain Monte Carlo (MCMC) output for the models described below and in the appendices are presented in figs. S13 to S44. The R code for the analyses is publicly available at Dryad (doi:10.5061/dryad.2280gb5ws).

Phenotypic response to genomic selection in the wild

Lay date. We assessed the phenotypic response to genomic selection in the wild by testing for a difference in lay dates between female recruits from the early and late selection line as well as between female selection line recruits and local female recruits. We included the lay dates of all females that recruited to the local study sites between 2018 and 2020, i.e., all females that had a lay date

in those years. For female selection line recruits (n = 9 for the early and n = 9 for the late selection line), we used the lay date in the first year of breeding as the number of females that bred in more than 1 year was too small to account for repeated measurements. For local female recruits, we focused on all first egg lay dates recorded within a year (irrespective of the female) to get the least biased representation of the year-specific distribution of lay dates (n = 433 lay dates over the three-year period). Because of the small number of female selection lines recruits, we did not account for year effects but used within-year standardized lay dates for the analysis. For the standardization, we used all observed lay dates of local female recruits within a year (i.e., excluding the lay dates of female selection line recruits) to derive estimates of the year-specific mean and SD that are representative for the local study population. We used these year-specific estimates to calculate year-specific z scores of lay dates for female selection line recruit and local female recruits, i.e., we subtracted the year-specific mean from each observed lay date within that year and divided the resulting difference by the year-specific SD. We detected a potential outlier observation for an early female selection line recruit (fig. S11 and see text S5 for details) and removed it from the statistical analysis (reducing the dataset to n = 8 for female recruits from the early and n = 9 for female recruits from the late selection line). However, repeating the statistical analysis while keeping the outlier observation included did not qualitatively change the results (fig. S12).

We specified a linear regression model to estimate the posterior distribution of standardized lay dates (Eq. 1). Within the regression model, we specified an intercept α and an effect of the selection line $\beta l_{\text{Line[i]}}$ with weakly regularizing priors

$$\begin{gathered} L_i \sim \text{Normal}(\mu_i, \sigma) \\ \mu_i = \alpha + \beta l_{\text{Line[i]}} \\ \alpha \sim \text{Normal}(0, 1.5) \\ \beta l_j \sim \text{Normal}(0, 1.5), \textit{for } j = 1..3 \\ \sigma \sim \text{Half} - \text{Normal}(0, 1) \end{gathered} \tag{1}$$

Phenological mismatch. We tested whether differential lay dates between female recruits from the early and late selection line translated into a difference in phenological mismatch. Using the same dataset as above (see the "Lay dates" section), we calculated the expected year-specific dates of highest resource demands for chick feeding by adding 33 days to the year-specific lay dates (*56*). We used the expected year-specific dates of highest resource demands and the year-specific caterpillar biomass peak dates (which were April date 37, 48, and 40 for 2018, 2019, and 2020, respectively) to calculate the year-specific phenological mismatch as the difference between both (i.e., we subtracted the year-specific caterpillar biomass peak dates from the expected year-specific dates of highest resource demands). We estimated the predicted posterior distributions for the phenological mismatch across years for female selection line recruits and local female recruits and estimated the difference in phenological mismatch between those groups. We standardized the calculated phenological mismatches within a year by estimating the year-specific SD of phenological mismatch excluding female selection line recruits and divided each calculated phenological mismatch within a year by the year-specific SD. (We did not mean center the data.) To estimate the posterior distribution of the standardized phenological mismatch, we specified a linear regression model (Eq. 2). Within the regression model, we specified an intercept α and an effect of the selection line $\beta l_{\text{Line[i]}}$ with weakly regularizing priors

$$\begin{gathered} \text{MM}_i \sim \text{Normal}\ (\mu_i, \sigma) \\ \mu_i = \alpha + \beta l_{\text{Line[i]}} \\ \alpha \sim \text{Normal}(0, 1.5) \\ \beta l_j \sim \text{Normal}(0, 1.5), \textit{for } j = 1..3 \\ \sigma \sim \text{Half} - \text{Normal}(0, 1) \end{gathered} \tag{2}$$

To gain insights on the magnitude of the phenological mismatch in the local study population within the years of the experiment, we estimate the year-specific phenological mismatch and the pairwise differences in year-specific phenological mismatch. The analysis is presented in text S6.

