JOURNAL OF Evolutionary Biology

doi: 10.1111/jeb.13246

Fluctuating selection and its (elusive) evolutionary consequences in a wild rodent population

T. BONNET*† & E. POSTMA†‡

*Research School of Biology, ANU College of Science, The Australian National University, Acton, ACT, Australia
†Department of Evolutionary Biology and Environmental Studies (IEU), University of Zurich, Zurich, Switzerland
‡Centre for Ecology and Conservation, University of Exeter, College of Life and Environmental Sciences, Penryn, Cornwall, UK

Keywords:

adaptation;
mammals;
natural selection;
quantitative genetics;
Individual-based;
simulation.

Abstract

Temporal fluctuations in the strength and direction of selection are often proposed as a mechanism that slows down evolution, both over geological and contemporary timescales. Both the prevalence of fluctuating selection and its relevance for evolutionary dynamics remain poorly understood however, especially on contemporary timescales: unbiased empirical estimates of variation in selection are scarce, and the question of how much of the variation in selection translates into variation in genetic change has largely been ignored. Using long-term individual-based data for a wild rodent population, we quantify the magnitude of fluctuating selection on body size. Subsequently, we estimate the evolutionary dynamics of size and test for a link between fluctuating selection and evolution. We show that, over the past 11 years, phenotypic selection on body size has fluctuated significantly. However, the strength and direction of genetic change have remained largely constant over the study period; that is, the rate of genetic change was similar in years where selection favoured heavier vs. lighter individuals. This result suggests that over shorter timescales, fluctuating selection does not necessarily translate into fluctuating evolution. Importantly however, individual-based simulations show that the correlation between fluctuating selection and fluctuating evolution can be obscured by the effect of drift, and that substantially more data are required for a precise and accurate estimate of this correlation. We identify new challenges in measuring the coupling between selection and evolution, and provide methods and guidelines to overcome them.

Introduction

Selection, the causal covariation between trait values and fitness, shapes biodiversity in time and space and explains the general match between organisms and their environment (Darwin, 1859; Endler, 1986). Linking the sources of natural and sexual selection to the dynamics of evolution (defined here as a change in mean breeding value for the trait of interest) has been a major goal of evolutionary biology during the last century (e.g. Fisher, 1958), but for most of the 20th century progress has been hampered by the lack of a unified framework to quantify selection (Wade, 2006). This changed with the development of regression-based methods to measure the strength and direction of selection (Lande, 1979; Lande & Arnold, 1983), which have enabled the estimation of selection gradients in a large variety of traits and biological systems (Kingsolver *et al.*, 2001; Stinchcombe *et al.*, 2008). This bonanza of estimates has shown that directional selection is stronger and more common than stabilizing selection, for both morphological and life-history traits (Kingsolver *et al.*, 2001; Hereford *et al.*, 2004; Hendry, 2017). At first sight, this pattern is contrary to expectations (Kingsolver & Diamond, 2011): as most traits are heritable (Mousseau & Roff, 1987; Postma, 2014), they are

Correspondence: Timothée Bonnet, Research School of Biology, 46 Sullivans Creek Road, The Australian National University, Acton, ACT 2601, Australia.e-mail: timotheebonnetc@gmail.com

predicted to evolve towards their fitness optimum, with directional selection progressively being replaced by stabilizing selection. In practice, however, most traits evolve only very slowly and within a limited phenotypic range (Hendry & Kinnison, 1999; Merilä *et al.*, 2001; Brookfield, 2016).

One explanation for this paradox is that fitness landscapes are not constant over time, and populations are evolving towards a continuously changing fitness optimum (Fisher & Ford, 1947; Lande, 1976). Whereas at any point in time directional selection may be strong, average selection gradients may be weaker, and if selection fluctuates not only in strength but also in direction, average selection may even be zero (Fig. 1a–c). Fluctuating selection may thus slow down longer-term evolutionary adaptation, or even bring it to a halt (Jones *et al.*, 2004; Estes & Arnold, 2007), and it thereby constitutes an appealing explanation for the commonly observed lack of evolutionary change, that is evolutionary stasis, as well as for the commonness of directional selection (Merilä *et al.*, 2001; Robinson *et al.*, 2008; Bell, 2010). However, although fluctuating selection as an explanation for 'macro-evolutionary' stasis is gaining theoretical and empirical support (Estes & Arnold, 2007; Uyeda *et al.*, 2011; Voje *et al.*, 2015), our understanding of the importance of fluctuations in selection in shaping the evolutionary dynamics of natural populations on a much shorter timescale, for example from one generation to the next, is still limited. A few robust examples of temporal variation in selection exist (e.g. Grant & Grant, 2002; Husby *et al.*, 2011; Bergland *et al.*, 2014; Milesi *et al.*, 2016), but an assessment of the general micro-evolutionary relevance of fluctuating selection is hampered by the lack of a clear answer to two questions: (i) Does phenotypic selection commonly fluctuate, in strength and/or direction? (Hendry, 2017, pp. 47–51) (ii) And if it does, do these fluctuations translate into fluctuations, in speed and/or direction, of genetic change?

The first question seemingly received a positive answer with the publication of a synthetic review of

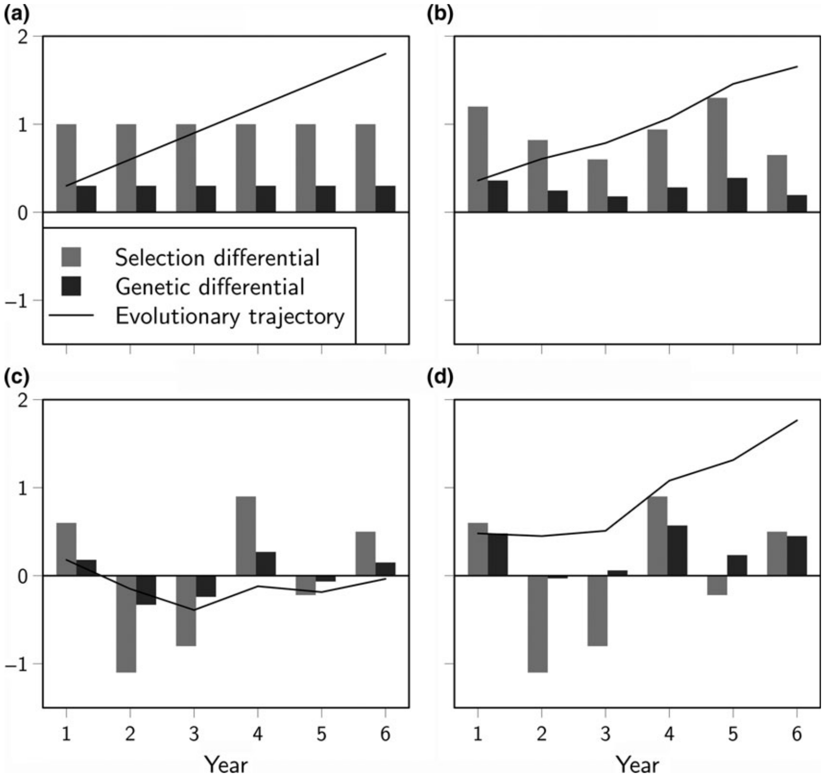

Fig. 1 Evolutionary change under constant and fluctuating selection regimes. In (a), selection is constant across years. Following the breeder's equation, the change in breeding values (i.e. genetic differential or the response to selection) is equal to the product of the selection differential and the narrow-sense heritability, which is here set to 0.3. The resultant cumulative response to selection, that is the evolutionary trajectory, is described by a straight line. In (b), selection fluctuates but does not reverse, and mean selection and the rate of evolution are only slightly reduced compared to (a). In (c), selection fluctuates and reverses, resulting in fluctuating and reversing evolution, and thereby evolutionary stasis over the time frame considered. In (d), selection fluctuates and reverts as in (c), but selection is partly noncausal and mediated by an unobserved environmental factor (i.e. a key assumption of the breeder's equation is violated). As a consequence, selection and evolution are uncoupled, and despite fluctuating selection, the rate of evolution is similar to (a).

temporal replicates of selection from 89 studies, which concluded that phenotypic selection does indeed vary and reverse its direction among years (Siepielski *et al.*, 2009). However, Morrissey & Hadfield (2012) showed that most of these fluctuations can be ascribed to sampling variation, that is, the stochasticity that causes the realized value of a parameter to differ from the parameter of the data-generating process in finite populations. When sampling variation is accounted for, directional selection is in fact remarkably constant over time, both in magnitude and direction: instead of estimating the variance of the distribution of temporal estimates of selection, as in Siepielski *et al.* (2009), tests for fluctuating selection must estimate the variance of the temporal distribution of selection (Morrissey & Hadfield, 2012). As of yet, Chevin *et al.* (2015) are among the few to have done this: using random regression mixed models which explicitly estimate the variance in selection gradients, they found that phenotypic selection on laying date fluctuated over a short time period in a population of great tits (*Parus major* Linnaeus, 1758). The generality of this finding, however, needs to be confirmed across a wider range of species, populations and traits, using the same, statistically robust approach.

