The quantitative genetics of fitness in a wild bird population

Maria Moiron[1,2*], Anne Charmantier[1†], Sandra Bouwhuis[2†]

[1] CEFE, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France

[2] Institute of Avian Research, Wilhelmshaven, Germany

*Author for correspondence: mariamoironc@gmail.com

[†] Contributed equally as last authors

Running headline: Genetics of fitness in the wild

Keywords: adaptive potential, animal model, heritability, lifetime reproductive success, latent scale

ABSTRACT

Additive genetic variance in fitness equals the change in mean fitness due to selection. It is a prerequisite for adaptation, as a trait must be genetically correlated with fitness in order to evolve. Despite its relevance, additive genetic variance in fitness has not often been estimated in wild populations. Here, we investigate additive genetic variance in lifetime fitness, as well as its underlying components, in common terns (*Sterna hirundo*). Using a series of animal models applied to 28 years of data comprising ca. 6000 pedigreed individuals, we find nominally zero additive genetic variance in the Zero-inflated component of lifetime fitness, and low but unreliable variance in the Poisson component. We also find low but likely nonzero additive genetic variance in adult annual reproductive success, but not in survival. As such, our study (i) suggests heritable variance in common tern fitness to result mostly from heritable variance in reproductive success, rather than in early-life or adult survival, (ii) shows how studying the genetic architecture of fitness in natural populations remains challenging, and (iii) highlights the importance of maintaining long-term individual-based studies such that a major research aim in evolutionary ecology will come within better reach in the next decade.

INTRODUCTION

Fisher's Fundamental Theorem of Natural Selection postulates that "*the rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time*" (Fisher 1930). The additive genetic variance of fitness has therefore been considered the single most useful statistic quantifying selection (Burt 1995). Genetic variation in fitness is also a prerequisite for adaptation and evolution, as a trait must be genetically correlated with fitness to evolve (Robertson 1966; Price 1970). Hence, understanding the quantitative genetics of individual variation in fitness is arguably one of the most important aims in evolutionary ecology (Burt 1995; Ellegren and Sheldon 2008; Walsh and Blows 2009; Gomulkiewicz and Shaw 2013; Shaw and Shaw 2014; Hendry et al. 2018).

All else being equal, a consequence of Fisher's Fundamental Theorem of Natural Selection would be that selection would drive additive genetic variance in fitness to zero. As such, one would predict the level of additive genetic variance in traits to be inversely correlated with the traits' correlation with fitness. However, this process and prediction can be counterbalanced by new variation arising from other sources, such as immigration by individuals with other genotypes, or mutation. Indeed, Houle et al. (1996) found that traits that were more closely associated with fitness had higher mutational variances, most likely do to them being underpinned by a larger number of loci (Houle 1992; Houle et al. 1996; Merilä and Sheldon 1999). Considerable debate therefore has surrounded the question of whether additive genetic variation in fitness should be low or not (e.g., Jones 1987; Burt 1995; Merilä and Sheldon 1999; Shaw and Shaw 2014). Unfortunately, empirical estimates of additive genetic variance in fitness from wild populations have so far not shed the much-needed light on this debate.

A recent review of 30 studies on humans, other animals and plants found that there were very few estimates of additive genetic variance in fitness (or fitness components) in the wild, and that those that are available varied substantially, with many estimates close to zero, and few large estimates (Hendry et al. 2018). To provide some examples: Kruuk et al. (2000) found no evidence for a significant heritability of lifetime fitness in Scottish red deer (*Cervus elaphus*). Similarly, additive genetic variance was estimated to be zero in both sexes of bighorn sheep (*Ovis Canadensis)* in Canada (Coltman et al. 2005) and very small in North American red squirrels (*Tamiasciurus hudsonicus*) (McFarlane et al. 2014, 2015). In birds, Gustafsson (1986) estimated a not-significantly-different-from-zero heritability of lifetime reproductive success in male and female collared flycatchers (*Ficedula albicollis*) in Sweden. In a later study from the same population, Merilä and Sheldon (2000) found significant additive genetic variance in lifetime reproductive success for females and not significant nonzero genetic variance for males (see also Brommer et al. (2007)). In female and male British great tits (*Parus major*), McCleery et al. (2004) found not significant close-to-zero heritability of lifetime reproductive success. Along the same lines, Wheelwright et al. (2014) found a zero heritability for lifetime reproductive success in female savannah sparrows (*Passerculus sandwichensis*) in Canada, while Teplitsky et al. (2009) found no-significant nonzero genetic variance in lifetime reproductive success for females and zero variance for males in a natural population of red-billed gulls (*Laurus novaehollandiae*) in New Zealand. Finally, de Villemereuil et al. (2019) showed that hihis (*Notiomystis cincta*) in New Zealand had negligible additive genetic variance in lifetime fitness, while Wolak et al. (2018) found that the song sparrows (*Melospiza melodia*) of Mandarte island in Canada harbour substantial additive genetic variance in female and male fitness.