Consequences of genomic selection for fitness

We assessed the fitness consequences as the lifetime number of fledglings a female selection line fledgling produced. The more common alternative to the approach taken here would be to assess fitness as the lifetime number of local recruits produces per female selection line recruit. However, such a "mixed fitness" measure is a function of the fitness of both the female's reproductive success and the offsprings' survival (from zygote to recruitment) and, hence, is prone to bias (*57*). Our genomic selection experiment allows us to unequally look at fitness from fledgling to fledgling as we can assign an expected lay date phenotype already at the fledgling stage, which is not possible for local fledglings. Our fitness measure is a "mixed-fitness" measure in its own right, but the potential bias on our fitness measure resulting from offspring survival is notably reduced when the period from zygote to fledgling is considered rather than the period from zygote to recruitment.

We derived an estimate of the fitness consequences of the genomic selection experiment in two steps; we considered proxies for the survival of female selection line fledglings (e.g., local recruitment probability) and proxies for reproductive success (e.g., the lifetime number of fledglings produced by female selection line fledglings that recruited into the local study population). We, lastly, combined estimates of local recruitment probability and the lifetime number of fledglings produced by female selection line fledglings to estimate a proxy of the total fitness of early and late selection line females.

Potential selective disappearance of early selection line females. One proxy for the survival of female selection line fledglings that we assessed is the potential selective disappearance of early selection line females before their first breeding event. This selective disappearance might arise as a consequence of environmental fitness costs for females that attempt to breed early [i.e., second scenario (*5*)]. For this, we tested whether early selection line females were less likely than late selection line females to recruit into the local study population when females were identified in late winter or during nest building in early spring. The analyses are described in text S2.

Local recruitment probability. Another proxy for the survival of female selection line fledglings we assessed is the local recruitment probability (which is later used for calculating the total fitness; see the "Consequences of genomic selection on total fitness" section). We consider fledglings that returned to the local study population in the year(s) following the year of fledging as a local recruit. For the analysis, we included records from all fledglings that fledged from 2017 to 2019. In those years, there were a total of 2347 fledglings at

the local study population of which 649, 752, and 946 fledged in 2017, 2018, and 2019, respectively. The 2347 fledglings included 318 early and 331 late selection line fledglings and 1698 local fledglings (table S35). We excluded 23 local fledglings for which no data on fledgling weight (measured on d15) were available. Local recruitment was encoded as a binary variable with one for fledglings that locally recruited and zero for all other fledglings. To estimate the posterior distribution of local recruitment probability at the fledgling level, we specified a generalized linear multivariate model assuming a binomial distribution over local recruitment probability (Eq. 3) following (*55*). The Binomial distribution is defined by two parameters: the constant probability of success p (here local recruitment) over each of n trials. Here, we consider local recruitment probability p_i at the level of individual fledglings (for i = 1..2324) and hence set the number of trials for each fledgling n_i to 1 (a special case of the Binomial distribution that is also referred to as Bernoulli distribution). We used a logit link function to bind the linear model for p_i to values between 0 and 1. Within the generalized linear multivariate model, we specified year- and brood-specific intercepts $\alpha_{Year[i]}$ and $\gamma_{Brood[i]}$ to account for the hierarchical structure of our data (i.e., brood nested within year), an effect of the selection line $\beta l_{Line[i]}$, an effect of a proxy for fledging date βd (April dates on d15), and an effect of fledgling weight βw. We standardized the fledgling weight and the proxy for fledging date (April date on d15) using z scores (i.e., we subtracted the mean fledgling weight or fledging date from each observation and divided the resulting difference by the SD of fledgling weight or fledging date). We overall used weakly regularizing priors, and for the year- and brood-specific intercepts, we specified the priors as a function of other parameters (termed hyperparameters), $\overline{\alpha}$, σ_α, and σ_γ, for which we also used weakly regularizing priors (termed hyperpriors). This specification of adaptive priors allowed us to pool information across years and broods meaning that the model adaptively learns about the prior that is common to the above specified intercepts.