In addition to showing statistically significant variation in selection, two more points must be addressed to assess the evolutionary relevance of fluctuating selection. First, the precise pattern of fluctuation matters: even in the presence of fluctuating selection, evolution will only come to a halt if the direction of selection changes regularly and the mean selection differential equals zero (Blanckenhorn, 2000; Hunt *et al.*, 2004; Morrissey & Hadfield, 2012; see Fig. 1b). Second, phenotypic selection, when defined as a nonzero phenotypic covariance between a trait and relative fitness, does not necessarily lead to an evolutionary response (see Fig. 1d). The breeder's equation assumes that fitness covaries with phenotypic variation blindly, and does not distinguish between whether this phenotypic variation is the result of genetic or nongenetic variation (Rausher, 1992). When this assumption is violated and apparent selection is disproportionately dominated by an environmental covariance between the trait of interest and fitness, estimates of phenotypic selection provide a poor predictor of genetic change (Price & Liou, 1989; Rausher, 1992; Morrissey *et al.*, 2010; Bonnet *et al.*, 2017). For instance, random infection of some individuals by a parasite may simultaneously drive among-individual variation in mass and variation in fitness, generating a covariation between mass and fitness, without mass causally affecting fitness.

Whereas the latter is one of the general explanations for apparent evolutionary stasis, it is particularly relevant within the context of fluctuating selection: as fluctuating selection is often thought to be driven by environmental fluctuations (Bell, 2010; Chevin & Haller, 2014), these may disproportionately shape (fluctuations in) the environmental component of selection. Fluctuations in the additive genetic covariance between the trait and fitness, that is in fluctuating evolution (Robertson, 1966; Price, 1970; Morrissey *et al.*, 2012), can result from fluctuating selection only if the fluctuations involve the causal effects of the focal trait on fitness.

Here we take advantage of the eleven-year-long monitoring of a population of snow voles (*Chionomys nivalis* Martins, 1842) to (i) quantify fluctuating selection on body size, (ii) describe the temporal dynamics of evolution in size and (iii) quantify the relationship between fluctuating selection and evolution. To this end, we first estimate directional selection on a year-to-year basis to quantify the variation in selection estimates. We then explicitly model these fluctuations of directional selection within a mixed model to account for sampling variance. Based on the sign of annual selection estimates, as well as on the ratio of the variance in selection over the mean strength of selection, we also assess the probability of selection reversal. These analyses are performed for total selection, as well as for fertility and viability selection separately. Second, we use a quantitative genetic framework to describe the general pattern of evolution over the study period and estimate the rate of evolution of size on a year-to-year basis. Third, we combine analyses of selection and estimates of evolutionary change to assess the coupling between variation in the strength and sign of selection and evolution. Finally, we perform a series of individual-based simulations to infer the statistical power of our test for fluctuating selection and its evolutionary relevance, which is crucial when it comes to interpreting any negative results.

Materials and methods

Study population

From 2006 to 2016, a wild population of snow voles (*Chionomys nivalis* Martins, 1842) has been monitored intensively. This population, which consists of 80–230 individuals (Table 1), is located in the Swiss Alps, near Chur (N46°48′, E9°34′; 2030 m.a.s.l.). The study area consists of 5 ha of scree with sparse vegetation, surrounded by meadows, forest and a steep cliff. Because the snow vole shows an overwhelming preference for rocky environments (Janeau & Aulagnier, 1997; Luque-larena *et al.*, 2002), the monitored population is ecologically fairly isolated. Nevertheless, it receives on average 8.6 immigrants per year (Table 1, see also García-Navas *et al.*, 2015).

Snow voles are live-trapped during two to five trapping sessions taking place between late May and early October. To this end, the study area is overlaid with a 10 × 10 m grid consisting of a total of 559 cells with stable geographic coordinates. A trapping session

Table 1 Number of phenotyped individuals, survivors to the next year and number of immigrants.

Year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Phenotyped individuals	183	193	139	163	131	56	66	116	130	118	128
Number of adults	64	66	62	46	69	36	32	40	52	59	65
Number of juveniles	112	126	75	103	59	15	34	75	77	55	63
Number of survivors	43	39	33	48	16	8	21	38	31	20	–
Number of immigrants	52	13	13	14	3	9	9	8	11	5	1

The number of phenotyped adults and juveniles includes all individuals with at least one body mass index measurement in a given year, that is with a measurement of both body mass and body length. This represents the sample size for the selection analyses based on total fitness (F) and viability (ϕ). The number of survivors to the next year represents the sample size for the selection analysis based on fertility (ρ) and is unknown for 2016. Immigrants are individuals with unknown parents and are counted only in the first year they appeared in the population. In 2006, the number of immigrants represents the size of the base population, whereas in other years the number of immigrants represents individuals immigrating in the population.

consists of four trapping nights, necessary to cover all four quarters of the study area. During each trapping session, a Longworth trap (catch-and-release trap, Penlon Ltd, Oxford, UK) filled with hay and baited with apple, hamster food and peanut butter is placed in every cell. Individuals captured for the first time are ear-clipped (2 mm diameter, thumb type punch, Harvard Apparatus, Holliston, MA, USA) and individually marked with a subcutaneous PIT tag (ISO transponder, Tierchip Dasmann, Tecklenburg, Germany). Ear clips are preserved in 95% ethanol + 5% TE. For each capture, we record individual identity, geographic coordinates, body mass, body length, tail length, sex and age.

Ear clips are stored at −20 °C until DNA extraction. All individuals are genotyped for 18 autosomal microsatellites using snow vole-specific primers (Wandeler *et al.*, 2008; García-Navas *et al.*, 2015). In addition, the sex of all individuals is confirmed by sequencing the *Sry* locus (Gubbay *et al.*, 1990; Wandeler *et al.*, 2008). Finally, the mitochondrial control region is sequenced, and all males are genotyped for one Y-linked microsatellite and three Y-linked insertion-deletions (Wandeler & Camenisch, 2011). Based on the autosomal microsatellite genotypes, we reconstruct the pedigree of the population using the maximum likelihood-based program COLONY (Wang, 2004; Jones & Wang, 2010) and the Bayesian R package MasterBayes (Hadfield *et al.*, 2006; R Core Team, 2015). The pedigree is then checked for consistency using the Y-linked markers and the mitochondrial haplotypes. This procedure allows the identification of most of the parental links (91%) as well as the identification of likely immigrants (individuals first captured as adults and with two unknown parents). This well-resolved pedigree is used to define annual and lifetime reproductive success, as well as to estimate the relatedness among all pairs of individuals.

A mark-recapture analysis has shown that between-session recapture probabilities are very high (adults: 92.4% ± 1.1; juveniles: 81.1% ± 3.0). Therefore, the between-year recapture probability is effectively 1, and the noncapture of an individual in a given year can be directly equated with death or permanent emigration without the need for mark-recapture modelling.

Fitness measures

We considered three measures of fitness: (i) survival from 1 year to the next, $\phi_{i,t}$, based on whether an individual i observed in year t is observed again in year $t + 1$ ($\phi_{i,t} = 1$) or not ($\phi_{i,t} = 0$); (ii) annual reproductive success, $\rho_{i,t+1}$, the number of juveniles born from i during the year $t + 1$, that is, after i survives to the next year (but irrespective of juvenile survival); (iii) an annualized measure of overall fitness, similar to that used in Qvarnström *et al.* (2006), $F_{i,t} = 2\phi_{i,t}+\rho_{i,t+1}$. $F_{i,t}$ is an appropriate measure of fitness in the context of studying evolution with overlapping generations because it captures the production of all the individuals present in year $t + 1$ by all the individuals present in year t (Fig. 2). In our measure of total fitness ($F_{i,t}$),

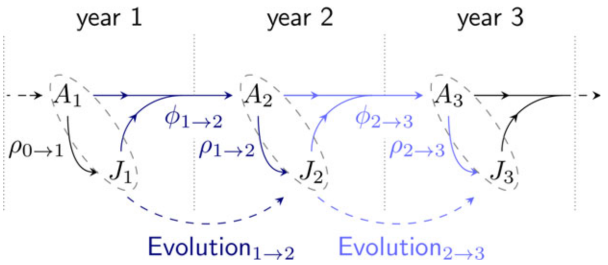

Fig. 2 Schematic representation of the snow vole life cycle, and of our definition of reproduction, survival and evolution. We are interested in predicting Evolution$_{t\to t+1}$, the genetic difference (i.e. the difference in mean breeding value) between all individuals present in year t [adults (A_t) and juveniles (J_t)] and all individuals present in year $t + 1$ (A_{t+1} and J_{t+1}). This genetic change is a response to viability selection from year t to year $t + 1$ ($\phi_{t\to t+1}$) and to fertility selection during year $t + 1$ ($\rho_{t\to t+1}$). Three years and two transitions are depicted. The colour (dark blue or light blue) shows which fitness components predict which evolutionary change.

survival is multiplied by two because an individual is twice as related with itself as with its offspring. The alternative of dividing the number of offspring by two would result in noninteger numbers, which cannot be fitted using standard generalized linear models.

Note that the distribution of this annualized measure of fitness F will never exactly comply with a Poisson distribution because 1 is an impossible value (an individual must survive, thus obtaining two fitness points, before it can reproduce). However, computer simulations (described in Appendix S1-1) confirm that selection measured as the covariance between a trait and F perfectly predicts genetic change from t to $t + 1$ when the heritability of the trait equals one (here and in the rest of the manuscript, heritability refers to the additive genetic heritability, i.e. narrow-sense heritability). Moreover, with a trait heritability smaller than one, the slope of evolution on selection was equal to the simulated heritability (see Appendix S1-1), as predicted by the breeder's equation.