Data constraints might partially explain the paucity of studies testing for the heritability of fitness in the wild and the heterogeneity among estimates of additive genetic variance, although steadily increasing datasets collected from long-term study populations gradually alleviate the problem (Clutton-Brock and Sheldon 2010). This increased data availability was recently accompanied by the development of (i) statistical tools designed to deal with the non-Gaussian distributions that often characterize fitness data (de Villemereuil et al. 2016; de Villemereuil 2018), as well as (ii) theoretical frameworks that facilitate the evolutionary inference of quantitative genetic parameters based on these data distributions (Morrissey and Bonnet 2019). Fitness components often follow a non-Gaussian error distribution as a result from the temporal sequence of survival (Binary data) and reproductive events (Poisson data). Modelling them accordingly, by applying generalized linear models (de Villemereuil et al. 2016; de Villemereuil 2018; Bonnet et al. 2019), offers an added benefit. Parameter estimates from a model with a Poisson error distribution fitted to absolute fitness data readily inform about the increase in fitness within a generation, while back-transformed estimates on the observed data scale inform about the increase in fitness between generations (Morrissey and Bonnet 2019). As such, estimates of the additive genetic variance for absolute fitness on the latent scale data are equivalent to evolvability estimates (i.e., the additive genetic variance in relative fitness) directly on the data scale for relative fitness, and, therefore, provide evidence for Fisher's rate of evolution (Hansen et al. 2011; de Villemereuil et al. 2016). To date, however, only four studies have modelled the quantitative genetics of fitness in wild populations assuming a Poisson (McFarlane et al. 2014, 2015; Wolak et al. 2018) or a zero-Inflated Poisson distribution (de Villemereuil et al. 2019). Now is the time to make use of the recent statistical tools and theoretical frameworks (de Villemereuil et al. 2016; Bonnet et al. 2019;

Morrissey and Bonnet 2019) to evaluate the existence and magnitude of additive genetic variance in fitness in natural populations.

Here, we present phenotypic and pedigree data obtained from a 28-year individual-based study on a serially monogamous and migratory seabird, the common tern (*Sterna hirundo*). Applying a series of "animal models" to data from almost 6000 pedigreed individuals, we investigate additive genetic variance for lifetime fitness (assessed as the total number of fledglings produced by a locally-born fledgling), and its underlying components: juvenile survival and adult lifetime reproductive success, with the latter again being decomposed in annual reproductive success and annual adult survival.

METHODS

Study System

Fitness and pedigree data were collected between 1992 and 2019 as part of a long-term study of a common tern population located at the Banter See on the German North Sea coast (53°36´N, 08°06´E). Common terns from this population spend their winters in western Africa and return to the breeding colony in early spring to breed or prospect potential breeding locations (Becker and Ludwigs 2004). The Banter See colony consists of six concrete islands, each of which is surrounded by a 60-cm wall. Walls are equipped with 44 elevated platforms, each containing an antenna which reads transponder codes. The individual-based study at the Banter See was initiated in 1992, when 101 adult birds were caught and marked with individually numbered subcutaneously injected transponders. Since 1992, all locally hatched birds are similarly marked with a transponder shortly before fledging and the presence and reproductive performance of marked individuals is monitored following a standard protocol (Becker and Wendeln 1997). As part of this protocol, the

colony is checked for new clutches every 2–3 days throughout the breeding season (Zhang et al. 2015). Parents are identified using portable antennae placed around each nest for 1–2 days during incubation, which is shared by both partners. Pairs can rear up to three chicks per brood (mean successful brood size 0.41 ± 0.65 SD chicks), and can produce replacement clutches after loss of eggs or chicks. True second clutches are extremely rare (Becker and Zhang 2011).

