



# Fluctuating optimum and temporally variable selection on breeding date in birds and mammals

Pierre de Villemereuil<sup>a,b,1</sup>, Anne Charmantier<sup>a</sup>, Debora Arlt<sup>c</sup>, Pierre Bize<sup>d</sup>, Patricia Brekke<sup>e</sup>, Lyanne Brouwer<sup>f,g,h</sup>, Andrew Cockburn<sup>f</sup>, Steeve D. Côté<sup>i</sup>, F. Stephen Dobson<sup>i</sup>, Simon R. Evans<sup>k,l</sup>, Marco Festa-Bianchet<sup>m,f</sup>, Marlène Gamelon<sup>n</sup>, Sandra Hamel<sup>o</sup>, Johann Hegelbach<sup>p</sup>, Kurt Jerstad<sup>q</sup>, Bart Kempenaers<sup>r</sup>, Loeske E. B. Kruuk<sup>f</sup>, Jouko Kumpula<sup>s</sup>, Thomas Kvalnes<sup>n</sup>, Andrew G. McAdam<sup>t</sup>, S. Eryn McFarlane<sup>u,v</sup>, Michael B. Morrissey<sup>w</sup>, Tomas Pärt<sup>o</sup>, Josephine M. Pemberton<sup>v</sup>, Anna Qvarnström<sup>u</sup>, Ole Wiggo Røstad<sup>x</sup>, Julia Schroeder<sup>y</sup>, Juan Carlos Senar<sup>z</sup>, Ben C. Sheldon<sup>k</sup>, Martijn van de Pol<sup>g</sup>, Marcel E. Visser<sup>g</sup>, Nathaniel T. Wheelwright<sup>aa</sup>, Jarle Tufto<sup>bb</sup>, and Luis-Miguel Chevin<sup>a,1</sup>

<sup>a</sup>Centre d'Écologie Fonctionnelle et Évolutive, CNRS, Université de Montpellier, Université Paul Valéry Montpellier 3, École Pratique des Hautes Études | Paris Science et Lettres, Institut de Recherche pour le Développement, 34000 Montpellier, France; <sup>b</sup>Institut de Systématique, Évolution, Biodiversité, École Pratique des Hautes Études | Paris Sciences et Lettres, Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, Université des Antilles, 75005 Paris, France; <sup>c</sup>Department of Ecology, Swedish University of Agricultural Sciences, 75007 Uppsala, Sweden; <sup>d</sup>School of Biological Sciences, University of Aberdeen, AB24 2TZ Aberdeen, United Kingdom; <sup>e</sup>Institute of Zoology, Zoological Society of London, NW1 4RY London, United Kingdom; <sup>f</sup>Division of Ecology and Evolution, Research School of Biology, The Australian National University, Canberra, ACT 2600 Australia; <sup>g</sup>Department of Animal Ecology, Netherlands Institute of Ecology, 6700 AB Wageningen, The Netherlands; <sup>h</sup>Department of Animal Ecology and Physiology, Institute for Water and Wetland Research, Radboud University, 6500 GL Nijmegen, The Netherlands; <sup>i</sup>Département de Biologie et Centre d'Études Nordiques, Université Laval, Québec, G1V 0A6 QC, Canada; <sup>j</sup>Department of Biological Sciences, Auburn University, Auburn, AL 36849; <sup>k</sup>Edward Grey Institute, Department of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom; <sup>l</sup>Centre for Ecology and Conservation, University of Exeter, Penryn TR10 9FE, United Kingdom; <sup>m</sup>Département de biologie, Université de Sherbrooke, J1K 2R1 Sherbrooke, Québec, Canada; <sup>n</sup>Centre for Biodiversity Dynamics, Department of Biology, Norwegian University of Science and Technology, 7491 Trondheim, Norway; <sup>o</sup>Département de Biologie, Université Laval, Québec, G1V 0A6 QC, Canada; <sup>p</sup>Institute of Evolutionary Biology and Environmental Studies, University of Zurich, CH-8057 Zurich, Switzerland; <sup>q</sup>Jerstad Viltforvaltning, 4516 Mandal, Norway; <sup>r</sup>Department of Behavioural Ecology and Evolutionary Genetics, Max Planck Institute for Ornithology, 82319 Seewiesen, Germany; <sup>s</sup>Terrestrial Population Dynamics, Natural Resources Institute Finland, FIN-999870, Inari, Finland; <sup>t</sup>Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO 80309; <sup>u</sup>Department of Ecology and Genetics, Uppsala University, 75236 Uppsala, Sweden; <sup>v</sup>Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3FL, United Kingdom; <sup>w</sup>School of Biology, University of St. Andrews, St. Andrews, Fife KY16 9TH, United Kingdom; <sup>x</sup>Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, 1432 Ås, Norway; <sup>y</sup>Department of Life Sciences, Imperial College London, SL5 7PY Ascot, Berks; <sup>z</sup>Behavioural and Evolutionary Ecology Research Unit, Museu de Ciències Naturals de Barcelona, E-08003 Barcelona, Spain; <sup>aa</sup>Department of Biology, Bowdoin College, Brunswick, ME 04011; and <sup>bb</sup>Centre for Biodiversity Dynamics, Department of Mathematics, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway

Edited by Peter R. Grant, Princeton University, Princeton, NJ, and approved September 24, 2020 (received for review May 7, 2020)

**Temporal variation in natural selection is predicted to strongly impact the evolution and demography of natural populations, with consequences for the rate of adaptation, evolution of plasticity, and extinction risk. Most of the theory underlying these predictions assumes a moving optimum phenotype, with predictions expressed in terms of the temporal variance and autocorrelation of this optimum. However, empirical studies seldom estimate patterns of fluctuations of an optimum phenotype, precluding further progress in connecting theory with observations. To bridge this gap, we assess the evidence for temporal variation in selection on breeding date by modeling a fitness function with a fluctuating optimum, across 39 populations of 21 wild animals, one of the largest compilations of long-term datasets with individual measurements of trait and fitness components. We find compelling evidence for fluctuations in the fitness function, causing temporal variation in the magnitude, but not the direction of selection. However, fluctuations of the optimum phenotype need not directly translate into variation in selection gradients, because their impact can be buffered by partial tracking of the optimum by the mean phenotype. Analyzing individuals that reproduce in consecutive years, we find that plastic changes track movements of the optimum phenotype across years, especially in bird species, reducing temporal variation in directional selection. This suggests that phenological plasticity has evolved to cope with fluctuations in the optimum, despite their currently modest contribution to variation in selection.**

adaptation | fluctuating environment | fitness landscape | meta-analysis | phenotypic plasticity

Natural environments vary on multiple timescales, with consequences for the ecology and evolution of species in the wild (1–6). Beyond directional trends (e.g., global warming) and periodic cycles (diurnal, seasonal, pluriannual), most envi-

ronmental variables exhibit random variation or noise (4, 6), the magnitude and temporal pattern of which are currently being altered by human activities (7, 8). From an evolutionary

## Significance

Many ecological and evolutionary processes strongly depend on the way natural selection varies over time. However, a gap remains when trying to connect theoretical predictions to empirical work on this question: Most theory assumes that adaptation involves tracking a moving optimum phenotype through time, but this is seldom estimated empirically. Here, we have assembled a large database of wild bird and mammal populations, to estimate patterns of fluctuations in the optimum breeding date and its influence on the variability of natural selection. We find that optimum fluctuations are prevalent. However, their influence on temporal variance in natural selection is partly buffered by tracking of the optimum phenotype through individual phenotypic plasticity.

Author contributions: P.d.V. and L.-M.C. designed research; P.d.V. performed research; P.d.V., J.T., and L.-M.C. analyzed data; P.d.V., A. Charmantier, D.A., P. Bize, P. Brekke, L.B., A. Cockburn, S.D.C., F.S.D., S.R.E., M.F.-B., M.G., S.H., J.H., K.J., B.K., L.E.B.K., J.K., T.K., A.G.M., S.E.M., M.B.M., T.P., J.M.P., A.Q., O.W.R., J.S., J.C.S., B.C.S., M.v.d.P., M.E.V., N.T.W., J.T., and L.-M.C. wrote the paper; A. Charmantier helped in gathering the data; and A. Charmantier, D.A., P. Bize, P. Brekke, L.B., A. Cockburn, S.D.C., F.S.D., S.R.E., M.F.-B., M.G., S.H., J.H., K.J., B.K., L.E.B.K., J.K., T.K., A.G.M., S.E.M., M.B.M., T.P., J.M.P., A.Q., O.W.R., J.S., J.C.S., B.C.S., M.v.d.P., M.E.V., N.T.W., and J.T. contributed data.

The authors declare no competing interest.

Published under the [PNAS license](#).

This article is a PNAS Direct Submission.

<sup>1</sup>To whom correspondence may be addressed. Email: pierre.devillemereuil@ephe.psl.eu or luis-miguel.chevin@cefe.cnrs.fr.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2009003117/-DCSupplemental>.

First published November 30, 2020.