Measuring body size

Although our aim is to gain a better understanding of the evolutionary relevance of fluctuating selection on body size, the use of absolute body mass measurements is complicated by the fact that juvenile age is unknown. As we have shown previously, not correcting for this age-related variation in body mass provides a misleading description of the causal relationship between body size and fitness, and it provides a poor prediction of adaptive evolution in this system (Bonnet *et al.*, 2017).

Rather than using growth curves to account for age-related variation in body mass (see Bonnet *et al.*, 2017), here we chose to instead divide body mass by body length to obtain a body mass index (BMI). We used mass over length rather than length squared, a more common BMI, because the distribution of the former was close to a Gaussian distribution, whereas the latter was right-skewed. BMI is more repeatable than body mass in juveniles (0.73 vs. 0.62). To obtain standardized selection gradients, we standardized BMI across all years by subtracting the mean and dividing by its standard deviation. An overview of the sample sizes is given in Table 1 Data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.6767m (Bonnet & Postma, 2018).

Selection analysis

Selection was estimated with a series of generalized linear models (GLMs) and generalized linear mixed models (GLMMs), regressing fitness measures on BMI. Mixed models were fitted with the R-package `MCMCglmm` (Hadfield, 2010). This package accounts for overdispersion when modelling Poisson traits.

Using the annualized measure of overall fitness, $F_{i,t}$, we first estimated selection on a year-by-year basis using a Poisson GLM with a log link, where the expected fitness of individual i at time t is predicted from:

$$\log(F_{i,t}) = \mu_{F,t} + \beta_{F,a,t}a_{i,t} + \beta_{F,s,t}s_i + \beta_{F,as,t}as_{i,t} + (\beta_{F,z,t})z_{i,t}, \quad (1)$$

where $a_{i,t}$ is the age (juvenile or adult) of individual i at year t, s_i is the sex of i, $z_{i,t}$ is the phenotype (BMI) of i at t, $\mu_{F,t}$ is the intercept of the regression, $\beta_{F,a,t}$ is the effect of age, $\beta_{F,s,t}$ is the effect of sex, $\beta_{F,as,t}$ is the interaction sex-by-age, and $\beta_{F,z,t}$ is the strength of selection on BMI. Because we used a log link, $\beta_{F,z,t}$ is a selection gradient *sensu* Lande & Arnold (1983; Smouse *et al.*, 1999; Firth *et al.*, 2015).

The standard deviation in the yearly estimates of selection (SD $(\hat{\beta}_{\mathrm{F,z,t}})$) gives a first idea about the temporal dynamic of selection, but as it includes sampling variance, it will always overestimate the real variation in selection (Morrissey & Hadfield, 2012).

Second, we estimated overall selection by fitting a Poisson GLM to pooled data from all the years, without taking into account temporal variation:

$$\log(F_{i,t}) = \mu_F + \beta_{F,a}a_{i,t} + \beta_{F,s}s_i + \beta_{F,as}as_{i,t} + \beta_{F,z}z_{i,t}. \quad (2)$$

Third, we directly estimated variation in selection by fitting a random regression to the full data set. Thus, we expanded model (2) to a Poisson GLMM by including a random intercept and a random slope for fitness as a function of BMI:

$$\log(F_{i,t}) = \mu_{F0} + \mu_{F,t} + \beta'_{F,a}a_{i,t} + \beta'_{F,s}s_i + \beta'_{F,as}as_{it} + (\beta'_{F,z} + \zeta_{F,t})z_{i,t}, \quad (3)$$

where $\beta'_{F,z}$ is the median selection estimate, $\mu_{F,t}$ is the random deviation of the global intercept (μ_{F0}) in year t, and $\zeta_{F,t}$ is the deviation of selection (i.e. the slope) in year t. The random effects $\mu_{F,t}$ and ζ_F are assumed to be multivariate normal with variances $\sigma^2_{F,\mu}$ and $\sigma^2_{F,\zeta}$, and a covariance $\sigma_{F,(\mu,\zeta)}$. The main parameter of interest in this equation is $\sigma^2_{F,\zeta}$, which captures temporal variation in selection and is free of sampling variance (Chevin *et al.*, 2015).

The median selection gradient estimate ($\beta'_{F,z}$) from model (3) differs from the estimate across all years ($\beta_{F,z}$) from model (2) if the estimate of $\sigma^2_{F,\zeta}$ is different from 0 and data are not perfectly balanced among years. Whereas the latter, $\beta_{F,z}$, is the best estimate of the overall selection, the former, $\beta'_{F,z}$, is the selection occurring in a 'standard' year. The ratio of $\sigma_{F,\zeta}/\left|\beta'_{F,z}\right|$ provides an indication of the likelihood of a reversal in the direction of selection. Assuming that the annual selection gradients follow a Gaussian distribution (as the random regression assumes), this ratio is similar to an inverse *Z*-value. Values around 0.6 indicate rare reversals (5% of the time), and values above 2 indicate common reversals (more than 31% of the time).

We repeated these analyses for annual reproductive success (ρ), again using a Poisson GLMM, and for over-winter survival (ϕ), using a logistic regression (i.e. a categorical GLMM with two levels). As was the case for *F*, because we use a log link, estimates of the strength of selection using ρ are selection gradients *sensu* Lande & Arnold (1983; Smouse *et al.*, 1999; Firth *et al.*, 2015). Although this is not the case for ϕ, the sign and strength of estimates of selection are still interpretable qualitatively. For Fig. 3c, we back-transformed the selection gradients to the scale of the data, following Janzen & Stern (1998). We did not back-transformation the selection parameters in Table 2 because it is unclear how to convert the variance in the slope to the scale of the data. The main parameters of interest, the variances in the slope of selection, are written $\sigma^2_{\phi,\zeta}$ and $\sigma^2_{\rho,\zeta}$, for viability and fertility, respectively.

Finally, we refitted model (3) with BMI standardized within years (subtracting the average and dividing by the standard deviation for each year), to evaluate whether the variation in selection comes from changes in phenotypic variance (resulting in a lower estimate of $\sigma_{F,\zeta}$), or rather from a moving fitness landscape (in which case $\sigma_{F,\zeta}$ would be unaffected).

Expected correlation between selection and evolution

We used individual-based computer simulations to explore the expected relationship between selection and evolution in a population similar to the snow vole population. Building on the model developed in Bonnet & Postma (2016), we simulated a heritable phenotypic trait, as well as reproductive success and survival from one year to the next, in discrete time, and considered two age-classes (nonreproductive juveniles and reproductive adults). For details of the simulation algorithm, see Appendix S1-1.

For every simulation replicate, we extracted the annual selection (standardized selection differentials) and annual evolution (change in mean breeding value for all individuals alive in one year and all individuals alive in the next year), and computed the correlation between them. We repeated this 500 times to estimate the expected distribution of the correlation given a set of parameter values. We tested the exactness of the simulation algorithm by comparing the ratio of evolution (R) over selection (S) to the heritability simulated (h^2) (Appendix S1-1). Besides, the equality $\frac{R}{S} = h^2$ holds when the assumptions of the breeder's equation are met, and our simulations provide a null model against which to test our hypothesis that the assumptions of breeder's equation are not valid in the snow vole population (Bonnet *et al.*, 2017).

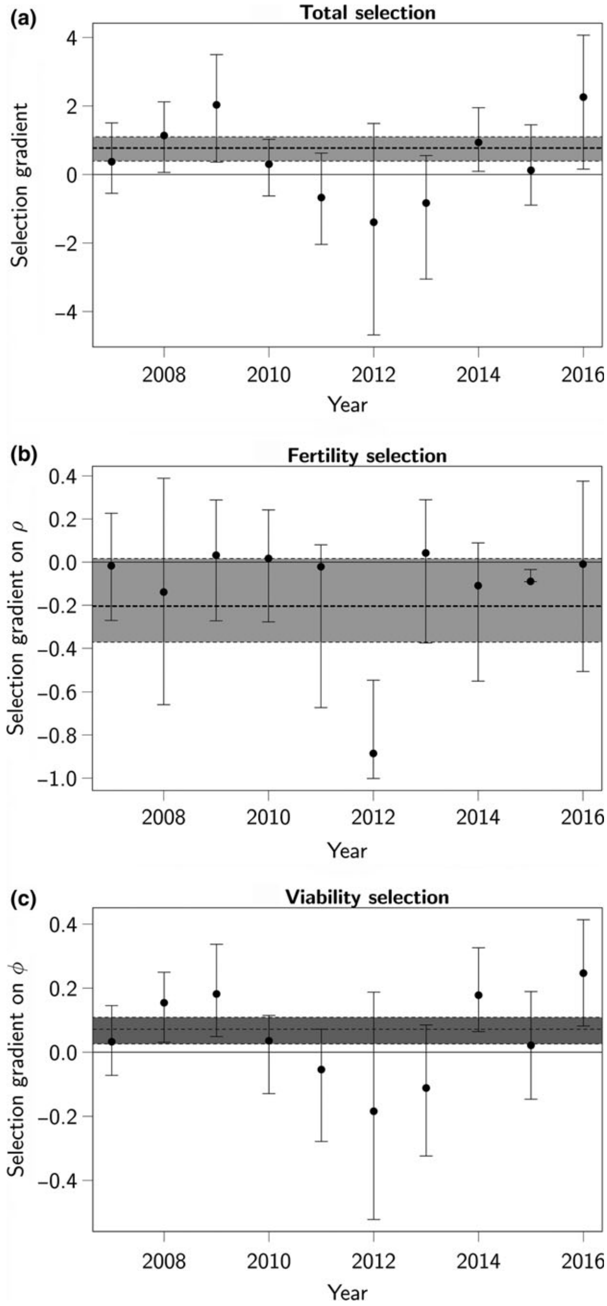

Fig. 3 Estimates of (a) total, (b) viability and (c) fertility selection gradients, year-by-year and across all years. Yearly estimates (black dots with 95% CI error bars) were obtained by fitting separate generalized linear models for each year. The overall estimate (dashed line with 95% CI depicted in grey) was produced by pooling all years together.