Fitness Data

Fitness data have been collected since 1992, with data up to 2019 being available for the analyses reported here. Our initial data selection included individuals that fledged between 1992 and 2016, because previous work showed that 97% of fledglings, if they returned, did so within the first 3 years (Vedder and Bouwhuis 2018). Although we cannot directly observe an individual's death, we can reliably assume it, because adult breeders at the Banter See are highly site-faithful, evidenced by the resighting probability of individuals that bred at least once being close to one (Szostek and Becker 2012), and 96% of breeders not skipping recording by the antenna system for two or more consecutive years after first reproduction (Bouwhuis et al. 2015; Zhang et al. 2015). Based on this knowledge, we removed all birds that were observed in 2018 and/or 2019 *and* were younger than 11 years old, because (i) they are known to not be, or cannot yet be assumed to be, dead, and (ii) lifetime fitness of individuals older than 10 years and those dead showed a high correlation ($r > 0.8$) in our dataset. Hence, we included birds that have completed their life histories ($n = 5836$), as well as birds that were still alive but older than 10 years ($n = 163$) to avoid introducing a cohort truncation bias by non-randomly removing longer-lived birds (Hadfield 2008; Morrissey et al. 2012). To control for any potential confounding effect, we modelled whether an individual was considered dead or alive as a fixed effect (see below).

We quantified lifetime fitness as the total number of local fledglings that a locally-hatched fledgling produced during its lifetime. In total, our data comprise the fitness of 5999 locally-hatched fledglings (Fig 1A). It can be decomposed in two major components: juvenile survival and adult lifetime reproductive success. Juvenile survival captures survival from fledgling to adulthood, whereas adult lifetime reproductive success captures adult survival and reproductive investment across life. These two fitness components correspond to the two mechanisms capturing the Zero-inflated Poisson distribution of lifetime fitness, and hence, we did not model them in two separated analyses. However, we further decomposed adult lifetime reproductive success (LRS) into its two components: annual reproductive success (ARS) and annual adult survival (AAS). ARS was measured as the number of fledglings that an individual produced each year between its first reproduction and last registration, assigning zeroes for years of skipped reproduction or registration (Fig 1B); AAS was adult survival (1/0) to the following breeding season. In total, our data for LRS comprised 793 individuals with 4453 and 5290 observations for ARS and AAS, respectively. The difference in the number of observations between ARS and AAS represents values of future-breeder prospecting individuals (i.e., detected alive but not yet breeding).

Pedigree

The pedigree was constructed by assigning all fledged offspring to their social parents, then pruned to remove individuals who are either not phenotyped or not ancestors to phenotyped individuals. For the purpose of this study, the pedigree comprised 6273 records. The maximum depth was five generations, the number of paternities and maternities 2509 and 2414, respectively. The numbers of full, maternal and paternal sibs were 2566, 10180 and 9718, respectively. This social pedigree

is a good approximation of the genetic pedigree, because common terns exhibit very low levels of extra-pair paternity (González-Solís et al. 2001).

Quantitative Genetic Models

We applied an animal model approach that combines the phenotypic information on individual fitness with information from the social pedigree (Kruuk 2004). As such, we fitted a series of univariate animal models where fitness, or one of its components, was the response variable.

To model lifetime fitness, we fitted a univariate animal model with a zero-inflated Poisson error distribution. We fitted random intercepts for individual identity linked to the pairwise relatedness matrix and for hatch-year (to account for cohort effects; e.g. Vedder & Bouwhuis 2018). Because we modeled lifetime fitness with a Zero-Inflated over-dispersed Poisson distribution, this model has a zero-inflated and a Poisson component, allowing us to explicitly estimate the covariance between the two components for each random effect. However, a model including additive genetic and hatch-year correlations between the zero-inflated and Poisson components of the trait did not provide a better fit to the data, because of which we did not model such correlations. As fixed effects, we modelled the trait intercept and whether the individual was alive or dead (categorical variable with two levels). Additionally, we performed a data simulation analysis to investigate whether we can effectively detect *small, but substantial* heritabilities and evolvabilities (*sensu* de Villemereuil et al. 2019) given our data and pedigree structure (see further details in Supplementary Material).