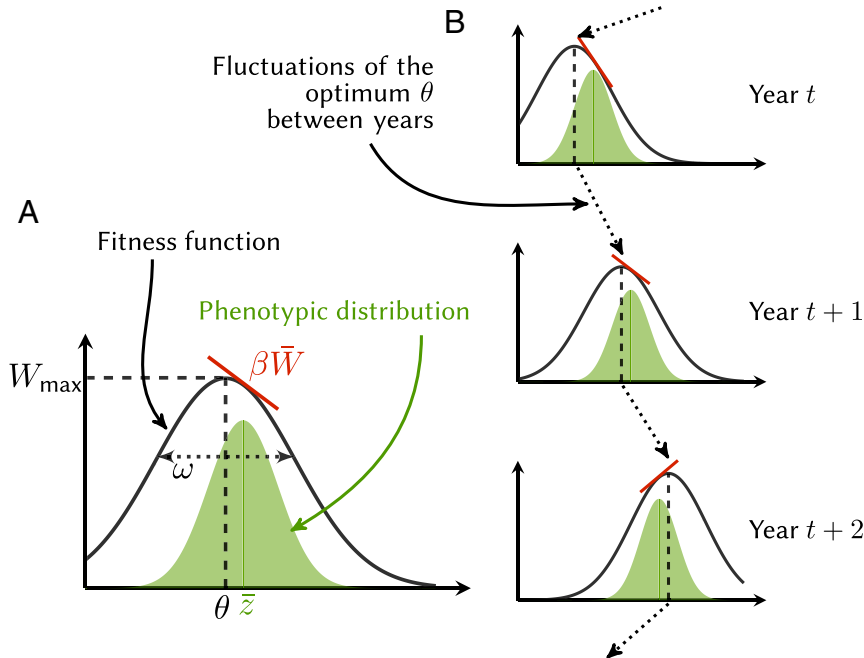
standpoint, these environmental fluctuations are important because they can lead to temporal variation in natural selection. This can in turn maintain genetic polymorphism and phenotypic/genetic variance of quantitative traits (9–12); select for traits that enhance evolvability [including the properties of mutations (13) or recombination (14, 15)]; and favor the evolution of specific mechanisms to cope with environmental fluctuations, from (transgenerational) phenotypic plasticity to bet hedging (12, 16–18). A perpetually fluctuating environment also prevents natural populations from being perfectly adapted to their current conditions at any time, resulting in a “lag load” (19) that may impact population dynamics and extinction risk (20–23). Over macroevolutionary time, temporal variation in selection is also invoked to reconcile observations of rapid responses to selection with the relative paucity of long-term evolutionary change (6, 24–26).

Most theoretical work on adaptation to fluctuating environments rests on the classical framework of “moving optimum models” (27), illustrated in Fig. 1. In this model, directional selection on a quantitative trait is proportional to the deviation of the mean phenotype from an environment-specific optimum phenotype (Fig. 1). Environmental fluctuations in the optimum phenotype can thus lead to temporal variation in directional selection, yet the two are not strictly equivalent, because changes in the expressed mean phenotype also affect temporal variation in deviations from the optimum and thus in selection. A mean phenotype that closely tracks movements of the optimum (via evolution or phenotypic plasticity) can thus buffer the influence of a fluctuating optimum on selection (28, 29).

The wealth of theoretical predictions on adaptation to fluctuating environments (11, 12, 16–18, 20–22, 25) has rarely been explicitly compared to empirical estimates, especially for poly-

genic, quantitative traits, which form the bulk of ecologically important traits such as body size, behavior, or phenology (see ref. 6 for a review on fluctuating selection on discrete traits or major genes). Recent meta-analyses of temporal variation in selection on quantitative traits (30, 31) have shown that—when carefully restricted to datasets for which measurement error was reported (31)—the direction of selection was largely consistent across years, despite evidence for some temporal variation in magnitude of the gradients (31). However, neither of these meta-analyses (30, 31) allowed direct connection with theory, because most theoretical predictions are expressed in terms of the variance and autocorrelation in the optimum (11, 12, 16–18, 20–22, 25), which cannot be recovered directly from variation in selection gradients (as shown by ref. 29). In addition, these meta-analyses (30, 31) could not ascribe temporal variation in selection gradients to movements of the fitness function versus changes in the phenotype distribution (as illustrated in Fig. 1).

Here, we investigate the extent of temporal variation in selection on breeding date. Breeding date can easily be compared across species and is likely to be under selection for an optimum phenotype, because reproducing either too early or too late should limit reproductive success (including offspring survival) and possibly survival of the parents. Changes in phenology (the seasonal timing of life history events) are a predominant phenotypic response to climate change (32–35). Thus, understanding natural selection on phenology is crucial for many eco-evolutionary projections of the effects of current anthropogenic climate change on wild populations (36). In addition, most phenological traits (including breeding time) are plastic in response to environmental variables such as temperature, and this plasticity is thought to have evolved to buffer the ecological



**Fig. 1.** Selection in the moving optimum model. (A) A fitness peak with an optimum (black curve) is modeled as a Gaussian fitness function following classical theory of adaptation. The maximum absolute fitness  $W_{\max}$  is reached at the optimal trait value  $\theta$ , and the width of the fitness peak is parameterized by  $\omega$ . A normal distribution of phenotypes is also shown underneath in green shading (note this distribution has its own scale of probability density, different from the fitness scale on the y axis, but we omit it for simplicity). The strength of directional selection is quantified by the linear selection gradient  $\beta$ , which measures the mean local slope of the relative fitness function, and is proportional to the slope of the red straight line. In this model of Gaussian fitness peak,  $\beta$  is proportional to the deviation of the mean phenotype from the optimum and inversely proportional to  $\omega^2 + 1$  (for SD-standardized traits), such that narrower fitness peaks cause stronger directional selection overall. (B) Temporal changes in the optimum  $\theta$  and in the mean phenotype (mode of the green distribution) jointly translate into changes in selection gradients  $\beta$ . Note that while the maximum fitness  $W_{\max}$  remains constant in this plot, it is allowed to vary in our models.

consequences of a moving optimum in a fluctuating environment (12, 16, 17, 37).

Instead of performing a meta-analysis of published selection estimates, we assembled a database combining 39 long-term datasets from natural populations (13 bird and 8 mammal species; *SI Appendix, Table S1*), over periods spanning from 9 to 63 y. Although parts of these datasets have been published previously, we obtained up-to-date versions by directly contacting the principal investigators. This has allowed us to analyze temporal variation in natural selection using the common framework illustrated in Fig. 1, using individual measurements of traits and fitness components. Based on key elements of the moving optimum theory of adaptation to a changing environment (27), we inquired the following: 1) Is there support for an optimum phenotype? 2) Is there support for a temporally fluctuating fitness function? 3) Does fluctuation of the fitness function translate into temporal variation in the direction and/or magnitude of selection? 4) What is the predictability (autocorrelation) of selection? 5) And to what extent is the effect of a moving optimum buffered by adaptive tracking by the mean phenotype, notably through phenotypic plasticity? While moving optimum models have previously been estimated in a couple of populations (38, 39), we here estimated such models systematically across a large number of populations and systems. This enabled us to report wild-population meta-estimates (robust overall estimators from “meta-analysis” models) of key parameters from the theory of selection in a variable environment.

## Results

**Selection Model.** Consistent with moving optimum models (27), we assumed that the relationship between breeding date and the fitness component exerting selection on it (annual reproductive success) involves a single fitness peak, with an optimum phenotype that fluctuates with the environment (Fig. 1). Denoting as  $W(z)$  the expected fitness component for an individual with breeding date  $z$ , we thus have

$$W(z) = W_{\max} \exp\left(-\frac{(z - \theta)^2}{2\omega^2}\right), \quad [1]$$

where  $\theta$  is the optimum breeding date, for which the expected fitness component is  $W_{\max}$ , and  $\omega$  describes the width of the fitness function. The fitness function in Eq. 1, being quadratic on the log scale (38, 40), uses as many parameters as the quadratic approximation often used in selection analysis (30, 41–43), but is

more realistic, notably because it precludes negative multiplicative fitness (38, 40). This makes it a reasonable approximation for any fitness peak with an optimum (hence its prevalence in theoretical work) (27, 44) and a biologically meaningful benchmark to draw generalizations about temporal variation in selection across populations and species, even if it does not perfectly match the actual fitness function for specific datasets (just like the effective population size allows comparing levels of drift even for non-Wright–Fisher populations).

In such a model, and assuming a normally distributed trait, the directional selection gradient measuring the strength of directional selection is (44)

$$\beta = \frac{\theta - \bar{z}}{\omega^2 + 1}, \quad [2]$$

where  $\bar{z}$  is the mean phenotype. Note that trait values are here divided by their SD  $\sigma_z$ , so  $\beta$  corresponds to a standardized, dimensionless gradient (41), also described as selection intensity ( $\theta$  and  $\omega$  are similarly standardized; for a nonstandardized trait, 1 should be replaced by  $\sigma_z^2$  in Eq. 2). Eq. 2 shows that  $\beta$  is proportional to the deviation of the mean phenotype from the optimum, as illustrated in Fig. 1. Fluctuations in directional selection ( $\beta$ ) can thus result from fluctuations in the optimum phenotype ( $\theta$ ), fluctuations in the mean phenotype ( $\bar{z}$ ), or both. Furthermore, fluctuations in the optimum might result in few to no fluctuations in directional selection, if the mean phenotype appropriately tracks changes in the optimum. For a given deviation from the optimum,  $\beta$  is larger if the fitness peak is narrower, leading to larger values of  $1/(\omega^2 + 1)$ . Note that the strength of stabilizing selection reducing phenotypic variance in any generation is also proportional to  $1/(\omega^2 + 1)$  (or  $1/(\omega^2 + \sigma_z^2)$  for an unstandardized trait), regardless of the deviation of the mean phenotype from the optimum (45, 46), such that the trait can be under both stabilizing and directional selection.