In a first step, we parameterized the simulations to closely match important properties of the snow vole

Table 2 Selection and temporal variation in total selection (F), fertility selection (ρ) and viability selection (ϕ) for body mass index.

Selection	β_z (SE)	$\mathrm{SD}_{\mathrm{year}}$	$\overline{\mathrm{SE}_{\mathrm{year}}}$	β'_z (SE)	σ_ζ (95% CI)	$\sigma_\zeta/\lvert\beta'_z\rvert$
Total	0.639 (0.18)	1.198	0.753	0.598 (0.309)	0.691 (0.461; 1.153)	1.156
Fertility	−0.204 (0.098)	0.277	0.160	−0.236 (0.219)	0.512 (0.385; 0.779)	2.167
Viability	0.433 (0.126)	0.843	0.533	0.439 (0.252)	0.642 (0.409; 1.024)	1.462

β_z(SE) is the selection gradient across all years and its standard error; $\mathrm{SD}_{\mathrm{year}}$ is the standard deviation of annual selection gradients; $\overline{\mathrm{SE}_{\mathrm{year}}}$ is the mean standard error of these annual estimates; β'_z(SE) is the selection gradient for the average year and its standard error; σ_ζ 95% CI is the standard deviation of the distribution of selection gradients and its 95% confidence interval; $\sigma_\zeta/\lvert\beta'_z\rvert$ is the ratio of the standard deviation in selection over the absolute median year selection, and indicates the likelihood of reversal in the direction of selection. All variables were estimated from generalized linear (mixed) models using standardized body mass index.

data set: sample size in every year; amount of genetic and environmental variance in size; strength and variability of selection. In a second step, we varied the heritability (while holding the phenotypic variance constant) or the monitoring duration, which allowed us to explore how these two parameters affect the correlation between selection and evolution (see Fig. 5 and Appendix S1-1.2).

Inference of evolution and the contribution of fluctuating selection

We estimated all quantitative genetic parameters by fitting animal models (Henderson, 1950, 1975, 1976; Kruuk, 2004) using `MCMCglmm`. This Bayesian package allows extracting and combining full posterior distributions of parameters. Unless stated otherwise, all calculations were made on the posterior distributions (rather than on point estimates) to propagate estimation uncertainty and account for covariation between parameters. For all models, we ran a MCMC chain long enough to obtain 1000 posterior samples, with a thinning interval large enough to keep the autocorrelation for any parameter below 10%, and added a burnin of about 20% of the total iterations. For fixed effects, we always used the default priors, which are normal distributions with means of zero and variances of 10^{10}. For random effects, we used inverse-Wishart priors, with parameters $V = 1$ and $v = 0.002$ for univariate models (equivalent to an inverse gamma distribution with parameters $\alpha = \beta = 0.001$), and with $V = \boldsymbol{I}(n)$ and $v = n + 1$, where n is the number of traits considered and $\boldsymbol{I}(\boldsymbol{n})$ is an identity matrix of dimension n, for multivariate models (see Appendix S1-2 for visual representations). All animal models included genetic groups (Quaas, 1988) to model gene flow from immigrants and to account for a difference in BMI breeding values between immigrants and the base population of the pedigree (Hadfield *et al.*, 2010; Wolak & Reid, 2017). To this end, we considered two groups, the base population and immigrants, and used the explicit fixed effect specification following Wolak & Reid (2017).

Because additive genetic variation in fitness is a prerequisite for a response to selection, we first estimated the genetic variance in our fitness proxy F, using a univariate animal model assuming a Poisson distribution with a log link. Overdispersion is accounted for by default in `MCMCglmm`. The model included an intercept, age, sex and their interaction, as well as date of capture as fixed effects, and additive genetic effects, individual identity (i.e. permanent environment effects), maternal identity and year as random effects. Additive genetic variance and heritability were estimated after transformation from the latent scale to the data scale, by integrating over all the random effects and fixed effects (Morrissey,2015; de Villemereuil *et al.*, 2016), using the `R` package `QGglmm` (de Villemereuil *et al.*, 2016).

We then used two approaches to infer the yearly rates of evolution in BMI: (i) a univariate approach based on best linear unbiased predictors (BLUPs) regression (Henderson, 1950; Hadfield *et al.*, 2010) and (ii) a multivariate approach based on the Robertson-Price identity (Price, 1970; Morrissey *et al.*, 2012; Bonnet *et al.*, 2017).

For the first approach, we fitted a univariate animal model to BMI data, including age, sex, their interaction and date of capture as fixed effects, and random additive genetic, permanent environment (i.e. individual identity), maternal (maternal identity) and year effects. For every two successive years, we computed the genetic change in BMI between the two sets of living individuals using BLUPs for breeding values (following Hadfield *et al.*, 2010). We simulated genetic drift down the pedigree of the snow vole population (following Hadfield *et al.*, 2010; and using the function `rbv()` in `MCMCglmm`, with genetic groups to account for immigration), and computed the range of genetic change between years that genetic drift can produce. We visualized the temporal dynamics of genetic evolution of BMI by fitting a time spline (i.e. a smoother) to the breeding values of all individuals alive in each year. The spline was fitted using a generalized additive model in the `R` package `mgcv` (Wood, 2011). We estimated a time spline for each posterior sample of the distributions of individual breeding values, to obtain the posterior distribution of evolution. We tested for the significance of evolution using the same approach but using linear regressions.

To quantify the role of variation in selection in shaping the population's evolutionary trajectory, we computed the correlations between the annual estimates of selection gradients and the change in breeding values to the next year. We used the posterior distribution of changes in breeding values, but only the point estimate of annual selection gradients, to obtain a posterior distribution of correlations.

For the second approach, we would ideally have estimated the genetic and environmental selection gradients for every year by fitting a multivariate animal model treating BMI in each year as a different trait. However, although we did initially fit such a model, because of data limitations it did not reach convergence and the priors dominated the posterior distribution. Instead, we split the data in two groups of years: those where our estimates of selection (as estimated above) were positive, and those where they were negative (see Reed *et al.*, 2016 for a similar approach). We considered BMI in these two groups of years as two different traits (M_+ and M_-, respectively). We subsequently fitted a trivariate animal model to the two BMI traits and our annualized measure of fitness (F). This model allows the estimation of an additive genetic covariance between BMI and fitness for the two year classes. Based on the Robertson-Price equation, these covariances provide a direct and unbiased expectation of the rate of evolution during the two groups of years (Robertson, 1966; Price, 1970, 1972; Frank, 2012; Morrissey *et al.*, 2012). By measuring fitness on a yearly basis, we removed the assumption of nonoverlapping generations. We compare and explain the advantages and drawbacks of both approaches in the discussion.

The trivariate animal model can be written as

$$[\boldsymbol{M_+}, \boldsymbol{M_-}, \boldsymbol{F}] \sim \boldsymbol{bX} + \boldsymbol{Z_1 a} + \boldsymbol{Z_2 m} + \boldsymbol{Z_3 p} + \boldsymbol{Z_4 y} + \boldsymbol{Ir},$$

where $\boldsymbol{X}$, $\boldsymbol{Z_1}$, $\boldsymbol{Z_2}$, $\boldsymbol{Z_3}$ and $\boldsymbol{Z_4}$ are design matrices relating BMI and fitness observations to the parameters to estimate, $\boldsymbol{b}$ is a matrix of fixed effects, $\boldsymbol{a}$, $\boldsymbol{m}$, $\boldsymbol{p}$ and $\boldsymbol{y}$ are random effects accounting for the variance associated with additive genetic, maternal, permanent environment and year effects, respectively. Residuals $\boldsymbol{r}$ are assumed to be normally distributed and independent, and are therefore associated with observations by an identity matrix $\boldsymbol{I}$. The fixed part of the model matches that used for each trait in univariate models (see above).