To model ARS, we assumed a Poisson error distribution with a log link function. We fitted random intercepts for individual identity linked to the pairwise relatedness matrix, individual identity not linked to the pedigree (to account for permanent environmental effects) and year of

observation (to account for temporal variation across years). As fixed effects, we modelled the trait intercept and age as a linear predictor (mean centered and variance standardized by subtracting the mean and dividing by the standard deviations), as fledgling production is known to linearly increase with age (Zhang et al. 2015).

To model AAS, we assumed a binary error distribution and fixed the residual variance to one. We fitted random intercepts for individual identity linked to the pairwise relatedness matrix, individual identity not linked to the pedigree (to account for permanent environmental effects) and year of observation (to account for temporal variation across years). As fixed effects, we modelled the trait intercept and age as linear predictor (mean centered and variance standardized by subtracting the mean and dividing by the standard deviations), as AAS is known to linearly decrease with age (Zhang et al. 2015; Vedder et al. 2021).

All quantitative genetic models were fitted using a Bayesian framework implemented in the statistical software R (v. 3.6.1, R Core Team 2019) using the R-packages *MCMCglmm* (Hadfield 2010) and *QGglmm* (de Villemereuil et al. 2016). Heritabilities (h^2) were conditional to the variance explained by fixed effects and estimated as the proportion of the total phenotypic variance explained by the additive genetic variance. Evolvabilities (I_A) were estimated by dividing the additive genetic variance by the squared population mean (Houle 1992; Hansen et al. 2011). Although we estimated the evolvability for AAS, care is required when drawing conclusions, because the population mean is regarded as arbitrary for binomial traits.

For all models we used parameter-expanded priors with an inverse Gamma distribution (Hadfield 2010). We fitted different priors for each fitness component (see Supplementary Material). The number of iterations and thinning intervals were chosen for each model so as to ensure that the minimum MCMC effective sample size for all parameters was 1000. Burn-in was

set to a minimum of 5000 iterations. The retained effective sample sizes yielded absolute autocorrelation values <0.1 and satisfied convergence criteria based on the Heidelberger and Welch convergence diagnostic (Heidelberger and Welch 1981). We drew inferences from the marginal posterior mode, mean and 95% credible intervals (95% CI) as inferences drawn from posterior modes and means might differ, particularly when model parameters are near a boundary (e.g., variance near zero). Additionally, we drew inferences on the importance of genetic variances and associated metrics based on criteria used by de Villemereuil et al. (2019): we considered an effect to be "small but likely non-zero" when an evolvability exceeded 0.01 for Poisson distributed traits (or Poisson component of fitness), and when a heritability exceeded 0.1 for binomial distributed traits (or zero-inflated component of fitness). Variance parameters were estimated on latent scales. Hence, to facilitate evolutionary inference (Bonnet et al. 2019; Morrissey and Bonnet 2019), we back-transformed the latent-scale posterior distributions of the quantitative genetic parameters to the observed data-scale (de Villemereuil et al. 2016).

RESULTS

Quantitative Genetics of Lifetime Fitness

Among the 5999 common tern chicks that fledged between 1992 and 2016, lifetime fitness ranged between 0 and 29 fledglings (Fig. 1A). 5231 (87.19%) fledglings obtained zero fitness, such that the distribution of fitness was strongly zero-inflated (Fig. 1A). Raw mean fitness was 0.72 ± 2.52 SD fledglings.

Data simulations showed that, given our data structure and pedigree, we would be able to detect what might be considered a *small, but substantial* signal for the Zero-inflated component of lifetime fitness (Figure S1): we generated a Zero-inflated component of fitness with a heritability (h^2) of 0.1 *(sensu* de Villemereuil et al. 2019), and found that the posterior modes and means

accurately estimated the simulated value of h^2 (average of 0.111 across the 100 simulations for the posterior mode and mean, see Supplementary Material). Our quantitative genetic analysis, however, suggested nominally zero additive genetic variance in the zero-inflated component of fitness, as the posterior mode and mean of the additive genetic variance were both in agreement and very close to zero (Table 1, Fig. 2E-F).