$$
\begin{gathered}
R_i \sim \text{Binomial}(1, p_i) \\
\text{logit}(p_i) = \alpha_{Year[i]} + \gamma_{Brood[i]} + \beta l_{Line[i]} + \beta d * \text{Date}_{[i]} + \beta w * \text{Weight}_{[i]} \\
\beta l_j \sim \text{Normal}(0, 1.5), \textit{for } j = 1..3 \\
\beta d \sim \text{Normal}(0, 1.5) \\
\beta w \sim \text{Normal}(0, 1.5) \\
\alpha_j \sim \text{Normal}(\overline{\alpha}, \sigma_\alpha), \textit{for } j = 1 \ldots 3 \\
\gamma_j \sim \text{Normal}(0, \sigma_\gamma), \textit{for } j = 1..95 \\
\overline{\alpha} \sim \text{Normal}(0, 1.5) \\
\sigma_\alpha \sim \text{Half} - \text{Normal}(0, 1) \\
\sigma_\gamma \sim \text{Half} - \text{Normal}(0, 1)
\end{gathered}
\tag{3}
$$

To aid approximation of the posterior distribution for local recruitment probability, we increased the target acceptance rate during sampling of the posterior to 99% (default: 95%) and reparametrized the model. Steep regions of the parameter space can be difficult to explore and, this way, harm the efficiency of the chains [which is a common problem in multilevel models (*55*)]. A reparameterization into a mathematically equivalent but numerically different version can increase the efficiency of chains. The alternative model (Eq. 4) constitutes a noncentered reparameterization of the initial model (Eq. 3) in which the parameters embedded within the adaptive priors of α_j and γ_j (i.e., the hyperparameters $\overline{\alpha}$, σ_α, and σ_γ) were moved out of the definition. For this, we defined some new variables $z_{Year[i]}$ and $x_{Brood[i]}$ that followed a standard Normal distribution and reconstructed the original variables by reversing the transformation within the definition of the linear model ($\overline{\alpha} + z_{Year[i]}\sigma_\alpha$ and $x_{Brood[i]}\sigma_\gamma$, respectively)

$$
\begin{gathered}
R_i \sim \text{Binomial}(1, p_i) \\
\text{logit}(p_i) = \overline{\alpha} + z_{Year[i]}\sigma_\alpha + x_{Brood[i]}\sigma_\gamma + \beta l_{Line[i]} + \beta d * \text{Date}_{[i]} + \beta w * \text{Weight}_{[i]} \\
\beta l_j \sim \text{Normal}(0, 1.5), \textit{for } j = 1..3 \\
\beta d \sim \text{Normal}(0, 1.5) \\
\beta w \sim \text{Normal}(0, 1.5) \\
z_j \sim \text{Normal}(0, 1), \textit{for } j = 1..3 \\
x_j \sim \text{Normal}(0, 1), \textit{for } j = 1..95 \\
\overline{\alpha} \sim \text{Normal}(0, 1.5) \\
\sigma_\alpha \sim \text{Half} - \text{Normal}(0, 1) \\
\sigma_\gamma \sim \text{Half} - \text{Normal}(0, 1)
\end{gathered}
\tag{4}
$$

In the above-described model for local recruitment probability (Eqs. 3 and 4), we included effects of fledgling weight and a proxy for fledging date as both are suggested to affect recruitment probability (*30*). However, fledgling weight might be affected by fledging date as we expect higher availability of food resources in early spring. To better understand whether both effects should be included, we compared different models that included effects of both fledging date and fledgling weight, an effect of fledgling weight only, and an effect of fledging date only. For model comparison, we used the Pareto-smoothed importance sampling (PSIS) cross-validation approximation implemented within the rethinking R package [for more details, see (*55*)]. PSIS provides feedback about its own reliability by emphasizing observations with very high weight (i.e., Pareto k values > 0.5) that might make the PSIS scores unreliable. Here, no Pareto k values > 0.5 were noted, and comparison showed that the model including effects for both fledging date and fledgling weight had the lowest PSIS score (PSIS with SEs = 1084.79 ± 61.72; table S36). However, the difference between the best and second-best model, the model including an effect for fledgling weight only, was 1.3 with an SE of the difference of 4.68, indicating that both models performed similarly well.