We are interested in distinguishing temporal variation in selection caused by fluctuation in the fitness function from that caused by changes in the mean phenotype (Fig. 1). To this aim, we directly estimated fluctuations of the fitness peak via a random effect for year  $t$  on the optimum  $\theta_t$  in a mixed model, which prevents conflating measurement error with the actual variance in selection (38, 39). We also investigated the temporal predictability of fluctuations in the optimum, by optionally allowing for temporal autocorrelation in the optimum, in the form of a first-order autoregressive process. As alternative models, we also considered fitness functions without an optimum, namely a

**Table 1. Statistical models considered, their characteristics, and relative statistical support for each taxonomic level (birds, 31 datasets; or mammals, 8 datasets; or all taxa together, 39 datasets)**

ID	Shape	Fluctuations	Autocorrelation	Statistical support		
				Bird	Mammal	Total
NoSel	Flat	✗	✗	0.034	0.08	0.043
ConstDir	Monotonic	✗	✗	0.12	0.082	0.112
ConstOpt	Gaussian	✗	✗	0.069	0.182	0.092
FluctDir	Monotonic	✓	✗	0.188	0.104	0.171
FluctOpt	Gaussian	✓	✗	0.194	0.211	0.198
FluctCorrDir	Monotonic	✓	✓	0.141	0.11	0.135
FluctCorrOpt	Gaussian	✓	✓	0.254	0.231	0.249

“NoSel” corresponds to a flat fitness function, i.e., no selection. “Const” models have a fitness function leading to constant selection, “Fluct” models have fluctuating optimum without correlation between years, while “FluctCorr” models have autocorrelated fluctuating optimum. In all models, the intercept was allowed to vary from year to year. Regarding the shape, “Dir” models correspond to a monotonic (directional) function, while “Opt” models include an optimum as described in Fig. 1 and Eq. 1. Relative statistical support is the average of the evidence weights [computed from the LOIC (47), following ref. 48] over the total number of tested models (note that relative statistical supports sum up to 1).

monotonic fitness function where the direction of selection does not change with the mean phenotype in the population (but can still change with the environment), and a flat fitness function causing no selection. The models are summarized in Table 1.

**Fluctuation of the Fitness Function Is Predominant.** We first investigated the support for fluctuating fitness functions, by using an information criterion akin to the Akaike Information Criterion (AIC), the Bayesian leave-one-out information criterion (LOOIC) (47). More specifically, we computed “weights of evidence” inspired by Akaike weights used in model averaging (48) (and summing to 1 across all compared models), which we used to compare the statistical support for different features of selection across datasets (see Table 1). The results of model selection for each dataset appear in *SI Appendix, Table S2*. We found little support for models without selection (flat fitness function, 3.4 and 8%, respectively, for birds and mammals). The statistical support for an optimum was dominant (optimum vs. directional models: 51.7 vs. 44.9% for birds and 62.4 vs. 29.6% for mammals). Similarly, the support for fluctuating fitness functions was also dominant (fluctuating vs. constant models: 77.7 vs. 22.3% for birds and 65.6 vs. 34.4% for mammals). Those results were qualitatively unchanged when considering a completely balanced setting using ConstDir/ConstOpt models as the sole contestants for “no fluctuation” and FluctCorrDir/FluctCorrOpt as the sole contestants for “fluctuating fitness functions.” For some datasets, especially the smaller ones and/or those where fitness was analyzed as a binary trait, there was considerable uncertainty regarding the best model(s), even when there was clear evidence for fluctuating fitness functions. For two datasets, the mountain goat (*Oreamnos americanus*, Oam) and the red-winged fairy wren (*Malurus elegans*, Mel), the support for an absence of selection was dominant (weight above 0.5), so we removed them from subsequent analyses to avoid commenting on spurious signals. In the rest of this paper, and for the sake of simplicity, we focus on the (maximal) model with an autocorrelated fluctuating optimum, unless otherwise noted. However, we also discuss the support for different aspects of the model when commenting on the results.

**The Optimum Fluctuates Differently between Birds and Mammals.** In datasets with predominant support for an optimum (relative support  $>0.5$  among models with selection), the peak width  $\omega$  was typically large (*SI Appendix, Figs. S1 and S2*), with a meta-estimate of 6.22 (95% higher posterior density credible interval [3.2, 9.4]) for birds and of 4.94 ([1.2, 9.2]) for mammals. Such values (in units of within-year phenotypic SD) correspond to weak stabilizing selection (fitness peak broader than phenotype distribution), consistent with previous estimates from the literature and with values commonly used in theory (42, 43, 49). A few notable exceptions had a narrow fitness peak with a low value of  $\omega$  (e.g., an Alpine swift dataset, *Tachymarptis melba*, Tme1; the eastern gray kangaroo, *Macropus giganteus*, Mgi; the oystercatcher, *Haematopus ostralegus*, Hos; and the reindeer, *Rangifer tarandus*, Rta). The lowest  $\omega$  was found in the hihi (*Notiomystis cincta*, Nci, 1.77 [1.56, 2.03]).

The mean location of the optimum  $\theta_t$  was often inferred to be significantly negative, implying that the average optimal timing was usually earlier than the average mean breeding date across years (Fig. 2). In the three cases when a point estimate was inferred to be positive, the sign of the estimate was uncertain (i.e., 95% credible intervals overlap zero), despite strong support for a model with an optimum for one of them (a blue tit, *Cyanistes caeruleus*, Cca10). The meta-estimate for birds was different from zero ( $-3.7$ ,  $[-7.5, -0.7]$ ), while that for mammals was not ( $-1.75$ ,  $[-6.4, 3.0]$ ; Fig. 2).

The magnitude of fluctuations in the optimum differed strongly between datasets, with 5 datasets (out of 20 with predominant support for an optimum) displaying low varia-

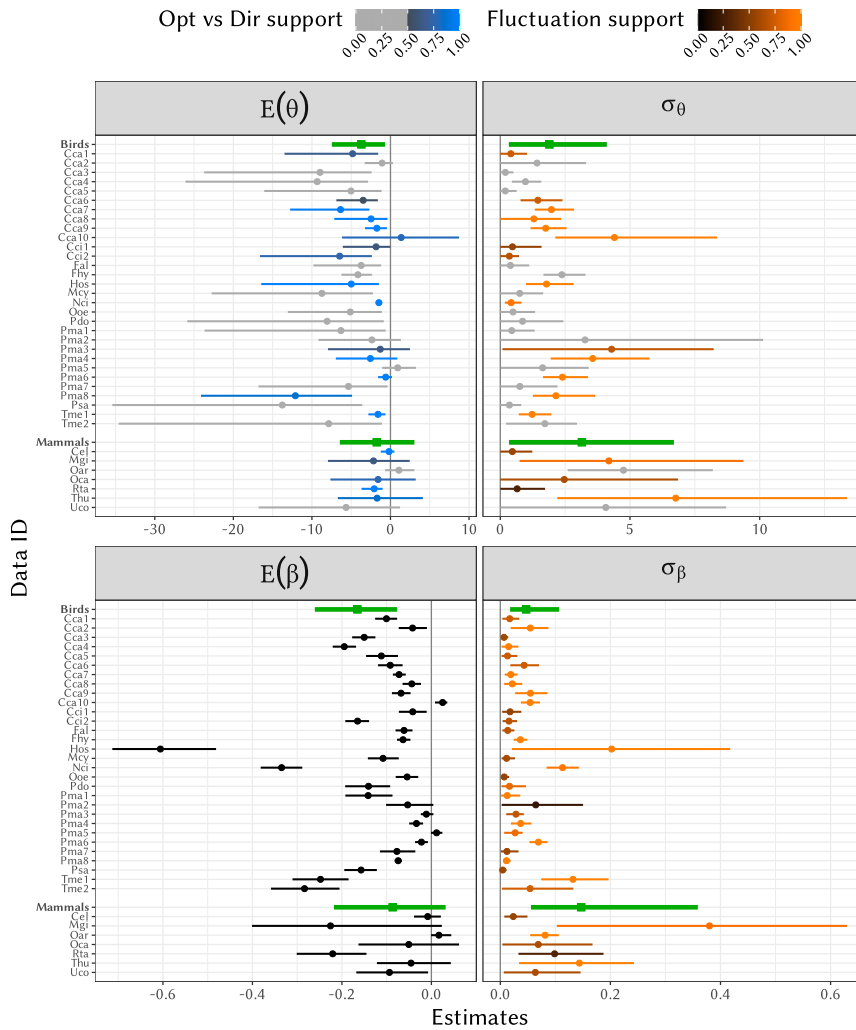
tion ( $\sigma_\theta < 0.5$ ; Fig. 2) and five inferred to have a large SD ( $\sigma_\theta > 3$ ; Fig. 2). Note that the latter also had  $E(\theta)$  not significantly different from zero, which could be linked to a greater uncertainty in the estimation of  $E(\theta)$  in the context of high levels of fluctuations. The meta-estimate for  $\sigma_\theta$  was higher for mammals (3.14, [0.34, 6.7]) than for birds (1.89, [0.33, 4.1]; Fig. 2). Interestingly, there was no obvious link between statistical support for fluctuations and the inferred SD of the optimum (orange scale in Fig. 2). Autocorrelation of the optimum was difficult to estimate, resulting in large 95% credible intervals overlapping zero most of the time ( $\varphi$  in *SI Appendix, Figs. S1 and S2, Bottom Left*). Still, six datasets had a significant estimate of temporal autocorrelation in the optimum, of which five were positive (blue tits, Cca7, 0.59[0.31, 0.84], Cca9, 0.42 [ $5.9 \times 10^{-4}$ , 0.80], Cca10, 0.94 [0.84, 0.99] and great tits, *Parus major*, Pma4, 0.74 [0.42, 0.97] and Pma8, 0.83 [0.64, 0.97], all from The Netherlands except Pma8). The only dataset with a significantly negative temporal autocorrelation was the hihi (Nci,  $-0.59[-0.98, -0.097]$ ). Overall, these differences between datasets resulted in a wide variation across datasets of the behavior of the fitness function over years (*SI Appendix, Fig. S3*).