The results for the Poisson component of lifetime fitness are less straightforward. Our analysis suggested small but likely non-zero evidence for the additive genetic variance. The posterior modes and means of V_A, h^2 and I_A of the Poisson component of fitness were nonzero (Table 1, Fig. 2A-C), but the associated lower 95% CI limits of V_A, h^2 and I_A converged towards zero. Data simulations showed that, given our data structure and pedigree, we would not be able to detect what might be considered a *small, but substantial* signal for the Poisson component of fitness: we generated a Poisson-component of fitness with an evolvability (I_A) of 0.010, and found that posterior means estimated an I_A of similar magnitude to the simulated value (average of 0.011 across the 100 simulations), while posterior modes largely underestimated I_A (average of 0.002 across simulations; see Supplementary Material, Fig. S1). However, this difference between the estimated posterior means and modes likely represents an artifact due to the asymmetric nature of the posterior distribution. With insufficient data for estimating variance components (due to low sample size or poor pedigree structure), posterior distributions close to the boundary (i.e., zero in this case) tend to broaden and the posterior mean tends to shift away from the boundary, resulting in an increased posterior mean value (He and Hodges 2008). To test for this potential artifact, we ran the same data simulations described above, but with a Poisson-component of lifetime fitness with an evolvability (I_A) of 0.00. We found very similar patterns (see Supplementary Material), supporting the notion that the difference between posterior means and modes was not biologically

informative. We also simulated data with a larger value of I_A (0.10). We found that posterior modes and means tended to accurately estimate the simulated value of I_A (average of ~0.135, see Supplementary Material). Altogether, our data simulations indicated that we would have power to detect larger values of additive genetic variance for the Poisson component of lifetime fitness, but we did not have sufficient power to detect a *small, but substantial* signal. As such, our analyses suggested there to be low to zero additive genetic variance in lifetime fitness, but we lack power to determine whether such variance is nominally zero or very small.

Quantitative Genetics of Fitness Components

We investigated the ARS and AAS of 793 fledglings that survived to adulthood and bred in our population (Table 2). Raw mean annual reproductive success was 0.70 ± 0.81 SD with a maximum of 3 fledglings (Fig. 1B). The posterior distribution of V_A for ARS showed a clear peak of density away from zero, with posterior modes and means of V_A, h^2 and I_A largely in agreement. The lower 95% CI, however, converged toward zero (Table 2, Fig. 3A-C), suggesting again a lack of power to detect the additive genetic variance in ARS with higher precision. Raw mean annual adult survival probability was 0.85 ± 0.36 SD. The posterior modes and means of all quantitative genetic parameters for AAS were very close to zero and in agreement (Table 2, Fig. 4A-C), with the lower 95% CI limit of all parameter estimates converging towards zero.

DISCUSSION

The most direct measure of the adaptive potential of a population is its standing additive genetic variance in fitness (Fisher 1930). Here, we estimated additive genetic variances in lifetime fitness and two of its key components in a wild population of common terns. Our findings indicated little

evidence for additive genetic variance in juvenile or adult survival in this population, while we may have detected non-zero additive genetic variance in annual reproductive success. However, simulations revealed a lack of statistical power, whereby most of our estimates were deemed little reliable. As such, our work demonstrated, once again (see early statement from Burt (1995)), that estimating additive genetic variance in fitness is very difficult in wild populations, partly due to the expected low values of genetic variation in fitness, and partly due to the challenges associated with collecting sufficient phenotypic and pedigreed data. We see this as an encouragement to sustain long-term monitoring programs with valuable fitness estimations.

Quantitative Genetics of Lifetime Fitness

In this study, we aimed at providing much-needed empirical estimates for key quantitative genetic parameters that have rarely been estimated in the wild, and did so by applying non-Gaussian models to estimate variation in fitness (Bonnet et al. 2019; Morrissey and Bonnet 2019). The exercise was particularly insightful because quantitative genetic parameters drawn from Poisson models can be readily interpreted in terms of evolutionary significance without back-transformation. Estimates of additive genetic variance for absolute fitness on the latent scale are equivalent to evolvability estimates directly on the data scale for relative fitness, and therefore, they provide evidence for Fisher's rate of evolution (Hansen et al. 2011; de Villemereuil et al. 2016; Morrissey and Bonnet 2019). Variance estimates on the latent-scale are insightful in terms of evolutionary inference, however, the data-scale is the only observable scale, and, therefore, of most direct interest.