As fledging date is suggested to affect recruitment probability (*30*), we designed the experiment with the aim to rear selection line individuals in a common environment in which early and late selection line individuals did not differ in their realized hatch date and subsequent fledging date. We formally tested this assumption, and the analysis is described in text S7. Moreover, we tested for a difference in fledgling weight of early and late selection line fledglings; hence, a difference might be a correlated response to genomic selection for lay date. The analysis is described in text S8.

The above-described model for local recruitment probability (Eq. 4) allowed us to assess the local recruitment probability of selection line fledglings in comparison to local fledglings but did not allow us to estimate the local recruitment probability for female selection line fledglings, as the sex of fledglings was only determined (via molecular markers; see above) for selection line fledglings, but nor for all other fledglings at the local study site. The 318 early selection line fledglings included 159 females and 158 males, and the 331 late selection line fledglings included 167 females and 164 males (note that for one early selection line fledgling, the molecular sex determination failed, such that one early selection line fledgling was excluded from the analysis). As the sex of local fledglings that

did not recruit remained unknown, we fitted another model that retained only data from selection line fledglings. This way, we could estimate the sex-specific local recruitment probability for selection line fledglings. We used the same model structure as for the above-described model (Eq. 4) but additionally specified an effect of sex using a weakly regularizing prior (i.e., normal distribution with mean = 0 and SD = 1.5). The posterior distributions of the recruitment probability for early and late female selection line fledglings were used to estimate the total fitness (see the "Consequences of genomic selection on total fitness" section).

Lifetime number of fledglings produced. We used the lifetime number of fledglings produced by female fledglings that recruited into the local study population as a proxy for reproductive success. We calculated the lifetime number of fledglings produced by female recruits for all females with known lay dates (n = 9 for female recruits from the early and n = 9 for female recruits from the late selection line and n = 433 for local female recruits; see the "Lay dates" section). We first calculated the lifetime number of fledglings produced within each year and performed a within-year standardization, then imputed the total number of fledglings for local female recruits where one or more broods were potentially affected by research-related manipulations (which is only the case for local females recruits, not for female selection line recruits) and, last, used the within-year standardized total number of fledglings for each year to calculate the sum across years resulting in the standardized lifetime number of fledglings (for details see text S9). This way, our final dataset included eight and nine female recruits from the early and late selection line, respectively, and 254 local female recruits.

We first focused on female selection line recruits to test for a difference in reproductive success between the selection lines (which is later used for calculating the total fitness; see the "Consequences of genomic selection on total fitness" section). We estimated the posterior distribution of the standardized lifetime number of fledglings produced by female selection line recruits by specifying a linear regression model (Eq. 5). Within the regression model, we specified an intercept α and an effect of the selection line $\beta l_{\text{Line[i]}}$ with weakly regularizing priors

$$\begin{aligned} \text{LNF}_i &\sim \text{Normal}(\mu_i, \sigma) \\ \mu_i &= \alpha + \beta l_{\text{Line[i]}} \\ \alpha &\sim \text{Normal}(0, 1.5) \\ \beta l_j &\sim \text{Normal}(0, 1.5), \textit{for } j = 1..3 \\ \sigma &\sim \text{Half} - \text{Normal}(0, 1) \end{aligned} \tag{5}$$