**Selection Varies in Strength, but Not in Direction.** The inferred selection gradients  $\beta_t$  were consistent between models with and without an optimum (computed following refs. 40 and 50) for the same dataset (*SI Appendix, Fig. S4*), so we hereafter focus only on results from the model with an optimum to avoid overfitting resulting from model selection.

The temporal mean of the standardized selection gradient  $E(\beta)$  was significantly negative (selection for earlier breeding) for most bird datasets (only three great tit datasets, Pma2, Pma3, and Pma5 were not significantly negative; and one, a blue tit dataset, Cca10, was significantly positive; Fig. 2). On the contrary, the temporal mean gradients for mammals were mostly not significant (with two exceptions, the reindeer, Rta and the Columbian ground squirrel, *Urocitellus columbianus*, Uco; Fig. 2). The meta-estimates for the temporal mean of standardized gradient reflected these individual results, being significantly negative for birds ( $-0.17$ ,  $[-0.26, -0.077]$ ) but not for mammals ( $-0.087$ ,  $[-0.22, 0.032]$ ; Fig. 2). Six datasets (the European oystercatcher, Hos; eastern gray kangaroo, Mgi; hihi, Nci; the reindeer, Rta; and two Alpine swift datasets, Tme1 and Tme2) had stronger mean selection gradients than the others (Fig. 2). Interestingly, large mean selection gradients over years (large absolute values of  $E(\beta)$ ) were sometimes associated with predominant support for an optimum and were then attributable to a narrow fitness peak (small  $\omega$ ) rather than to a large temporal mean deviation from the optimum (large  $E(\theta)$ ; *SI Appendix, Fig. S5*).

The magnitude of variation in directional selection, as quantified by  $\sigma_\beta$ , was highly different between datasets, although less so than for  $\sigma_\theta$ . Overall, variation in standardized gradients ranged from very small to large (0.004 to 0.38 for the posterior medians of  $\sigma_\beta$ ), with meta-estimates at 0.047 ([0.018, 0.11]) for birds and 0.15 ([0.056, 0.36]) for mammals (Fig. 2). Despite such possibly large variation, there was very little evidence for fluctuations in the sign of selection gradients (e.g., negative gradients becoming positive [*SI Appendix, Fig. S6* ], and 49% of datasets with strong support for no change of sign at all), and such fluctuations were more frequent (posterior median above 30%) for datasets with an especially small average gradient ( $-0.04 < E(\beta) < 0.02$ ). Again, there was no link between statistical support in favor of fluctuations and the inferred  $\sigma_\beta$  (Fig. 2, levels of orange), which suggests that moderate variation in selection could still be strongly supported by the data.

**Plasticity Causes Adaptive Tracking of the Optimum Phenotype.** To better understand the causes of variation in directional selection, we disentangled the relative contributions of fluctuations in the



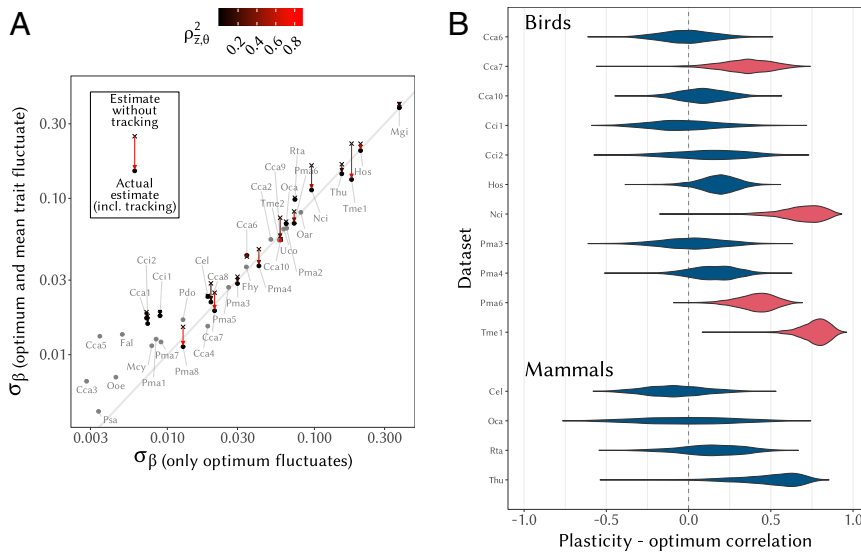
**Fig. 2.** Strength and variation of selection. The average location of the optimum  $E(\theta)$  (Top Left, where 0 represents the mean breeding time across years) and selection gradients  $E(\beta)$  (Bottom Left) are shown, together with their temporal SDs  $\sigma_\theta$  (Top Right) and  $\sigma_\beta$  (Bottom Right), for all datasets (points, posterior median; lines, 95% credible intervals). Meta-estimators for birds and mammals (computed on datasets with majority optimum support for Top Left and Top Right) are available at the bottom of each panel (in green, with squares and thicker lines). Note that the phenotypes were mean centered and scaled to a within-year variance of 1, so  $\theta$  and  $\beta$  are dimensionless. The evidence weight for an optimum (vs. directional models, excluding NoSel models) phenotype is indicated by a color on the blue scale (Top Left), while the orange scale (Top Right and Bottom Right) represents the evidence weight for fluctuating selection (more saturated colors for higher values, i.e., more support for the estimate). Datasets for which the optimum support was in the minority (<0.5) compared to directional models are grayed out in Top Left and Top Right. Estimates were computed from FluctCorrOpt models. The dataset codes are explained in SI Appendix, Table S1 and the values are provided in a CSV file on the GitHub repository.

optimum phenotype vs. in the mean phenotype (Fig. 1). From Eq. 2, the variance of selection gradients is

$$\sigma_\beta^2 = \frac{\sigma_\theta^2 + \sigma_{\bar{z}}^2 - 2\rho_{\bar{z},\theta}\sigma_\theta\sigma_{\bar{z}}}{(\omega^2 + 1)}. \quad [3]$$

Eq. 3 shows that the temporal variance in directional selection gradients  $\sigma_\beta^2$  results not only from fluctuations in the optimum, with variance  $\sigma_\theta^2$ , but also from year-to-year fluctuations in the annual mean phenotype  $\bar{z}$ , with variance  $\sigma_{\bar{z}}^2$ . Fluctuations in  $\bar{z}_t$  are explained by a combination of phenotypic plasticity (adaptive or not), responses to selection, and drift (neglecting the influence of dispersal). In addition,  $\sigma_\beta^2$  depends on the correlation  $\rho_{\bar{z},\theta}$  between the mean phenotype and the optimum (hereafter referred to as phenotypic tracking of the optimum). A positive  $\rho_{\bar{z},\theta}$  is indicative of adaptive change in the mean phenotype, as produced by adaptive phenotypic plasticity and/or genetic responses to natural selection.