There have been around 30 studies testing for additive genetic variance in fitness in the wild (see Introduction), with, to our knowledge, only four using non-Gaussian animal models

(McFarlane et al. 2014, 2015; Wolak et al. 2018; de Villemereuil et al. 2019), and only one testing for variance components of fitness using a zero-inflated Poisson distribution (de Villemereuil et al. 2019). Our estimate of the additive genetic variance for the Zero-inflated component of common tern lifetime fitness on the observed data-scale was very small (posterior mean and mode $V_{A\ data\text{-}scale} \sim 0.004$, Table 1), similarly to results for the hihi (de Villemereuil et al. 2019) (posterior median and mode $V_{A\ data\text{-}scale} \sim 0$). For the Poisson component, de Villemereuil et al. (2019) found a posterior median $V_{A\ data\text{-}scale}$ of 0.73 and posterior mode of 0.0078. Unfortunately, however, our estimate was not very reliable, such that there is no added value in comparing the two. Given that estimates of additive genetic variance in fitness can be readily interpreted as evidence for a potential for increased fitness between generations (Bonnet et al. 2019; Morrissey and Bonnet 2019), our findings of very low (or nominally zero) values of additive genetic variance for lifetime fitness imply that the adaptive potential of this wild population of common terns will be extremely limited.

Our estimate of the heritability of the Zero-inflated component of lifetime fitness (posterior mean and mode $h^2_{\ data\text{-}scale} \sim 0.03$, Table 1) was somewhat greater than that reported by de Villemereuil et al. (2019) (posterior median $h^2_{\ data\text{-}scale} = 0.003$ for Zero-inflated part). It is, however, important to highlight that back-transforming heritability estimates from the latent- to the data-scale will generally lead to very small estimates. This is because the data-scale heritability not only contains additive genetic variance but also distribution noise associated with Poisson or Bernouilli processes (de Villemereuil et al. 2016; Bonnet et al. 2019; Morrissey and Bonnet 2019). Hence, heritability estimates should be interpreted with care when making cross-study comparisons, and, overall, heritability of fitness is not a good metric to assess the adaptive potential of a population. Indeed, evolvability, i.e., additive genetic variance in relative fitness, is a more

informative metric for evolutionary potential (Fisher 1930; Queller 2017; Morrissey and Bonnet 2019). Notably, estimates of additive genetic variance for absolute fitness on the latent scale data are equivalent to evolvability estimates directly on the data scale for relative fitness.

Quantitative Genetics of Fitness Components

Additive genetic variance in lifetime fitness can theoretically be decomposed into the additive genetic variances in its underlying components. The two primary components of our measure of lifetime fitness are juvenile survival and adult lifetime reproductive success. Our zero-inflation in lifetime fitness is mainly due to low juvenile survival (i.e., 74% of fledglings did not locally recruit), while the Poisson process generating the observed fitness distribution is mostly capturing adult lifetime reproductive success. If we compare our nominally zero additive genetic variance in the Zero-inflated component of lifetime fitness (Table 1), with estimates from other studies that tested for additive genetic variance in juvenile survival, we observe some differences. For instance, the study of Wolak et al. (2018) on the song sparrow population of Mandarte Island reported evidence of non-zero V_A for juvenile survival. The natural history of common terns and song sparrows differ in many ways, yet one reason for this disparity could be a difference in emigration rates since the Mandarte population is isolated with very little juvenile emigration (Reid et al. 2021).