To test for an effect of lay date on the lifetime number of fledglings produced by fledglings that recruited into the local study population, we fitted another model using the data from local (i.e., nonselection line) female fledglings that recruited into the local study population. To estimate the posterior distribution of the standardized lifetime number of fledglings produced by local female recruits, we specified a linear regression model (Eq. 6). Within the regression model, we specified an intercept α and an effect of lay date βd with weakly regularizing priors

$$\begin{aligned} \text{LNF}_i &\sim \text{Normal}(\mu_i, \sigma) \\ \mu_i &= \alpha + \beta d * \text{Lay date}_{[i]} \\ \alpha &\sim \text{Normal}(0, 1.5) \\ \beta d &\sim \text{Normal}(0, 1.5) \\ \sigma &\sim \text{Half} - \text{Normal}(0, 1) \end{aligned} \tag{6}$$

In addition to the lifetime number of fledglings produced by female selection line fledglings that recruited into the local study population as a proxy for reproductive success, we also assessed the quality of fledglings produced in terms of fledgling weight, tarsus length, and P3 length (all measured at d15). The analyses are described in text S10.

Moreover, we tested whether the genomic selection experiment for early and late lay dates resulted in other correlated responses (in addition to lay dates) that might affect fitness in their own right such as the daily energy expenditure during chick feeding or daily chick feeding frequency. The analyses are described in texts S3 and S4, respectively.

Consequences of genomic selection on total fitness. To derive estimates of the total fitness, we multiplied the posterior distribution of the standardized lifetime number of fledglings produced by female selection line recruits with the posterior distributions of the recruitment probability of female selection line fledglings to estimate the posterior distribution of the total fitness of selection line fledglings while propagating the uncertainty of the estimates.

Supplementary Materials

This PDF file includes:
Texts S1 to S10
Figs. S1 to S44
Legends for tables S1 to S36

Other Supplementary Material for this manuscript includes the following:
Tables S1 to S36

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. H. E. Chmura, C. T. Williams, A cross-taxonomic perspective on the integration of temperature cues in vertebrate seasonal neuroendocrine pathways. *Horm. Behav.* **144**, 105215 (2022).
2. C. Parmesan, G. Yohe, A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**, 37–42 (2003).
3. S. J. Thackeray, P. A. Henrys, D. Hemming, J. R. Bell, M. S. Botham, S. Burthe, P. Helaouet, D. G. Johns, I. D. Jones, D. I. Leech, E. B. Mackay, D. Massimino, S. Atkinson, P. J. Bacon, T. M. Brereton, L. Carvalho, T. H. Clutton-Brock, C. Duck, M. Edwards, J. M. Elliott, S. J. G. Hall, R. Harrington, J. W. Pearce-Higgins, T. T. Høye, L. E. B. Kruuk, J. M. Pemberton, T. H. Sparks, P. M. Thompson, I. White, I. J. Winfield, S. Wanless, Phenological sensitivity to climate across taxa and trophic levels. *Nature* **535**, 241–245 (2016).
4. M. E. Visser, M. Lindner, P. Gienapp, M. C. Long, S. Jenouvrier, Recent natural variability in global warming weakened phenological mismatch and selection on seasonal timing in great tits (Parus major). *Proc. Royal Soc. B Biol. Sci.* **288**, 20211337 (2021).
5. M. E. Visser, P. Gienapp, Evolutionary and demographic consequences of phenological mismatches. *Nat. Ecol. Evol.* **3**, 879–885 (2019).
6. M. E. Visser, A. J. van Noordwijk, J. M. Tinbergen, C. M. Lessells, Warmer springs lead to mistimed reproduction in great tits (Parus major). *Proc. R. Soc. Lond. B* **265**, 1867–1870 (1998).
7. V. Reed, S. Grotan, B.-E. Jenouvrier, M. E. Saether, Population growth in a wild bird is buffered against phenological mismatch. *Science* **340**, 488–491 (2013).