The matrix of breeding values $\boldsymbol{a}$ follows a multivariate normal distribution

$$\boldsymbol{a} \sim \boldsymbol{MVN}(\boldsymbol{0}, \boldsymbol{A} \otimes \boldsymbol{G})$$

where $\boldsymbol{A}$ is the relatedness matrix between all individuals, and $\boldsymbol{G}$ is the additive genetic variance–covariance matrix between the three traits.

$$\boldsymbol{G} = \begin{pmatrix} \sigma^2_{\boldsymbol{A}}(\boldsymbol{M_+}) & \sigma_{\boldsymbol{A}}(\boldsymbol{M_+M_-}) & \sigma_{\boldsymbol{A}}(\boldsymbol{M_+F}) \\ \sigma_{\boldsymbol{A}}(\boldsymbol{M_+M_-}) & \sigma^2_{\boldsymbol{A}}(\boldsymbol{M_-}) & \sigma_{\boldsymbol{A}}(\boldsymbol{M_-F}) \\ \sigma_{\boldsymbol{A}}(\boldsymbol{M_+F}) & \sigma_{\boldsymbol{A}}(\boldsymbol{M_-F}) & \sigma^2_{\boldsymbol{A}}(\boldsymbol{F}) \end{pmatrix},$$

where $\sigma^2_A(M_+)$ and $\sigma^2_A(M_-)$ are the additive genetic variance for BMI in years with positive selection and negative selection, respectively; $\sigma_A(M_+M_-)$ is the additive genetic covariance between BMI in the two group of years; $\sigma^2_A(F)$ is the additive genetic variance in fitness across years, that is the genetic differential of fitness itself (Fisher, 1958), and finally, $\sigma_A(M_+F)$ and $\sigma_A(M_-F)$ are the additive genetic covariances between fitness and BMI in years with high and low selection, respectively. We computed the genetic gradients for both groups of years as $\beta_{A+} = \sigma_A(M_+F)/\sigma^2_A(M_+)$ and $\beta_{A-} = \sigma_A(M_-F)/\sigma^2_A(M_-)$. The additive genetic correlation between BMI in the two groups of years was computed as $\sigma_A(M_+M_-)/\sigma_A(M_+)\sigma_A(M_-)$.

Environmental selection differentials $\sigma_E(M_+F)$ and $\sigma_E(M_-F)$ were calculated as the sum of the covariances between BMI and fitness in the random effect variance–covariance matrices for permanent environment, maternal identity and the residuals. The environmental variances $\sigma^2_E(M_+F)$ and $\sigma^2_E(M_-F)$ were obtained by summing the variance components of the same random effects. Subsequently, the environmental selection gradients were obtained using $\beta_{E+} = \sigma_E(M_+F)/\sigma^2_E(M_+)$ and $\beta_{E-} = \sigma_E(M_-F)/\sigma^2_E(M_-)$. Finally, the phenotypic selection gradients were recovered using $(\sigma_A(M_+F)+ \sigma_E(M_+F))/(\sigma^2_A(M_+) + \sigma^2_E(M_+))$ and $(\sigma_A(M_-F) + \sigma_E (M_-F))/(\sigma^2_A(M_-) + \sigma^2_E(M_-))$. The phenotypic selection gradients represent selection on the phenotype *sensu* Lande & Arnold (1983), whereas the environmental and the additive genetic selection gradient represent the indirect action of selection (they are not selection in a strict sense) on the environmental and additive genetic part of phenotypic variation, respectively. These three gradients are equal if the assumptions of the breeder's equation are met, that is when the phenotypic covariation between the trait and fitness is causal and not in part the result of unmeasured environmental covariates (Rausher, 1992). For size-related traits, disproportionately large environmental selection gradients might be interpreted as the effect of nonheritable body condition shaping both mass and fitness, whereas the additive genetic selection gradient captures causal, direct selection on the trait.

To confirm the values of the genetic and environmental covariances between BMI and F, we additionally ran two bivariate animal models with M_+ and F, and M_- and F, respectively. In addition, to confirm the stability of the genetic covariance through time (see Results), we refitted the trivariate model, but instead of distinguishing between years with positive and negative phenotypic selection, we treated BMI in every second year as one trait (equivalent to M_-), and BMI in the other years as another trait (equivalent to M_{+}). Finally, we assessed potential biases in $\beta_A - \beta_E$ generated by splitting the data set into two groups of years on the basis of the direction of phenotypic selection: Using the R-package `pedantics` (Morrissey & Wilson, 2010), we simulated phenotypes and fitness values with β_A equal to β_E, split the data based on the sign of the selection differential in every year, and fitted animal models to estimate the expected distribution of $\beta_{A+} - \beta_{E+}$ and $\beta_{A-} - \beta_{E-}$. See Appendix S1-4 for details.

Results

Yearly estimates of selection

Annual selection gradients varied considerably (standard deviation = 1.198) around the overall selection gradient (0.639 ± SE 0.18; Fig. 3a). Estimates of total selection were mostly positive, but appeared to have been negative in 3 years. Although the standard deviation of the yearly estimates was greater than the overall selection gradient, a large proportion of this variation must be attributable to sampling error. Indeed, yearly selection was estimated with much less precision than overall selection, as is reflected by a mean standard error of the yearly estimates of 0.753. Fertility and viability selection gradients showed similar patterns (Fig. 3b,c): the standard deviations of the estimates of viability and fertility selection were high, but so were the mean standard errors of these estimates (Table 2).

Fluctuation of selection

Fitting eqn (3), we estimated $\sigma_{F,\zeta} = 0.691$ [95% CI (0.461; 1.153)] and $\sigma_{F,\zeta}/\left|\beta'_{F,z}\right| = 1.156$. Assuming a normal distribution of selection gradients and a ratio of standard deviation over mean of 1.156, a reversal of the direction of selection would be occasional (about once every 5 years). Refitting eqn (3) with BMI standardized within years yielded a similar estimate of $\sigma_{F,\zeta}$ [0.674 (0.433; 1.162)]. Variance in fertility selection was estimated as $\sigma_{\rho,\zeta} = 0.512$ [95% CI (0.385; 0.779)], more than twice the absolute median selection gradient (Table 2), meaning that selection was very likely to change direction. Variance in viability selection was estimated as $\sigma_{\phi,\zeta} = 0.642$ [95% CI (0.409; 1.024)]. The correlations between random intercepts and random slopes were close to zero for all three models ($\sigma_{F,(\mu,\zeta)} = -0.11$, $\sigma_{\rho,(\mu,\zeta)} = 0.08$, $\sigma_{\phi,(\mu,\zeta)} = -0.16$), suggesting appropriate estimation of the variance components (the correlation is close to 1 or −1 when the model fit is (quasi-)singular).

Fluctuation of evolution

There was a small but significant amount of additive genetic variation in our proxy of annual fitness: on the latent scale of the Poisson model, the additive genetic variation was estimated to be 0.299 (0.086; 0.692). On the scale of the data, this translates into an additive genetic variance of 0.280 (0.001; 0.994) and a heritability of 1.13% (0.06%; 5.01%). This is comparable to the heritability of lifetime fitness in Bonnet *et al.* (2017), which used a lifetime rather than annual measure of fitness. We found significant additive genetic variation in BMI [167 g^2/m^2 (98; 307)]; heritability = 16.7% (8.9%; 26.3%). In this population, there is evidence for adaptive evolution towards lower body size from 2006 to 2014 (Bonnet *et al.*, 2017). This pattern is also found for BMI (Fig. 4), with a decrease in mean breeding value of −3.69 (−8.13; 0.51) on a mean trait value of 288, and a 3% probability that the trend is not negative. The trend might have reversed between 2014 and 2016 (Fig. 4), when there is some evidence that the breeding values for BMI have increased by 2.51 (−1.1; 8.42) (6.5% probability that the change is not positive). The decrease in breeding values from 2006 to 2014 is unlikely to have been produced solely by genetic drift, with a probability that drift generated a decrease that is at least as large of $p_{\mathrm{MCMC}} = 0.064$ (see also Bonnet *et al.*, 2017), whereas drift could have produced the rebound from 2014 to 2016 ($p_{\mathrm{MCMC}} = 0.24$).

From selection to evolution

Given the heritability of BMI and the duration of the snow vole monitoring, the correlation between selection gradients and change in breeding values from 1 year to the next is expected to be strongly positive on average, but also highly variable: individual-based simulations show that the distribution of the correlation between selection and evolution has its mode at 0.68 with 95% CI (−0.12; 0.94) (see Fig. 5 and

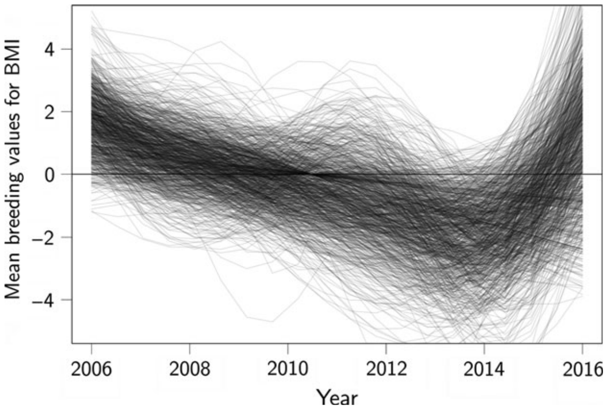

Fig. 4 Temporal dynamics of mean breeding values for body mass index (BMI). Each line was obtained from a different MCMC posterior sample, by fitting a time spline to the mean of estimated breeding values among individuals alive in any given year.