Adult lifetime reproductive success is the sum of annual reproductive events across the life of an individual, and hence, can be decomposed into annual reproductive success and annual adult survival. Given the lack of substantial additive genetic variance for adult annual survival (Table 2), the decomposition of adult lifetime reproductive success into its components revealed that annual reproductive success was the main mechanism underlying genetic variation in adult lifetime

reproductive success. This pattern resembles the one from Mandarte's song sparrows, where quantitative genetics analyses also demonstrated moderate levels of V_A in ARS (especially for males) and close to zero V_A in AAS, indicating that heritable ARS was the primary component of heritable adult LRS also in that natural population (Wolak et al. 2018). In the case of the common terns, the lack of additive genetic variance in annual adult survival appears in line with their life history, as common terns are relatively long-lived seabirds (that show an average and maximum individual lifespan in this dataset of 8 and 23 years, respectively), generally favoring (canalization of) survival over reproduction (Vedder et al. 2017, 2019).

Limitations of studying quantitative genetics of fitness in the wild

Despite the fundamental relevance of knowledge on additive genetic variance in fitness in the context of understanding adaptation and evolutionary potential, Hendry et al. (2018) found that there were very few estimates of additive genetic variance for fitness in the wild, and that those estimates available were heterogeneous, with many estimates close to zero, and very few large estimates (e.g., Gustafsson 1986; Kruuk et al. 2000; Merilä and Sheldon 2000; Coltman et al. 2005; McFarlane et al. 2014).

Data constraints might partially explain the paucity of studies testing for the heritability of fitness in the wild. Animal models are data-hungry and rely on high quality pedigree information. Researchers therefore are faced with the challenge of collecting hard-to-quantify lifetime fitness data from an unbiased sample of the population (i.e., avoiding a "missing fraction" bias) that comprises a sufficiently large number of individuals of known relatedness (Burt 1995; Merilä and Sheldon 1999; Hendry et al. 2018). In addition, even when a large dataset and pedigree are available, additive genetic variance in fitness is generally expected to be low (as theoretically

predicted by Fishers' natural theorem), such that the power to detect small, close to zero, additive genetic variation in fitness may be low as well. Non-zero but non-significant estimates and zero estimates might simply represent bounded estimates (i.e., there is a border effect, preventing to estimate very low values with precision). As pointed out by Burt (1995), "it is very difficult to get an estimate that is statistically distinguishable from zero, and the sample sizes required to do so might easily lead to despair". In light of the multiple constraints posed by data requirements and expected low values, negative results with respect to additive genetic variation in fitness should be taken and discussed with care. Data simulations aiming at determining the statistical power of a given dataset and pedigree structure, will help to distinguish a true negative result from an unreliable zero estimate (e.g., de Villemereuil et al. 2019). Overall, the field of quantitative genetics in the wild needs more and better estimates stemming from a broad taxonomic range, and to systematically associated those with simulations to assess the power of the dataset.

Conclusion

Our quantitative genetic study of fitness in a wild population of common terns revealed low and unreliable, to zero levels of additive genetic variance in lifetime fitness and two underlying components. Those analyses, however, were overshadowed by a lack of statistical power to detect additive genetic variation in fitness more accurately and precisely. The continuation of long-term individual-based studies should be safeguarded (also see Clutton-Brock and Sheldon 2010), such that the maturation of long-term studies will offer improved opportunities for testing genetic variation in natural populations, which, thanks to the recent development of appropriate statistical and theoretical frameworks (de Villemereuil et al. 2016; Bonnet et al. 2019; Morrissey and Bonnet 2019), will help to improve our understanding of the genetics of fitness in the wild. Ultimately, a

robust quantification of the standing additive genetic variation for fitness will inform us about the rate of adaptation of populations between and within generations, and allow a better understanding of their viability in the face of the deleterious environmental effects that current climate and global changes pose.

AUTHOR CONTRIBUTIONS

M.M. conceived the study with input from S.B. and A.C. M.M. designed and conducted the analyses, and wrote the manuscript. S.B. manages the tern data and collated the dataset. All authors contributed to editing the final paper.

ACKNOWLEDGMENTS

We are indebted to Peter H. Becker for setting up the long-term common tern study and thank the numerous field workers that have contributed to compiling the data. We also thank Tim Janicke for providing enlightening discussions during manuscript conception and preparation, and Timothée Bonnet for helping with the interpretation of the model parameters. M.M. was funded by a Marie Curie Individual Fellowship (PLASTIC TERN, Grant Agreement Number 793550), A.C. by the European Research Council (Starting grant ERC-2013-StG-337365-SHE). Authors declare no conflict of interest.