8. M. D. Burgess, K. W. Smith, K. L. Evans, D. Leech, J. W. Pearce-Higgins, C. J. Branston, K. Briggs, J. R. Clark, C. R. du Feu, K. Lewthwaite, R. G. Nager, B. C. Sheldon, J. A. Smith, R. C. Whytock, S. G. Willis, A. B. Phillimore, Tritrophic phenological match–mismatch in space and time. *Nat. Ecol. Evol.* **2**, 970–975 (2018).
9. E. G. Simmonds, E. F. Cole, B. C. Sheldon, T. Coulson, Phenological asynchrony: A ticking time-bomb for seemingly stable populations? *Ecol. Lett.* **23**, 1766–1775 (2020).
10. P. R. Grant, B. R. Grant, Predicting microevolutionary responses to directional selection on heritable variation. *Evolution* **49**, 241–251 (1995).
11. L. E. B. Kruuk, J. Slate, A. J. Wilson, New answers for old questions: The evolutionary quantitative genetics of wild animal populations. *Annu. Rev. Ecol. Evol. Syst.* **39**, 525–548 (2008).
12. L. E. B. Kruuk, J. Slate, J. M. Pemberton, S. Brotherstone, F. Guinness, T. Clutton-Brock, Antler size in red deer: Heritability and selection but no evolution. *Evolution* **56**, 1683–1695 (2002).
13. B. C. Sheldon, L. E. B. Kruuk, J. Merila, Natural selection and inheritance of breeding time and clutch size in the collared flycatcher. *Evolution* **57**, 406–420 (2003).
14. D. H. Nussey, E. Postma, P. Gienapp, M. E. Visser, Selection on heritable phenotypic plasticity in a wild bird population. *Science* **310**, 304–306 (2005).
15. P. Marrot, A. Charmantier, J. Blondel, D. Garant, Current spring warming as a driver of selection on reproductive timing in a wild passerine. *J. Anim. Ecol.* **87**, 754–764 (2018).
16. J. Merilä, B. C. Sheldon, L. E. B. Kruuk, Explaining stasis: Microevolutionary studies in natural populations. *Genetica* **112-113**, 199–222 (2001).
17. B. Pujol, S. Blanchet, A. Charmantier, E. Danchin, B. Facon, P. Marrot, F. Roux, I. Scotti, C. Teplitsky, C. E. Thomson, I. Winney, The missing response to selection in the wild. *Trends Ecol. Evol.* **33**, 337–346 (2018).
18. S. Verhulst, J.-Å. Nilsson, The timing of birds' breeding seasons: A review of experiments that manipulated timing of breeding. *Phil. Trans. R. Soc. B* **363**, 399–410 (2008).
19. L. te Marvelde, S. V. Schaper, M. E. Visser, A single long day triggers follicle growth in captive female great tits (Parus major) in winter but does not affect laying dates in the wild in spring. *PLOS ONE* **7**, e35617 (2012).
20. L. Salis, S. P. Caro, R. A. Hut, L. Vernooij, M. E. Visser, Manipulation of photoperiod perception advances gonadal growth but not laying date in the great tit. *J. Avian Biol.* **50**, e02197 (2019).
21. L. te Marvelde, M. E. Visser, Manipulation of life-history decisions using leptin in a wild passerine. *PLOS ONE* **7**, e34090 (2012).
22. S. Daan, C. Dijkstra, R. Drent, T. Meijer, Food supply and the annual timing of avian reproduction in *Volume I: Proceedings XIX International Ornithological Congress, 1986* (Univ. of Ottawa Press, 1989), vol. 1, pp. 392–407.
23. P. Gienapp, M. E. Visser, Possible fitness consequences of experimentally advanced laying dates in Great Tits: Differences between populations in different habitats. *Funct. Ecol.* **20**, 180–185 (2006).
24. P. Gienapp, M. P. L. Calus, V. N. Laine, M. E. Visser, Genomic selection on breeding time in a wild bird population. *Evol. Lett.* **3**, 142–151 (2019).
25. A. Charmantier, D. Garant, L. E. B. Kruuk, *Quantitative Genetics in the Wild* (Oxford University Press, New York, NY, 2014).
26. M. E. Goddard, B. J. Hayes, Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* **10**, 381–391 (2009).
27. C. E. Regan, J. M. Pemberton, J. G. Pilkington, P. T. Smiseth, A. J. Wilson, Linking genetic merit to sparse behavioral data: Behavior and genetic effects on lamb growth in Soay sheep. *Behavioral Ecology* **31**, arz166 (2019).
28. S. Clark, J. van der Werf, Genomic best linear unbiased prediction (gBLUP) for the estimation of genomic breeding values. *Methods Mol. Biol.* **1019**, 321–330 (2013).
29. I. Verhagen, P. Gienapp, V. N. Laine, E. M. Grevenhof, A. C. Mateman, K. Oers, M. E. Visser, Genetic and phenotypic responses to genomic selection for timing of breeding in a wild songbird. *Funct. Ecol.* **33**, 1708–1721 (2019).
30. N. Verboven, M. E. Visser, Seasonal variation in local recruitment of great tits: The importance of being early. *Oikos* **81**, 511 (1998).
31. M. E. Visser, S. P. Caro, K. van Oers, S. V. Schaper, B. Helm, Phenology, seasonal timing and circannual rhythms: Towards a unified framework. *Phil. Trans. R. Soc. B* **365**, 3113–3127 (2010).