The dots in Fig. 3A show the estimated standard deviations of selection gradients  $\sigma_\beta$ , plotted against their hypothetical values if we solely include fluctuations in the optimum, by assuming  $\sigma_{\bar{z}} = 0$  in the numerator of Eq. 3. Even for datasets with moderate or weak support for an optimum (gray dots), fluctuations of the optimum are a very good predictor of variation in selection gradients, as the points are close to the identity line (in light gray, which corresponds to the assumption that all variance in  $\beta$  originates from variance in the optimum  $\theta$ ). In cases where the optimum causes little variation in  $\beta$  (bottom left), the actual  $\sigma_\beta$  was inflated relative to this identity line. This inflation originates from mild fluctuations in the mean phenotype (with magnitude  $\sigma_{\bar{z}}$ ), which become nonnegligible relative to small values of  $\sigma_\theta$  and therefore contribute to variation in deviations from the optimum. The crosses in Fig. 3A show, for datasets with predominant support for an optimum, the hypothetical standard deviations of selection gradients in the absence of phenotypic tracking of the optimum, that is, keeping only  $\sigma_{\bar{z}}^2$  and  $\sigma_\theta^2$  in the numerator



**Fig. 3.** Phenotypic tracking of fluctuations in the optimum. (A) SD of the selection gradient  $\beta_t$  (dots, actual values  $\sigma_\beta$ ; crosses, computation assuming no tracking, i.e.,  $\rho_{z,\theta} = 0$  in Eq. 3) against the SD expected when using optimum fluctuations only (i.e.,  $\sigma_z = 0$  in Eq. 3). Arrows show the direction of the change when accounting for tracking, and the red scale indicates the actual value of  $\rho_{z,\theta}^2$ . Note that long arrows tend to be red, while short arrows tend to be gray. For datasets with minority support for an optimum compared to the directional models, only grayed-out dots are displayed. The identity line is depicted in gray. (B) For the 15 datasets with predominant support for an optimum and repeated measures, posterior distributions (coming from propagated Bayesian uncertainty) of the correlation coefficients between shifts in the optimum and shifts in the average phenology for individuals measured in 2 consecutive years. In light red: The distribution does not contain zero in the 95% highest density posterior interval. The dataset codes are explained in *SI Appendix, Table S1*.

of Eq. 3, while setting  $\rho_{z,\theta} = 0$ . The arrows connecting crosses to dots thus represent the influence of phenotypic tracking on variation in selection gradients: The longer the arrow, the more  $\rho_{z,\theta}$  becomes important to understand  $\sigma_\beta$  (Eq. 3). These arrows are pointing down in most cases, indicating that realized  $\sigma_\beta$  were smaller than expected when assuming independent fluctuations in the optimum and mean phenotype. The length of the downward-facing arrows can thus be interpreted as the degree to which temporal variation in selection was reduced by phenotypic tracking of the optimum causing a positive  $\rho_{z,\theta}$  (color of the arrows in Fig. 3A).

An obvious candidate mechanism for phenotypic tracking of the optimum is adaptive phenotypic plasticity (51, 52). Using only individuals with repeated measures in subsequent years (on a subset of 15 datasets with both predominant support for an optimum and sufficient repeated-individual data), we were able to distinguish plastic from genetic changes in mean breeding date. We detected plastic phenotypic tracking of fluctuations in the optimum (Fig. 3B), especially in four datasets for which the correlation between plastic phenotypic change and change in the optimum was significantly positive (in red in Fig. 3B; note that Cca7 and Pma6 are both located in Hoge Veluwe in The Netherlands). The meta-estimate of the correlation across the 11 bird datasets was relatively strong and significant for birds (0.25 [0.072, 0.44];  $P = 0.0095$ ), contrary to the meta-estimate across the 4 mammal datasets (0.13 [-0.17, 0.43];  $P = 0.35$ ). Note, however, that American red squirrel (*Tamiasciurus hudsonicus*, Thu) had a large correlation (0.53), which, despite being nonsignificant using a sample-based  $P$ -value ( $P = 0.0675$ ), had a 95% higher posterior density interval nonoverlapping zero ([0.056, 0.78]). These results suggest that phenotypic plasticity indeed plays an important role in tracking the optimum phenotype, at least in bird species.

### Discussion

We investigated fluctuations of fitness functions and temporal variation in selection, as estimated by the relationship between individual breeding date and yearly reproductive output. Our

unique database, comprising 39 datasets of wild populations of birds and mammals, allowed for an unprecedented estimation of parameters that appear in a wealth of theoretical predictions for adaptation to changing environments (11, 12, 16–18, 20–22, 25), answering our key questions laid out in the Introduction. In summary, we found predominant support for 1) models with a fitness peak against the alternatives and 2) fluctuations of the fitness function over time. This translated into 3) variation in the strength but not direction of selection, with a strong dependence on taxa (mammal/bird), species, and population. We found 4) uncertainty in the estimation of autocorrelation in the optimum and directional selection, owing to the high data requirements of these estimates. But we showed 5) substantial plastic phenotypic tracking of the optimum phenotype between years for bird species. Beyond our case study on reproductive phenology, the range of parameters we estimated here can serve as a much-needed benchmark of biologically realistic values for theoretical studies of adaptation to changing and fluctuating environments.

Our results corroborate a consensus in the bird literature that natural selection on phenology tends to favor earlier breeding (35), with a significantly negative meta-estimate for the directional selection gradients (Fig. 2). This pattern, which has been documented before (35, 39, 51, 53–60), was, however, not found in mammals overall, despite two individually significant datasets (Fig. 2), previously shown to be under such negative selection (61, 62). We also found support for the presence of an optimum phenotype (total statistical support of 54% for models with an optimum; Table 1), with slightly more support in mammals, perhaps in relation to the difference in significance of the selection gradient above. Support for an optimum is consistent with the intuition that breeding too early or too late should be detrimental in the temperate locations constituting most of our database, characterized by marked seasonality with stressful conditions in winter and summer (61, 62). This raises the question, especially for birds: Why are breeding dates in these populations not closer to their expected evolutionary equilibrium, instead displaying consistent deviations from their optimum? Among several possible explanations for this “paradox of stasis”

(63), a particularly relevant one for breeding time involves body condition (64). Nonheritable aspects of physiological condition (e.g., nutritional status) are known to influence both the timing of breeding and reproductive output, such that individuals in better condition tend to breed earlier and have more offspring (64). This causes the optimal breeding date to be displaced to a later time than the optimum set by the external environment (e.g., date of peak in resource abundance), such that apparent directional selection—mediated by condition—persists even at evolutionary equilibrium (64). Another mechanism with a similar outcome is when competition for breeding territories produces frequency-dependent selection favoring individuals that breed earlier than others in the population, regardless of the actual date (65). In that light, the difference between birds and mammals, in both the significance of mean selection gradients and support for an optimum, could stem from differences in how interindividual competition is happening over time, with possibly shorter periods of stronger competition when birds feed the chicks. Note that temporal variation in condition, or in its relationship with breeding date and reproductive success, could also contribute to the estimated variation in selection to some extent. A promising approach for partitioning out this effect would be to include a proxy for physiological condition in a multivariate selection analysis. More broadly speaking, trade-offs with other components of fitness not included in our estimate of selection, such as maternal survival or future performance (66), could also affect our inference of natural selection and its variation.

Our analysis indicates that the strength of natural selection on a phenological trait, one of the best-studied phenotypic categories in evolutionary ecology, varies in time in most investigated wild populations of birds and mammals (Fig. 2). Models including variation in the strength of selection and/or fluctuations of an optimum phenotype had statistical support above 75% (all taxa together; Table 1), and the SD of standardized selection gradients was relatively large, up to 0.38. However, we found little variation in the direction of selection, consistent with findings of a previous study based on a meta-analysis (31). Nevertheless, theoretical work has shown that randomly varying selection can have substantial eco-evolutionary impacts, even when the direction of selection does not fluctuate. Indeed, environmental stochasticity causes randomness in evolutionary trajectories, increasing both the average magnitude and stochastic variance of phenotypic mismatches with optimum, in turn leading to higher extinction probability in a novel or changing environment (20–22). These studies have shown that the demographic load (expressed as a reduction in log mean fitness) caused by a fluctuating optimum is proportional to  $\frac{\sigma_{\theta}^2}{2(\omega^2+1)}$  (for a SD-standardized trait), which we here estimate as 0.199 ( $[1.6 \times 10^{-5}, 0.99]$ ) for birds and 0.401 ( $[0.0067, 1.6]$ ) for mammals, equivalent to an 18% (respectively 33%) decrease in mean fitness.

Environmental fluctuations might not result in detectable variation in natural selection if populations track their fluctuating optimum over time. In datasets for which an optimum was well supported, we found that fluctuations in the optimum strongly influenced temporal variation in selection gradients (Fig. 3A), but that the latter was considerably attenuated by phenotypic tracking of the optimum. We demonstrated that this phenotypic tracking is largely caused by plastic responses of individuals that reproduce in consecutive years (Fig. 3B), with four datasets showing a significant correlation (from 0.36 to 0.78) between changes in the optimum and plastic change in the mean phenotype. A significant meta-estimate of this correlation was found for birds (no perfect tracking—correlation of 1—was detected, as would be expected) (67). The meta-estimate was not significant for the tested mammal datasets, which were mainly ungulates. Although difficult to generalize based on only four datasets, it is possible that because in mammals gestation periods are

often longer than for birds and annual fitness is often measured based on offspring recruitment (*SI Appendix, Table S1*), tracking selection through plasticity might be particularly challenging for mammals. An exception to this trend was the only nonungulate (American red squirrel, Thu), for which tracking was partially supported, consistent with previous findings in this species (23). It is possible that the natural history of this species—food hoarding (68) and year-round social cues of density (69)—provides access to cues of upcoming natural selection that are typically not available to other species.