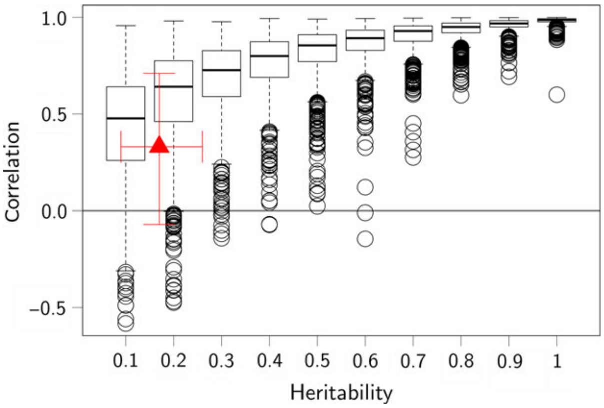

Fig. 5 Realized correlation between selection and evolution as a function of simulated heritability. Selection was measured as a standardized selection differential on annualized fitness, and evolution was measured as the difference between mean breeding values of individuals present in one year and those presents in the next year. Simulations consisted of eleven years (as in the snow vole data set). The empirical estimates for the heritability and the correlation are drawn in red, together with their confidence intervals.

Appendix S1-1.2). This variability is due to strong genetic drift combined with the relatively small number of years. Increasing the heritability of the trait increases the expected correlation and reduces its variability (Fig. 5), whereas increasing the duration of the monitoring reduces variability only (Appendix S1-1.2). Empirically, the correlation is estimated with much uncertainty and is not statistically significantly different from zero ($p_{\mathrm{MCMC}} = 0.08$). Nevertheless, the most likely value is positive [mode 0.33, 95% CI (−0.07; 0.71)] and does not lie in the extreme tail of the theoretical distribution (Fig. 5).

As expected, in years with positive selection (based on selection gradients from year-by-year GLMs, see above), the selection gradient reconstructed from our trivariate animal model was positive, whereas it was negative for years with negative selection gradients (Fig. 6). Importantly however, the genetic gradients were negative in both groups of years (Fig. 6) and did not differ from each other [$\beta_{A+} - \beta_{A-} = -0.0011$, 95% CI (−0.0164; 0.0112), $p_{\mathrm{MCMC}} = 0.72$].

The environmental gradients, on the other hand, differed from each other [$\beta_{E+} - \beta_{E-} = 0.0218$, 95% CI (0.0009; 0.0355), $p_{\mathrm{MCMC}} = 0.036$], with β_{E+} being significantly positive, and β_{E-} slightly negative. Moreover, during years of positive selection, the genetic and environmental gradients were of opposite sign (Fig. 6), and significantly different [$\beta_{A+} - \beta_{E+} = -0.0260$, 95% CI (−0.0454; −0.0028), $p_{\mathrm{MCMC}} = 0.034$]. On the other hand, during years of negative selection, the genetic and environmental gradients were both negative (Fig. 6) and not significantly different [$\beta_{A-} - \beta_{E-} = -0.0045$, 95% CI (−0.0282; 0.0205), $p_{\mathrm{MCMC}} = 0.824$]. Finally, the genetic correlation between BMI in positive selection years and BMI in negative selection years was strongly positive [0.61, 95% CI (0.22; 0.83)]. The stability of these results was confirmed by splitting the data set differently (see Appendix S1-3).

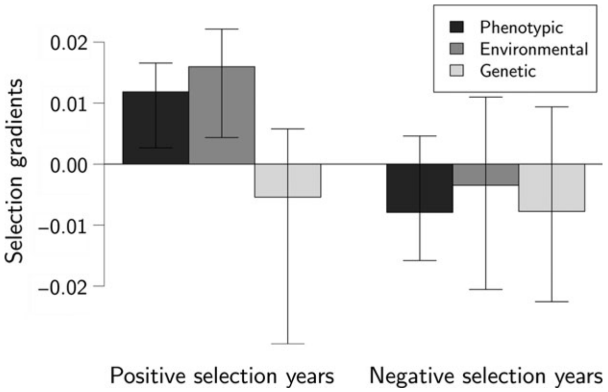

Fig. 6 Phenotypic selection gradients and their decomposition into environmental and genetic gradients for years with positive selection on body mass index (BMI) and for years with negative selection on BMI. Error bars show 95% confidence intervals.

Discussion

Here we have shown that selection on BMI fluctuates in a natural population of snow voles. In addition, we have shown that BMI has evolved, but that both the rate and direction of evolution do not appear to be tightly coupled with the dynamics of selection. Below we discuss the methodological challenges posed by the quantification of variation in selection and its evolutionary relevance, and our contribution to their resolution. We then discuss whether our analyses can inform us about the mechanisms of fluctuating selection and what is needed to answer the questions that are beyond the reach of our analyses. Finally, we discuss the importance of timescale when studying variation in selection and evolution.

The modelling of evolution and selection

The random regression method of Chevin *et al.* (2015) provides a statistically rigorous way to quantify and test for the significance of variation in selection. On its own, however, a random regression does not address the evolutionary relevance of fluctuating selection. To establish the latter, two additional issues need to be investigated: (i) variation in the strength of selection will reverse the direction of evolution only if it fluctuates not only in strength, but also in direction (see Fig. 1b,c); (ii) as selection does not always lead to an evolutionary response (Rausher, 1992; Merilä *et al.*, 2001; Morrissey *et al.*, 2010), fluctuating selection does not necessarily translate into fluctuating evolution (see Fig. 1d).

To address the first issue, we considered where the distribution of selection gradients, estimated by a random regression, is located relative to zero. If this distribution is centred around zero, selection reversal is frequent, whereas if the distribution does not overlap much with zero, selection reversal is rare. We evaluated the likelihood of selection reversal by calculating the ratio of the standard deviation of selection gradients over the absolute median selection gradient ($\sigma_\zeta / |\beta'_z|$). As this ratio increases, the fluctuation of selection becomes increasingly biologically relevant, and a reversal becomes increasingly likely. However, even if the distribution of selection gradients is symmetric (which it does not have to be), as our estimate of the distribution of selection gradients is based on a finite number of years, it is unlikely to comply with an inverse *Z*-distribution. Furthermore, as selection gradients may show temporal autocorrelation, the appropriate number of degrees of freedom is unclear. Furthermore, as selection gradients may show temporal autocorrelation, the appropriate number of degrees of freedom is unclear. Hence, we are reluctant to translate this ratio into a probability of reversal. Nevertheless, it gives a qualitative assessments of the likelihood of reversal that could be developed further into a more quantitatively rigorous measure.

To address the second issue, we estimated the coupling between variation in selection and variation in genetic change. This exercise proved to be challenging and provided somewhat mixed results. In a first approach, we computed the correlation between selection and year-to-year changes in breeding values by relating the full distribution of the change in BLUPs for breeding values to point estimates of selection gradients. Therefore, the uncertainty accompanying the selection estimates was not propagated to this correlation. In contrast, the trivariate animal model estimates selection and evolution within the same model, along with their respective uncertainties. This allows for the integration of uncertainty in both selection and evolution when comparing genetic and environmental gradients, and to take into account the nonindependence of their posterior distributions. Unfortunately however, this multivariate approach is particularly data hungry, and the snow vole population is too small to estimate year-specific genetic parameters. As a consequence, we were forced to compare groups of years with negative and positive selection, although this approach generates a bias in the estimated difference between genetic and environmental gradients (Appendix S1-4). Fortunately, in our particular case, the bias is in the direction opposite to our findings, and our analyses are hence statistically conservative. Nevertheless, the presence of biases makes this approach risky, and its correct interpretation relies on computationally intense simulations. In conclusion, whenever the population size allows for it, and to avoid the aforementioned problems and biases, we advocate the use of year-specific multivariate animal models for assessing the coupling of selection and evolution.

Coupling of selection and evolution

Simple algebra shows that a positive correlation between selection and evolution is expected. For a trait *z*, a selection gradient is the ratio of the phenotypic covariance between trait and relative fitness, over the phenotypic variance in the trait:

$$\beta_P = \frac{\sigma_P(z, F)}{\sigma_P^2(z)}.$$

Assuming a standard quantitative genetic model in which there is no correlation or interaction between the genetic effects and the environmental effects (i.e. an absence of genotype–environment correlations and interactions), z can be decomposed into additive genetic effects and environmental effects $z = a + e$. Similarly, the phenotypic covariance ($\sigma_P(z, F)$, i.e. the selection differential) can be decomposed into an additive genetic ($\sigma_A(z, F)$) and an environmental covariance ($\sigma_E(z, F)$). Therefore, the phenotypic selection gradient (β_P) can be written as:

$$\beta_P = \frac{\sigma_A(z, F) + \sigma_E(z, F)}{\sigma_P^2(z)}.$$

According to the Robertson-Price identity (Robertson, 1966; Price, 1970), $\sigma_A(z, F)$ is the expected rate of genetic change. From the above, it follows that the phenotypic selection gradient is likely to be positively correlated with evolution (provided the latter is nonzero). Even if their signs are opposite in all years, years with more positive selection gradients will go with less negative genetic change, and vice versa.

Using computer simulations, we found that for our data set, if the relationship between trait and fitness (i.e. selection) is causal (Reed *et al.*, 2016), the correlation between evolution and selection is expected to be relatively strong and positive [0.68, 95% CI (−0.12; 0.94)]. Nevertheless, this correlation has a 7.8% (SE 0.2%) probability to be zero or negative because of the potentially large effect of genetic drift.