32. P. Gienapp, V. N. Laine, A. C. Mateman, K. van Oers, M. E. Visser, Environment-dependent genotype-phenotype associations in avian breeding time. *Front. Genet.* **8**, 102 (2017).
33. P. M. VanRaden, C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, F. S. Schenkel, Invited review: Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* **92**, 16–24 (2009).
34. Z. Lin, B. J. Hayes, H. D. Daetwyler, Genomic selection in crops, trees and forages: A review. *Crop Pasture Sci.* **65**, 1177 (2014).
35. K. Schöpke, H. H. Swalve, Review: Opportunities and challenges for small populations of dairy cattle in the era of genomics. *Animal* **10**, 1050–1060 (2016).
36. K. van Oers, A. W. Santure, I. De Cauwer, N. E. van Bers, R. P. Crooijmans, B. C. Sheldon, M. E. Visser, J. Slate, M. A. Groenen, Replicated high-density genetic maps of two great tit populations reveal fine-scale genomic departures from sex-equal recombination rates. *Heredity* **112**, 307–316 (2014).
37. M. E. Visser, L. te Marvelde, M. E. Lof, Adaptive phenological mismatches of birds and their food in a warming world. *J. Ornithol.* **153**, 75–84 (2012).
38. M. E. Visser, C. M. Lessells, The costs of egg production and incubation in great tits (Parus major). *Proc. R. Soc. Lond. B* **268**, 1271–1277 (2001).
39. M. I. Avery, J. R. Krebs, Temperature and foraging success of Great Tits *Parus major* hunting for spiders. *Ibis* **126**, 33–38 (1984).
40. S. C. Kendeigh, V. R. Dol'nik, V. M. Gavrilov, Avian energetics in *Granivorous Birds in Ecosystemss: Their Evolution, Populations, Energetics, Adaptations, Impact and Control*, J. Pinowski, S. C. Kendeigh, Eds. (Cambridge Univ. Press, 1977), pp. 127–204.
41. L. te Marvelde, S. L. Webber, H. A. J. Meijer, M. E. Visser, Energy expenditure during egg laying is equal for early and late breeding free-living female great tits. *Oecologia* **168**, 631–638 (2012).
42. P. Monaghan, R. G. Nager, D. C. Houston, The price of eggs: Increased investment in egg production reduces the offspring rearing capacity of parents. *Proc. R. Soc. Lond. B* **265**, 1731–1735 (1998).
43. A. J. V. Noordwijk, R. H. McCleery, C. M. Perrins, Selection for the timing of great tit breeding in relation to caterpillar growth and temperature. *J. Anim. Ecol.* **64**, 451–458 (1995).
44. J. M. Reid, P. Acker, Conceptualizing the evolutionary quantitative genetics of phenological life-history events: Breeding time as a plastic threshold trait. *Evol. Lett.* **6**, 220–233 (2022).
45. M. E. Visser, N. Verboven, Long-term fitness effects of fledging date in great tits. *Oikos* **85**, 445–450 (1999).
46. J. J. C. Ramakers, P. Gienapp, M. E. Visser, Phenological mismatch drives selection on elevation, but not on slope, of breeding time plasticity in a wild songbird. *Evolution* **73**, 175–187 (2019).
47. A. Husby, M. E. Visser, L. E. B. Kruuk, Speeding up microevolution: The effects of increasing temperature on selection and genetic variance in a wild bird population. *PLOS Biol.* **9**, e1000585 (2011).
48. A. J. Wilson, D. Réale, M. N. Clements, M. M. Morrissey, E. Postma, C. A. Walling, L. E. B. Kruuk, D. H. Nussey, An ecologist's guide to the animal model. *J. Anim. Ecol.* **79**, 13–26 (2010).
49. P. J. Drent, K. van Oers, A. J. van Noordwijk, Realized heritability of personalities in the great tit (*Parus major*). *Proc. R. Soc. Lond. B* **270**, 45–51 (2003).
50. T. J. Greives, S. A. Kingma, B. Kranstauber, K. Mortega, M. Wikelski, K. Oers, A. C. Mateman, G. A. Ferguson, G. Beltrami, M. Hau, Costs of sleeping in: Circadian rhythms influence cuckoldry risk in a songbird. *Funct. Ecol.* **29**, 1300–1307 (2015).
51. R Core Team, R: A language and environment for statistical computing (R Foundation for Statistical Computing, 2021); (https://www.R-project.org/.
52. R. McElreath, *rethinking: Statistical Rethinking book package* (2020).
53. J. Gabry, R. Češnovar, cmdstanr: R Interface to "CmdStan" (2020);https://mc-stan.org/cmdstanr,https://discourse.mc-stan.org).
54. Stan Development Team, Stan Modeling Language Users Guide and Reference Manual (2019);https://mc-stan.org).
55. R. McElreath, *Statistical rethinking: A Bayesian course with examples in R and Stan* (CRC Press/Taylor & Francis Group, ed. 2, 2020).
56. L.-M. Chevin, M. E. Visser, J. Tufto, Estimating the variation, autocorrelation, and environmental sensitivity of phenotypic SELECTION. *Evolution* **69**, 2319–2332 (2015).
57. C. E. Thomson, J. D. Hadfield, Measuring selection when parents and offspring interact. *Methods Ecol. Evol.* **8**, 678–687 (2017).