Even when plastic phenotypic tracking was strong, the mean breeding time was consistently late relative to the optimum, thus questioning the adaptiveness of plasticity in these populations. Given that environmental cues strongly associated with phenological plasticity have been detected in all of the populations with substantial support for plastic tracking (60, 70–72), it is likely that such cues allow tracking of the optimum, but are somehow biased toward later phenology. A possible reason may be that the mean phenology is lagging behind an advancing optimum caused by warming climate and that the reaction norm for plasticity is shallower than that for the optimum (67, 73). For example, the significant positive autocorrelation signal observed in five of our datasets can be explained by a significant trend over years (without much impact on the estimate of  $\sigma_{\theta}$  for all five, but resulting in nonsignificant autocorrelation in two cases; *SI Appendix, Fig. S7*). Another possibility is that cue reliability has been reduced under climate change and habitat degradation, causing originally adaptive phenotypic plasticity to become less suitable for tracking the optimum phenotype. This scenario, which is predicted to cause evolution of the environmental cues used by organisms to plastically adjust their phenotypes (74), remains to be investigated further.

## Materials and Methods

**Data Collection.** We assembled a collection of surveys of wild populations for which episodes of fertility selection on reproductive phenology were monitored over multiple years, allowing estimation of parameters of fluctuating selection. To enter the database, a dataset had to include information on both 1) a trait relating to reproductive phenology, such as lay or parturition date, and 2) a measure of fitness for this selection episode, such as number of viable offspring or survival of offspring, which quantifies the output of a reproductive event. We also retained only datasets with a sufficiently large number of years (at least 9 y). The final collected database includes  $N_d = 39$  datasets, with 21 different species (13 birds and 8 mammals) and 32 different locations. The number of years varied between 9 and 63 (average 33.2) and the average number of females breeding per year between 15.7 and 236.3 (average 64.8) for a total of between 353 and 12,357 breeding events (average 1,880). More detailed information on each dataset is available in *SI Appendix, Table S1*.

**Data Formatting.** All datasets were formatted consistently. In the case of multiple breeding events per breeding season, we used the date of the first event as the phenological trait (onset of breeding); otherwise, we used the start date of the unique breeding event. For each dataset, this phenological trait was centered to the overall mean across years for the dataset and standardized by dividing by the average within-year phenotypic SD also for the dataset. As a measure of reproductive output for each female and breeding event, we used the number of fledglings summed over the entire breeding season for bird species and the number of offspring at weaning, or alive after 1 y, for mammals with large numbers of offspring. For mammals with one (occasionally two) offspring per breeding event, we used the survival to weaning or to 1 y after birth. Whether a dataset was using weaning or the 1-y threshold as the reference was decided in agreement with the contributors and is shown in *SI Appendix, Table S1*. All records with a missing value for either the phenological trait or the fitness measure were removed. A dummy identification (ID) was assigned for each record missing a female ID.

## Statistical Analyses.

**Fitness function.** Expanding on ref. 38, we contrasted three shapes of the fitness function relating the phenological trait to fitness in each breeding

season: 1) a flat function corresponding to no selection (“NoSel” model), 2) a monotonic function for which the direction of selection is independent of the mean phenotype (“Dir” models), and 3) a Gaussian optimum (“Opt” models). Denoting as  $W(z)$  the expected number of offspring of an individual with phenotype  $z$ , these fitness functions took the following mathematical forms when fitness consisted of a count of offspring:

$$1) \quad W(z) = \exp(a), \quad [4a]$$

$$2) \quad W(z) = \exp(a + bz), \quad [4b]$$

$$3) \quad W(z) = W_{\max} \exp\left(-\frac{(z - \theta)^2}{2\omega^2}\right). \quad [4c]$$

Note that for the exponential fitness function in Eq. 4b, the directional selection gradient is the parameter  $b$  (40), regardless of the phenotype distribution. For the Gaussian fitness peak in Eq. 4c, the parameter  $\omega$  describes the width of the fitness function, with smaller  $\omega$  causing stronger stabilizing selection, while  $\theta$  is the optimal timing for reproduction, and directional selection depends on the mean deviation from the optimum, as illustrated in Fig. 1. Since the phenological traits were standardized,  $\theta$  and  $\omega$  are in units of within-year phenotypic SD. When fitness measures consisted of survival of one offspring, we replaced the exponential in Eqs. 4a and 4b with an inverse-logit, while for Eq. 4c we retained the Gaussian fitness peak, but obtained  $W_{\max} \in [0, 1]$  from a continuous latent scale on real numbers via a logit link. The realized reproductive output was then obtained from this expected fitness using a Poisson or binomial distribution, depending on whether the fitness measures were a number or individual survival of offspring, respectively. The Poisson distribution could further be zero truncated or zero inflated, if posterior predictive checks on a Poisson model were showing a bad fit for the zero category. Furthermore, we included female IDs as a random effect on the intercept ( $a$  in Eqs. 4a and 4b and  $W_{\max}$  in Eq. 4c), to account for repeated measurements.

**Models of fluctuating selection.** To investigate temporally variable selection (“Fluct” models throughout, e.g., “FluctOpt” and “FluctDir”), we allowed the fitness function to vary from year to year, using random effects for time in the relevant parameters, as in refs. 38 and 39. For models with an optimum, a random effect for year was included for both  $W_{\max}$  and  $\theta$  (on the log or logit scale for  $W_{\max}$ ). We did not allow  $\omega$  to vary between years, because it is a difficult parameter to infer, and within-year sample sizes were likely not enough to bear with its estimation for each year. We can thus think of our estimates as fluctuations of an effective optimum with constant width, even though the true optimum may vary in width to some extent. For models without an optimum, we used random effects for years on the  $a$  and  $b$  parameters. The random effects (following a Gaussian distribution) allowed us to infer the SD over years of  $\theta$  and  $W_{\max}$  (on the log or logit scale),  $\sigma_\theta$  and  $\sigma_{W_{\max}}$ , and of  $a$  and  $b$ ,  $\sigma_a$  and  $\sigma_b$ . Models with only variation in the intercept ( $W_{\max}$  or  $a$ ) are referred to as “Const” models, because although the function varies in intercept from year to year, the actual selection process is assumed constant. Temporal autocorrelation, in the form of a first-order autoregressive process (AR1) with slope  $\varphi$ , was optionally introduced in the random effects for the  $\theta$  and  $b$  parameters (referred to as “FluctCorr” models).

The combination of fitness functions and patterns of fluctuations led to seven alternative parameterizations, which are summarized in Table 1. To compare the magnitude of selection and its fluctuation across models with alternative fitness functions, we computed the selection gradients  $\beta_t$  (estimated for each year  $t$  if fluctuations are assumed) from both kinds of statistical models with selection. For models with monotonic directional selection (ConstDir, FluctDir, FluctCorrDir), the selection gradient is simply the slope of the linear model  $\beta_t = b_t$  when using the log-link and was computed for logit-link as

$$\beta_t = b_t \left(1 - \frac{\overline{W}_t^2}{\overline{W}_t}\right), \quad [5]$$

where  $\overline{W}_t$  and  $\overline{W}_t^2$  are, respectively, the population mean fitness and mean squared fitness, computed over all available individuals each year, adapted from ref. 50. For models including an optimum, the directional selection gradient in year  $t$  is as in Eq. 2. Note that with an optimum, variation in directional selection gradients must account for year-to-year variation in the mean phenotype  $\bar{z}_t$  (Fig. 1).

**Prior distributions.** Diffuse, zero-centered normal distributions (with variance  $10^6$ ) were chosen as priors for  $\log(W_{\max})$ ,  $\theta$ ,  $a$ , and  $b$ , while for  $\text{logit}(W_{\max})$  in the binomial model, we used a weakly informative normal distribution with mean 0 and SD of 1. In contrast, we used a slightly

informed prior for  $\omega$ , because we do not expect the fitness peak to be narrow relative to the phenotypic SD since this would lead to extremely strong stabilizing selection, with most phenotypes having a fitness near zero, except in the immediate vicinity of the optimal timing for reproduction. We thus used a Gamma distribution parameterized so that 95% of the prior distribution lies between 1 and 10 SDs of the trait (standardized to 1), leading to a shape parameter of 3.36 and a rate parameter of 0.78. The variances of the random effects added to  $\log(W_{\max})$ ,  $a$ , and  $b$  were assigned a weakly informative standard normal distribution prior, while the prior variance of  $\sigma_\theta$  was specified indirectly via an independent exponential prior of rate 1 on  $c = \sigma_\theta/\omega$ . Finally, the zero-inflation probability  $p_{zi}$  was assigned a uniform prior between 0 and 1 and the autoregressive coefficient  $\varphi$  a uniform prior between  $-1$  and  $1$ .

**Statistical implementation.** We implemented the models using Hamiltonian Monte Carlo (HMC) as available in the Stan framework (75). We ran 10 chains, each with 2,000 iterations following a burn-in of 1,000 iterations. After a thinning every 5 iterations, we obtained a total of 4,000 iterations. Divergent transitions can happen during HMC and hamper safe interpretation of the output. Given the high number of models to be analyzed, we kept models with divergent transitions, although only if at low rates (less than 2.5% of the iterations), increasing the adapt.delta parameter in Stan as needed to reach this threshold. Convergence was checked graphically and using the potential scale reduction factor diagnostic (76). Effective sample size was kept above 200 for all parameters.