The observed correlation between selection and evolution among years was not significantly different from zero nor from the theoretical expectation (see Results). Nevertheless, there are good reasons to think that phenotypic selection on size may not translate into consistent evolution: (i) across-year selection favours larger sizes (Fig. 3), whereas evolution is towards smaller sizes (Fig. 4); and (ii) in years of positive selection, the genetic gradient differs in sign from the environmental and phenotypic gradients (Fig. 6), a pattern also seen when analysing all years together (Bonnet *et al.*, 2017). Therefore, our finding that the correlation between selection and evolution does not deviate significantly from the null expectation may be the result of a lack of statistical power, and not of the lack of an environmental bias.

What drives fluctuations in selection?

Although our random regression and quantitative genetic models give a thorough description of the dynamics of selection and evolution in this population, they do not provide direct insight into the underlying mechanisms. We have shown that selection fluctuates, and thus that the relationship between size and fitness changes at the population level, but why does selection change? Different processes may lead to the same distribution of directional selection gradients, and based on the analysis of selection gradients alone, it is difficult to distinguish fluctuations due to a moving fitness optimum from those due to a change in the distribution of phenotypes among years (Chevin & Haller, 2014). The latter could have played a role here as we find substantial variation between years in both the mean phenotype (ranging between 277 and 312 g/m) and its variance (ranging between 1779 and 4573 g^2/m^2). Nevertheless, we can rule out that change in the phenotypic distribution played a major role in the fluctuation of selection because the estimate of variation in the slope of selection was almost identical in models where the phenotype was standardized among years vs. within years. Fluctuation in selection was therefore the result of variation in the fitness landscape, but we do not know what drove this variation.

If we are to gain a deeper understanding of the dynamics of the fitness landscape and the ecological drivers of selection, we ultimately need to move beyond the estimation of variance parameters, towards a more mechanistic understanding of the genetic and ecological sources of phenotypic variation and their covariance with fitness (Morrissey & Hadfield, 2012). Good examples where we know the detailed ecological driver of variation in selection are still scarce. Some notable exceptions include beak size in Darwin finches (Grant & Grant, 2002), reproductive timing in great tits (Husby *et al.*, 2011), and insecticide resistance in *Culex* mosquitoes (Milesi *et al.*, 2016). All of these, as well as this study, rely on individual-based long-term monitoring, difficult and costly to upkeep, but necessary to disentangle the causes and consequences of selection in natural populations (Clutton-Brock & Sheldon, 2010). The snow vole monitoring is more complete and spans over more generations (approximately nine) than most longitudinal studies of wild populations, but our simulations highlight that this is not sufficient yet to fully describe the evolutionary consequences of fluctuating selection on size. Future studies might hence consider specifically targeting highly heritable traits (Fig. 5) to obtain a stronger and less variable expected correlation between selection and evolution, whereas they wait for more data to accumulate.

Alternatively, meta-analyses of many replicated estimates of selection may reveal preponderant drivers of selection across species and ecosystems, even if individual studies are often short-term and lack resolution (e.g. Caruso *et al.*, 2017; Siepielski *et al.*, 2017).

Timescale

Despite fluctuations in the strength and direction of phenotypic selection, the rate and direction of evolution was constant and nonzero over most of the study period. Thereby, our findings are at odds with the idea that fluctuating selection causes short-term evolutionary stasis. Nevertheless, fluctuating selection may be a driver of short-term evolutionary dynamics in other natural populations, where the selection measured by regression-based methods is causal and not dominated by an environmental covariation between traits and fitness. Moreover, it is unlikely that fluctuating selection will not be evolutionary relevant on longer timescales, in the snow voles and in other species. Indeed, over geological timescales, bounded fluctuations of phenotypic evolution are increasingly attributed to responses to fluctuating selection, rather than to sampling variation and evolutionary stasis (Uyeda *et al.*, 2011; Voje *et al.*, 2015). Unless the environment is constant, causal selective pressures are likely to change over longer time periods, either because the fitness landscape changes or because the phenotypic distribution changes through evolutionary adaptation or phenotypic plasticity.

Fluctuating selection and evolution might go undetected because the time frame is too short. For instance in the snow vole population, adaptive evolution and the causal selective pressure causing it are probably related to a short-term climatic anomaly which goes against long-term changes induced by global climate change. On the other hand, we may have missed some fluctuating selection and evolution because the temporal resolution at which selection is estimated is too low, smoothing out very short-term changes in selection and the rate of genetic change. The latter is not unlikely in the snow vole population, where the causal selective pressure varies seasonally: viability selection is null early in the reproductive season and increases throughout summer (Bonnet *et al.*, 2017).

Conclusion

Whereas our results do not argue against the evolutionary relevance of fluctuating selection in general, they warn against interpreting any phenotypic fluctuating selection in terms of fluctuating evolution: as the dynamics of selection and evolution can be uncoupled on certain timescales, fluctuating selection does not necessarily provide a general explanation for evolutionary stasis. Thereby, we have highlighted the danger of relying on temporally replicated phenotypic estimates of selection to understand and predict the evolutionary dynamics of natural populations. Instead, quantifying the evolutionary relevance of fluctuating selection requires a joint analysis of selection and evolution.

Acknowledgments

Thanks to two anonymous reviewers for constructive corrections and suggestions. Thanks to Jarrod D. Hadfield, Lukas F. Keller, Marc Kéry and Pirmin Nietlisbach for comments on earlier versions of this work. Thanks to the many field helpers over 11 years. The snow vole monitoring was authorized by the Amt für Lebensmittelsicherheit und Tiergesundheit, Chur, Switzerland, and supported by the Claraz-Donation. This work was funded by Swiss National Science Foundation project grants 31003A_141110 and 31003A_159462/1 awarded to EP, and by the Forschungskredit of the University of Zurich, grant no. FK-16-097. The authors declare no conflict of interest.

References

Bell, G. 2010. Fluctuating selection: the perpetual renewal of adaptation in variable environments. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**: 87–97.

Bergland, A.O., Behrman, E.L., O'Brien, K.R., Schmidt, P.S. & Petrov, D.A. 2014. Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in *Drosophila*. *PLoS Genet.* **10**: e1004775.

Blanckenhorn, W. 2000. The evolution of body size: what keeps organisms small? *Q. Rev. Biol.* **75**: 385–407.

Bonnet, T. & Postma, E. 2016. Successful by chance? The power of mixed models and neutral simulations for the detection of individual fixed heterogeneity in fitness components. *Am. Nat.* **187**: 60–74.

Bonnet, T. & Postma, E. 2018. Data from: Fluctuating selection and its (elusive) evolutionary consequences in a wild rodent population. *Dryad Digital Repository*. doi:10.5061/dryad.6767m.

Bonnet, T., Wandeler, P., Camenisch, G. & Postma, E. 2017. Bigger is fitter? Quantitative genetic decomposition of selection reveals an adaptive evolutionary decline of body mass in a wild rodent population. *PLoS Biol.* **15**: e1002592.

Brookfield, J.F. 2016. Why are estimates of the strength and direction of natural selection from wild populations not congruent with observed rates of phenotypic change? *BioEssays* **38**: 1–8.

Caruso, C.M., Martin, R.A., Sletvold, N., Morrissey, M.B., Wade, M.J., Augustine, K.E. *et al.* 2017. What are the environmental determinants of phenotypic selection? A meta-analysis of experimental studies. *Am. Nat.* **190**: 363–376.

Chevin, L.M. & Haller, B.C. 2014. The temporal distribution of directional gradients under selection for an optimum. *Evolution* **68**: 3381–3394.

Chevin, L.M., Visser, M.E. & Tufto, J. 2015. Estimating the variation, autocorrelation, and environmental sensitivity of phenotypic selection. *Evolution* **69**: 2319–2332.

Clutton-Brock, T. & Sheldon, B.C. 2010. Individuals and populations: the role of longterm, individual-based studies of animals in ecology and evolutionary biology. *Trends Ecol. Evol.* **25**: 562–573.

Darwin, C. 1859. *On the Origin of Species*. John Murray, London, UK.

Endler, J.A. 1986. *Natural Selection in the Wild*. Princeton University Press, Princeton, NJ.

Estes, S. & Arnold, S.J. 2007. Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. *Am. Nat.* **169**: 227–244.

Firth, J.A., Hadfield, J.D., Santure, A.W., Slate, J. & Sheldon, B.C. 2015. The influence of nonrandom extra-pair paternity on heritability estimates derived from wild pedigrees. *Evolution* **69**: 1336–1344.

Fisher, R. 1958. *The Genetical Theory of Natural Selection*, 2nd edn. Dover Publications, New York, NY.

Fisher, R.A. & Ford, E.B. 1947. The spread of a gene in natural conditions in a colony of the moth *Panaxia dominula* L. *Heredity* **1**: 143–174.

Frank, S.A. 2012. Natural selection. IV. The price equation. *J. Evol. Biol.* **25**: 1002–1019.

García-Navas, V., Bonnet, T., Waldvogel, D., Wandeler, P., Camenisch, G. & Postma, E. 2015. Gene flow counteracts the effect of drift in a Swiss population of snow voles fluctuating in size. *Biol. Conserv.* **191**: 168–177.

Grant, P.R. & Grant, B.R. 2002. Unpredictable evolution in a 30-year study of Darwin's finches. *Science* **296**: 707–711.

Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Munsterberg, A. *et al.* 1990. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* **346**: 245–250.

Hadfield, J.D. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *J. Stat. Soft.* **33**: 1–22.