Acknowledgments: We thank M. Kaandorp, B. van Lith, and P. de Vries for supporting the fieldwork; the animal caretakers at the NIOO-KNAW for taking care of the birds; M. van der Sluijs and colleagues in the molecular laboratory at NIOOK-NAW for the support with the molecular sex determination; M. Kaandorp and L. Vernooij for database support; and K. van Oers for useful discussions. We would like to thank the board of the National Park "de Hoge Veluwe" for their permission to work within their reserve. We thank A. Wilson, A. Phillimore, and one anonymous reviewer for constructive comments that greatly improved the manuscript. **Funding:** This work was supported by a European Research Council Advanced grant (339092–E-Response) to M.E.V.
Ethics statement: This study was performed under the approval by the Animal Experimentation Committee (DEC), Amsterdam, The Netherlands, protocol NIOO14.10 and addendum 2 to this protocol. **Author contributions:** Conceptualization: M.E.V. and P.G. Methodology: M.E.V., J.J.C.R., and M.L. Investigation: M.L., J.J.C.R., B.T., I.V., and A.C.M. Visualization: M.L. Supervision: M.E.V. Writing—original draft: M.L. Writing—review and editing: M.L., M.E.V., P.G., J.J.C.R., M.L., I.V., and A.C.M. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary

Materials. Data and the R code used for analysis are publicly available at Dryad (doi:10.5061/dryad.2280gb5ws).

Submitted 29 August 2022
Accepted 1 May 2023
Published 7 June 2023
10.1126/sciadv.ade6350