**Model selection.** The models were compared using a cross-validation procedure, namely approximate leave-one-out with Pareto smooth importance sampling (LOO-PSIS) (47). An information criterion can be derived from LOO-PSIS, named LOOIC, which was used to compare models. LOOIC is akin to the Widely Applicable Information Criterion (WAIC) (but does not rely on asymptotic assumptions) (47) and can be interpreted in a similar fashion to other information criteria such as AIC. To compute the overall statistical support, across datasets, for each model in Table 1, we derived weights of evidence inspired by Akaike weights used in model averaging (48), but based on LOOIC. The relative support for model  $i$  across datasets was defined as

$$w_i = \frac{1}{N_d} \sum_{j=1}^{N_d} \frac{\exp(-\Delta_{ij}/2)}{\sum_{k=1}^7 \exp(-\Delta_{kj}/2)}, \quad [6]$$

where  $\Delta_{ij}$  is the difference between the LOOIC of the best model and that of the focal model  $i$  ( $k$  iterates over the seven models), both for dataset  $j$ , and  $N_d$  is the total number of datasets as defined above. We repeated the same analysis using only birds and then only mammals datasets, adjusting  $N_d$  in Eq. 6 as needed.

This procedure of using weights of evidence was preferred over a simple computation of the proportion of datasets for which each model was the best model because the latter would necessarily be less precise. For instance, when several models (say, all those with fluctuating selection) have very similar LOOIC scores, but differ substantially from the remainder of the models for a given dataset (e.g., Cca1 in *SI Appendix, Table S2*), it is not particularly meaningful to select only the slightly best model; instead we want to measure how well each model is supported relative to all others. This is what  $w_i$  does: It attributes a score to each model, reflecting the relative support the model offers to the data, compared to other models.

**Post hoc analysis.** We computed the posterior distributions of the selection gradients  $\beta_t$  using the HMC samples of all parameters involved, to propagate uncertainty in these estimates toward the  $\beta_t$  estimates. To do that while accounting for uncertainty in estimating  $\bar{z}_t$  for models with an optimum (Eq. 2), we implemented a Monte Carlo sampling of the mean phenotype in each year, assuming a normal sampling distribution of the mean. We thus used the Monte Carlo and HMC samples of  $\bar{z}_t$ ,  $\theta_t$ , and  $\omega^2$  to propagate uncertainty in estimates of  $\beta_t$ . We then directly used estimates of  $\beta_t$  to compute the mean selection gradient  $E(\beta)$  and its SD over the years  $\sigma_\beta$ . Note that this strategy will cause a slight regression toward the mean and thus a slight underestimation of  $\sigma_\beta$  in general, but this is conservative with respect to the estimation of the prevalence and magnitude of fluctuating selection.

To obtain “meta-estimates” (i.e., robust overall estimates across all datasets, accounting for different uncertainties between datasets), we generated 100 tables (each composed of one row for each dataset), drawing from the posterior samples of  $E(\theta)$ ,  $\sigma_\theta$ ,  $E(\beta)$ ,  $\sigma_\beta$ , and  $\omega$ . We used the multiple-imputation framework of the R package *brms* (77) to perform a mixed-model analysis of each of these parameters using the taxon (bird or mammal) as a fixed effect and species and population as random effects.



We used the taxon-level intercepts of such models as the meta-estimates and report their posterior median and 95% credible interval. For  $E(\theta)$ ,  $\sigma_\theta$ , and  $\omega$ , we used only datasets with a majority statistical support for optimum models, compared to directional models.

To study the influence of phenotype optimum tracking by plastic responses at the individual level, we selected individuals that reproduced in two consecutive years and computed the difference in average phenology between years in this subset (again, using Monte Carlo simulations to account for uncertainty thereafter). We retained only datasets with at least five individuals in common between consecutive years, for at least 10 y in total, and with a majority statistical support for an optimum. Although proper measurement of phenotypic plasticity requires data about an environmental cue that induces the plastic response, the phenotypic change caused by plasticity (i.e., the plastic response) can be inferred accurately without this information provided that other processes such as ontogeny, habitat choice, or senescence can be ignored. This assumption is generally a good approximation for phenological traits and was used for instance by ref. 78 to estimate selection on plasticity, even though there is some evidence for senescence of reproductive phenology and its plasticity in the wild (see ref. 79 for an example on blue tits). We then computed the correlation between plastic changes in mean individual phenotype and changes in optimum phenotype across years, still accounting for uncertainty. To test for the significance of an overall trend in these correlations, we sampled Monte Carlo and HMC iterations amounting to the sample size of each dataset and did so 100 times. We then inferred the meta-estimate of the correlation using a mixed model in brms, as described above, using taxon as a fixed effect and study ID as a random effect.

**Data Availability.** Estimates, code, and data to reproduce the analysis can be found in Github at <https://github.com/devillemereuil/MetaFluctSel> (80).

- P. Inchausti, J. Halley, The long-term temporal variability and spectral color of animal populations. *Evol. Ecol. Res.* **4**, 1033–1048 (2002).
- P. R. Grant, B. R. Grant, Unpredictable evolution in a 30-year study of Darwin's finches. *Science* **296**, 707–711 (2002).
- R. Lande, S. Engen, B.-E. Saether, *Stochastic Population Dynamics in Ecology and Conservation* (Oxford University Press, 2003).
- D. A. Vasseur, P. Yodzis, The color of environmental noise. *Ecology* **85**, 1146–1152 (2004).
- M. R. Robinson, J. G. Pilkington, T. H. Clutton-Brock, J. M. Pemberton, L. E. B. Kruuk, Environmental heterogeneity generates fluctuating selection on a secondary sexual trait. *Curr. Biol.* **18**, 751–757 (2008).
- G. Bell, Fluctuating selection: The perpetual renewal of adaptation in variable environments. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 87–97 (2010).
- T. M. L. Wigley, R. L. Smith, B. D. Santer, Anthropogenic influence on the autocorrelation structure of hemispheric-mean temperatures. *Science* **282**, 1676–1679 (1998).
- G. J. Boer, Changes in interannual variability and decadal potential predictability under global warming. *J. Clim.* **22**, 3098–3109 (2009).
- J. Felsenstein, The theoretical population genetics of variable selection and migration. *Annu. Rev. Genet.* **10**, 253–280 (1976).
- P. W. Hedrick, Genetic variation in a heterogeneous environment. I. Temporal heterogeneity and the absolute dominance model. *Genetics* **78**, 757–770 (1974).
- J. J. Bull, Evolution of phenotypic variance. *Evolution* **41**, 303–315 (1987).
- J. Tufto, Genetic evolution, plasticity, and bet-hedging as adaptive responses to temporally autocorrelated fluctuating selection: A quantitative genetic model. *Evolution* **69**, 2034–2049 (2015).
- A. G. Jones, S. J. Arnold, R. Bürger, The mutation matrix and the evolution of evolvability. *Evolution* **61**, 727–745 (2007).
- B. Charlesworth, Directional selection and the evolution of sex and recombination. *Genet. Res.* **61**, 205–224 (1993).
- R. Bürger, Evolution of genetic variability and the advantage of sex and recombination in changing environments. *Genetics* **153**, 1055–1069 (1999).
- S. Gavrillets, S. M. Scheiner, The genetics of phenotypic plasticity. VI. Theoretical predictions for directional selection. *J. Evol. Biol.* **6**, 49–68 (1993).
- R. Lande, Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. *J. Evol. Biol.* **22**, 1435–1446 (2009).
- C. A. Botero, F. J. Weissing, J. Wright, D. R. Rubenstein, Evolutionary tipping points in the capacity to adapt to environmental change. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 184–189 (2015).
- J. M. Smith, What determines the rate of evolution? *Am. Nat.* **110**, 331–338 (1976).
- M. Lynch, R. Lande, "Evolution and extinction in response to environmental-change" in *Workshop on Biotic Interactions and Global Change*, P. M. Kareiva, J. G. Kingsolver, R. B. Huey, Eds. (Sinauer Associates, Sunderland, MA, 1993), pp. 234–250.
- R. Lande, S. Shannon, The role of genetic variation in adaptation and population persistence in a changing environment. *Evolution* **50**, 434–437 (1996).
- L. M. Chevin, O. Cotto, J. Ashander, Stochastic evolutionary demography under a fluctuating optimum phenotype. *Am. Nat.* **190**, 786–802 (2017).
- A. G. McAdam, S. Boutin, B. Dantzer, J. E. Lane, Seed masting causes fluctuations in optimum litter size and lag load in a seed predator. *Am. Nat.* **194**, 574–589 (2019).
- P. D. Gingerich, Rates of evolution: Effects of time and temporal scaling. *Science* **222**, 159–161 (1983).
- S. Estes, S. J. Arnold, Resolving the paradox of stasis: Models with stabilizing selection explain evolutionary divergence on all timescales. *Am. Nat.* **169**, 227–244 (2007).
- J. C. Uyeda, T. F. Hansen, S. J. Arnold, J. Pienaar, The million-year wait for macroevolutionary bursts. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 15908–15913 (2011).
- M. Kopp, S. Matuszewski, Rapid evolution of quantitative traits: Theoretical perspectives. *Evol. Appl.* **7**, 169–191 (2014).
- T. E. Reed, R. S. Waples, D. E. Schindler, J. J. Hard, M. T. Kinnison, Phenotypic plasticity and population viability: The importance of environmental predictability. *Proc. Biol. Sci.* **277**, 3391–3400 (2010).
- L. M. Chevin, B. C. Haller, The temporal distribution of directional gradients under selection for an optimum. *Evolution* **68**, 3381–3394 (2014).
- A. M. Siepielski, J. D. DiBattista, S. M. Carlson, It's about time: The temporal dynamics of phenotypic selection in the wild. *Ecol. Lett.* **12**, 1261–1276 (2009).
- M. B. Morrissey, J. D. Hadfield, Directional selection in temporally replicated studies is remarkably consistent. *Evolution* **66**, 435–442 (2012).
- C. Parmesan, G. Yohe, A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**, 37–42 (2003).
- M. B. Davis, R. G. Shaw, J. R. Etterson, Evolutionary responses to changing climate. *Ecology* **86**, 1704–1714 (2005).
- C. Parmesan, Ecological and evolutionary responses to recent climate change. *Annu. Rev. Ecol. Syst.* **37**, 637–669 (2006).
- V. Radchuk et al., Adaptive responses of animals to climate change are most likely insufficient. *Nat. Commun.* **10**, 3109 (2019).
- C. J. Tansey, J. D. Hadfield, A. B. Phillimore, Estimating the ability of plants to plastically track temperature-mediated shifts in the spring phenological optimum. *Global Change Biol.* **23**, 3321–3334 (2017).
- 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).
- L. M. Chevin, M. E. Visser, J. Tufto, Estimating the variation, autocorrelation, and environmental sensitivity of phenotypic selection. *Evolution* **69**, 2319–2332 (2015).
- M. Gamelon et al., Environmental drivers of varying selective optima in a small passerine: A multivariate, multiphasic approach. *Evolution* **72**, 2325–2342 (2018).
- M. Morrissey, I. B. J. Goudie, Analytical results for directional and quadratic selection gradients for log-linear models of fitness functions. bioRxiv:040618 (22 February 2016).
- R. Lande, S. J. Arnold, The measurement of selection on correlated characters. *Evolution* **37**, 1210–1226 (1983).
- J. G. Kingsolver et al., The strength of phenotypic selection in natural populations. *Am. Nat.* **157**, 245–261 (2001).
- H. E. Hoekstra et al., Strength and tempo of directional selection in the wild. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9157–9160 (2001).
- R. Lande, Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30**, 314–334 (1976).