Hadfield, J.D., Richardson, D.S. & Burke, T. 2006. Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. *Mol. Ecol.* **15**: 3715–3730.

Hadfield, J.D., Wilson, A.J., Garant, D., Sheldon, B.C. & Kruuk, L.E.B. 2010. The misuse of BLUP in ecology and evolution. *Am. Nat.* **175**: 116–125.

Henderson, C.R. 1950. Estimation of genetic parameters. *Ann. Math. Stat.* **21**: 309–310.

Henderson, C.R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**: 423–447.

Henderson, C.R. 1976. Simple method for computing inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* **32**: 69–83.

Hendry, A.P. 2017. *Eco-evolutionary Dynamics*. Princeton University Press, Princeton, NJ.

Hendry, A.P. & Kinnison, M.T. 1999. Perspective: the pace of modern life: measuring rates of contemporary microevolution. *Evolution* **53**: 1637–1653.

Hereford, J., Hansen, T. & Houle, D. 2004. Comparing strengths of directional selection: how strong is strong? *Evolution* **58**: 2133–2143.

Hunt, J., Bussière, L.F., Jennions, M.D. & Brooks, R. 2004. What is genetic quality?. *Trends Ecol. Evol.* **19**: 329–333.

Husby, A., Visser, M.E. & Kruuk, L.E.B. 2011. Speeding up microevolution: the effects of increasing temperature on selection and genetic variance in a wild bird population. *PLoS Biol.* **9**: e1000585.

Janeau, G. & Aulagnier, S. 1997. Snow vole – *Chionomys nivalis* (Martins 1842). *J. Mount. Ecol.* **4**: 1–11.

Janzen, F.J. & Stern, H.S. 1998. Logistic regression for empirical studies of multivariate selection. *Evolution* **52**: 1564–1571.

Jones, O.R. & Wang, J. 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. *Mol. Ecol. Resour.* **10**: 551–555.

Jones, A.G., Arnold, S.J. & Bürger, R. 2004. Evolution and stability of the G-matrix on a landscape with a moving optimum. *Evolution* **58**: 1639–1654.

Kingsolver, J.G. & Diamond, S.E. 2011. Phenotypic selection in natural populations: what limits directional selection? *Am. Nat.* **177**: 346–357.

Kingsolver, J.G., Hoekstra, J.M., Berrigan, D., Vignieri, S.N., Hill, C.E., Hoang, A. *et al.* 2001. The strength of phenotypic selection in natural populations. *Am. Nat.* **157**: 245–261.

Kruuk, L.E.B. 2004. Estimating genetic parameters in natural populations using the “animal model”. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **359**: 873–890.

Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30**: 314–334.

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution* **33**: 402–416.

Lande, R. & Arnold, S.J. 1983. The measurement of selection on correlated characters. *Evolution* **37**: 1210–1226.

Luque-larena, J.J., López, P. & Gosálbez, J. 2002. Microhabitat use by the snow vole *Chionomys nivalis* in alpine environments reects rock-dwelling preferences. *Can. J. Zool.* **80**: 36–41.

Merilä, J., Sheldon, B.C. & Kruuk, L.E.B. 2001. Explaining stasis: microevolutionary studies in natural populations. *Genetica* **112**: 199–222.

Milesi, P., Lenormand, T., Lagneau, C., Weill, M. & Labbé, P. 2016. Relating fitness to long-term environmental variations in natura. *Mol. Ecol.* **25**: 5483–5499.

Morrissey, M.B. 2015. Evolutionary quantitative genetics of nonlinear developmental systems. *Evolution* **69**: 2050–2066.

Morrissey, M.B. & Hadfield, J.D. 2012. Directional selection in temporally replicated studies is remarkably consistent. *Evolution* **66**: 435–442.

Morrissey, M.B. & Wilson, A.J. 2010. PEDANTICS: an R package for pedigree-based genetic simulation and pedigree manipulation, characterization and viewing. *Mol. Ecol. Resour.* **10**: 711–719.

Morrissey, M.B., Kruuk, L.E.B. & Wilson, A.J. 2010. The danger of applying the breeder’s equation in observational studies of natural populations. *J. Evol. Biol.* **23**: 2277–2288.

Morrissey, M.B., Parker, D.J., Korsten, P., Pemberton, J.M., Kruuk, L.E.B. & Wilson, A.J. 2012. The prediction of adaptive evolution: empirical application of the secondary theorem of selection and comparison to the breeder’s equation. *Evolution* **66**: 2399–2410.

Mousseau, T.A. & Roff, D.A. 1987. Natural selection and the heritability of fitness components. *Heredity* **59**: 181–197.

Postma, E. 2014. Four decades of estimating heritabilities in wild vertebrate populations: improved methods, more data, better estimates? In: *Quantitative Genetics in the Wild*, 1st edn (A. Charmentier, D. Garant & L.E.B. Kruuk, eds), pp. 16–33. Oxford University Press, Oxford, UK.

Price, G.R. 1970. Selection and covariance. *Nature* **227**: 520–521.

Price, G. 1972. Extension of covariance selection mathematics. *Ann. Hum. Genet.* **35**: 85–490.

Price, T. & Liou, L. 1989. Selection on clutch size in birds. *Am. Nat.* **134**: 950–959.

Quaas, R. 1988. Additive genetic model with groups and relationships. *J. Dairy Sci.* **71**: 1338–1345.

Qvarnström, A., Brommer, J.E. & Gustafsson, L. 2006. Testing the genetics underlying the co-evolution of mate choice and ornament in the wild. *Nature* **441**: 84–86.

R Core Team 2015. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Rausher, M.D. 1992. The measurement of selection on quantitative traits: biases due to environmental covariances between traits and fitness. *Evolution* **46**: 616–626.

Reed, T.E., Gienapp, P. & Visser, M.E. 2016. Testing for biases in selection on avian reproductive traits and partitioning direct and indirect selection using quantitative genetic models. *Evolution* **70**: 2211–2225.

Robertson, A. 1966. A mathematical model of the culling process in dairy cattle. *Anim. Prod.* **8**: 95–108.

Robinson, M.R., Pilkington, J.G., Clutton-Brock, T.H., Pemberton, J.M. & Kruuk, L.E.B. 2008. Environmental heterogeneity generates fluctuating selection on a secondary sexual trait. *Curr. Biol.* **18**: 751–757.

Siepielski, A.M., Dibattista, J.D. & Carlson, S.M. 2009. It’s aobut time: the temporal dynamics of phenotypic selection in the wild. *Ecol. Lett.* **12**: 1261–1276.

Siepielski, A.M., Morrissey, M.B., Buoro, M., Carlson, S.M., Caruso, C.M., Clegg, S.M. *et al.* 2017. Precipitation drives global variation in natural selection. *Science* **355**: 959–962.

Smouse, P.E., Meagher, T.R. & Lobak, C.J. 1999. Parentage analysis in *Chaemeclirium luteum* (L.) Gray (Liliaceae): why do some males have higher contributions? *J. Evol. Biol.* **12**: 1069–1077.

Stinchcombe, J.R., Agrawal, A.F., Hohenlohe, P.A., Arnold, S.J. & Blows, M.W. 2008. Estimating nonlinear selection gradients using quadratic regression coefficients: double or nothing? *Evolution* **62**: 2435–2440.

Uyeda, J.C., Hansen, T.F., Arnold, S.J. & Pienaar, J. 2011. The million-year wait for macroevolutionary bursts. *Proc. Natl. Acad. Sci. USA* **108**: 15908–15913.

de Villemereuil, P., Schielzeth, H., Nakagawa, S. & Morrissey, M. 2016. General methods for evolutionary quantitative genetic inference from generalised mixed models. *Genetics* **204**: 1281–1294.

Voje, K.L., Holen, Ø.H., Liow, L.H. & Stenseth, N.C. 2015. The role of biotic forces in driving macroevolution: beyond the Red Queen. *Proc. Biol. Sci.* **282**: 1–9.

Wade, M.J. 2006. Natural selection. In: *Evolutionary Genetics: Concepts and Case Studies* (C.W. Fox & J.B. Wolf, eds), pp. 399–413. Oxford University Press, Oxford, UK.

Wandeler, P. & Camenisch, G. 2011. Identifying Y-chromosomal diversity by longtemplate PCR. *Mol. Ecol. Resour.* **11**: 835–841.

Wandeler, P., Ravaioli, R. & Bucher, T.B. 2008. Microsatellite DNA markers for the snow vole (*Chionomys nivalis*). *Mol. Ecol. Resour.* **8**: 637–639.

Wang, J. 2004. Sibship reconstruction from genetic data with typing errors. *Genetics* **166**: 1963–1979.

Wolak, M.E. & Reid, J.M. 2017. Accounting for genetic differences among unknown parents in microevolutionary studies: how to include genetic groups in quantitative genetic animal models. *J. Anim. Ecol.* **86**: 7–20.

Wood, S. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Series B. Stat. Methodol.* **73**: 3–36.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article: **Appendix S1** (1) Individual-based simulations; (2) Prior visualization; (3) Alternative splitting of the dataset; (4) Estimation of the bias introduced by splitting the dataset on the basis of the direction of phenotypic selection.

Data deposited at Dryad: https://doi.org/10.5061/dryad.6767m

Received 6 November 2017; revised 16 January 2018; accepted 18 January 2018