45. B. D. H. Latter, Selection in finite populations with multiple alleles. II. Centripetal selection, mutation, and isoallelic variation. *Genetics* **66**, 165–186 (1970).
46. R. Bürger, *The Mathematical Theory of Selection, Recombination, and Mutation* (John Wiley & Sons, Chichester, UK, 2000).
47. A. Vehtari, A. Gelman, J. Gabry, Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432 (2017).
48. K. P. Burnham, D. R. Anderson, Multimodel inference understanding AIC and BIC in model selection. *Sociol. Methods Res.* **33**, 261–304 (2004).
49. T. Johnson, N. Barton, Theoretical models of selection and mutation on quantitative traits. *Philos. Trans. Biol. Sci.* **360**, 1411–1425 (2005).
50. J. Janzen, H. S. Stern, Logistic regression for empirical studies of multivariate selection. *Evolution* **52**, 1564–1571 (1998).
51. A. Charmantier *et al.*, Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science* **320**, 800–803 (2008).
52. M. E. Visser, S. P. Caro, K. van Oers, S. V. Schaper, B. Helm, Phenology, seasonal timing and circannual rhythms: Toward a unified framework. *Philos. Trans. Biol. Sci.* **365**, 3113–3127 (2010).
53. A. Van Noordwijk, R. McCleery, C. Perrins, Selection for the timing of great tit breeding in relation to caterpillar growth and temperature. *J. Anim. Ecol.* **64**, 451–458 (1995).
54. 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 Biol. Sci.* **265**, 1867–1870 (1998).
55. B. C. Sheldon, L. E. B. Kruuk, J. Merilä, Natural selection and inheritance of breeding time and clutch size in the collared flycatcher. *Evolution* **57**, 406–420 (2003).
56. P. Gienapp, E. Postma, M. E. Visser, Why breeding time has not responded to selection for earlier breeding in a songbird population. *Evolution* **60**, 2381–2388 (2006).
57. C. Teplitsky, J. A. Mills, J. W. Yarrall, J. Merilä, Indirect genetic effects in a sex-limited trait: The case of breeding time in red-billed gulls. *J. Evol. Biol.* **23**, 935–944 (2010).
58. T. Pärt, J. Knappe, M. Low, M. Öberg, D. Arlt, Disentangling the effects of date, individual, and territory quality on the seasonal decline in fitness. *Ecology* **98**, 2102–2110 (2017).
59. P. M. Sirkkiä *et al.*, Climate-driven build-up of temporal isolation within a recently formed avian hybrid zone. *Evolution* **72**, 363–374 (2018).
60. P. de Villemereuil, A. Rutschmann, J. G. Ewen, A. W. Santure, P. Brekke, Can threatened species adapt in a restored habitat? No expected evolutionary response in lay date for the New Zealand hihi. *Evol. Appl.* **12**, 482–497 (2019).
61. J. E. Lane, L. E. B. Kruuk, A. Charmantier, J. O. Murie, F. S. Dobson, Delayed phenology and reduced fitness associated with climate change in a wild hibernator. *Nature* **489**, 554–557 (2012).
62. H. Holand *et al.*, Stabilizing selection and adaptive evolution in a combination of two traits in an arctic ungulate. *Evolution* **74**, 103–115 (2020).
63. J. Merilä, B. Sheldon, L. Kruuk, Explaining stasis: Microevolutionary studies in natural populations. *Genetica* **112**, 199–222 (2001).
64. T. Price, M. Kirkpatrick, S. Arnold, Directional selection and the evolution of breeding date in birds. *Science* **240**, 798–799 (1988).
65. J. Johansson, H. G. Smith, N. Jonzén, Adaptation of reproductive phenology to climate change with ecological feedback via dominance hierarchies. *J. Anim. Ecol.* **83**, 440–449 (2014).
66. D. Schluter, T. D. Price, L. Rowe, P. R. Grant, Conflicting selection pressures and life history trade-offs. *Proc. R. Soc. Lond. B Biol. Sci.* **246**, 11–17 (1991).
67. P. Gienapp, T. E. Reed, M. E. Visser, Why climate change will invariably alter selection pressures on phenology. *Proc. Biol. Sci.* **281**, 20141611 (2014).
68. S. Boutin *et al.*, Anticipatory reproduction and population growth in seed predators. *Science* **314**, 1928–1930 (2006).
69. B. Dantzer *et al.*, Density triggers maternal hormones that increase adaptive offspring growth in a wild mammal. *Science* **340**, 1215–1217 (2013).
70. M. E. Visser, L. J. M. Holleman, P. Gienapp, Shifts in caterpillar biomass phenology due to climate change and its impact on the breeding biology of an insectivorous bird. *Oecologia* **147**, 164–172 (2006).
71. S. Bonamour, L. M. Chevin, A. Charmantier, C. Teplitsky, Phenotypic plasticity in response to climate change: The importance of cue variation. *Philos. Trans. R. Soc. B* **374**, 20180178 (2019).
72. L. D. Bailey *et al.*, Bird populations most exposed to climate change are less responsive to climatic variation. [bioRxiv:2020.08.16.252379](https://doi.org/10.1101/2020.08.16.252379) (16 August 2020).
73. M. E. Visser, Keeping up with a warming world; assessing the rate of adaptation to climate change. *Proc. Biol. Sci.* **275**, 649–659 (2008).
74. R. Lande, L. M. Chevin, Evolution of environmental cues for phenotypic plasticity. *Evolution* **69**, 2767–2775 (2015).
75. M. D. Hoffman, A. Gelman, The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15**, 1593–1623 (2014).
76. A. Vehtari, A. Gelman, D. Simpson, B. Carpenter, P. C. Bürkner, Rank-normalization, folding, and localization: An improved R for assessing convergence of MCMC. [arXiv:1903.08008](https://arxiv.org/abs/1903.08008) Stat (19 March 2019).
77. P. C. Bürkner, Advanced Bayesian multilevel modeling with the R package brms. [arXiv:1705.11123](https://arxiv.org/abs/1705.11123) Stat (31 May 2017).
78. J. E. Brommer, E. Kløen, Exploring the genetics of nestling personality traits in a wild passerine bird: Testing the phenotypic gambit. *Ecol. Evol.* **2**, 3032–3044 (2012).
79. S. Bonamour, L.-M. Chevin, D. Réale, C. Teplitsky, A. Charmantier, Age-dependent phenological plasticity in a wild bird. *J. Anim. Ecol.* **89**, 2733–2741 (2020).
80. de Villemereuil *et al.*, Code and data for “Fluctuating optimum and temporally variable selection on breeding date in birds and mammals.” Github. <https://github.com/devillemereuil/MetaFluctSel>. Deposited 24 April 2020.