DATA ARCHIVING

Data will be archived in the Dryad Digital Repository upon acceptance.

REFERENCES

Becker, P. H., and J.-D. L. Ludwigs. 2004. Sterna hirundo Common Tern. Pp. 93–139 *in* Parkin D, ed. The Birds of the Western Palearctic Update. Oxford University Press, Oxford, UK.

Becker, P. H., and H. Wendeln. 1997. A New Application for Transponders in Population Ecology of the Common Tern. Condor 99:534–538.

Becker, P. H., and H. Zhang. 2011. Renesting of Common Terns Sterna hirundo in the life history perspective. J. Ornithol. 152:213–225.

Bonnet, T., M. B. Morrissey, and L. E. B. Kruuk. 2019. Estimation of Genetic Variance in Fitness, and Inference of Adaptation, When Fitness Follows a Log-Normal Distribution. J. Hered. 110:393–395.

Bouwhuis, S., O. Vedder, and P. H. Becker. 2015. Sex-specific pathways of parental age effects on offspring lifetime reproductive success in a long-lived seabird. Evolution (N. Y). 69:1760–1771.

Brommer, J. E., M. Kirkpatrick, A. Qvamström, and L. Gustafsson. 2007. The intersexual genetic correlation for lifetime fitness in the wild and its implications for sexual selection. PLoS One 2:1–6.

Burt, A. 1995. The evolution of fitness. Evolution. Evolution (N. Y). 49:1–8.

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

Coltman, D. W., P. O'Donoghue, J. T. Hogg, and M. Festa-Bianchet. 2005. Selection and genetic (co)variance in bighorn sheep. Evolution (N. Y). 59:1372–1382.

de Villemereuil, P. 2018. Quantitative genetic methods depending on the nature of the

440 phenotypic trait. Ann. N. Y. Acad. Sci. 1422:29–47.
441 de Villemereuil, P., A. Rutschmann, K. D. Lee, J. G. Ewen, P. Brekke, and A. W. Santure. 2019.
442 Little Adaptive Potential in a Threatened Passerine Bird. Curr. Biol. 29:889-894.e3.
443 de Villemereuil, P., H. Schielzeth, S. Nakagawa, and M. Morrissey. 2016. General methods for
444 evolutionary quantitative genetic inference from generalized mixed models. Genetics
445 204:1281–1294.
446 Ellegren, H., and B. C. Sheldon. 2008. Genetic basis of fitness differences in natural populations.
447 Nature 452:169–175.
448 Fisher, R. 1930. The Genetical Theory of natural Selection. Oxford.
449 Gomulkiewicz, R., and R. G. Shaw. 2013. Evolutionary rescue beyond the models. Philos. Trans.
450 R. Soc. B Biol. Sci. 368.
451 González-Solís, J., E. Sokolov, and P. H. Becker. 2001. Courtship feedings, copulations and
452 paternity in common terns, Sterna hirundo. Anim. Behav. 61:1125–1132.
453 Gustafsson, L. 1986. Lifetime Reproductive Success and Heritability: Empirical Support for
454 Fisher's Fundamental Theorem. Am. Nat. 128:761–764.
455 Hadfield, J. 2010. MCMC Methods for Multi-response Generalized Linear Mixed Models : The
456 MCMCglmm R Package. J. Stat. Softw. 33:1–22.
457 Hadfield, J. D. 2008. Estimating evolutionary parameters when viability selection is operating.
458 Proc. R. Soc. B Biol. Sci. 275:723–734.
459 Hansen, T. F., C. Pélabon, and D. Houle. 2011. Heritability is not Evolvability. Evol. Biol.
460 38:258–277.
461 He, Y., and J. S. Hodges. 2008. Point estimates for variance-structure parameters in Bayesian
462 analysis of hierarchical models. Comput. Stat. Data Anal. 52:2560–2577.

463 Heidelberger, P., and P. D. Welch. 1981. A spectral method for confidence interval generation
464 and run length control in simulations. Commun. ACM 24:233–245.
465 Hendry, A. P., D. J. Schoen, M. E. Wolak, and J. M. Reid. 2018. The Contemporary Evolution of
466 Fitness. Annu. Rev. Ecol. Evol. Syst. 49:457–476.
467 Houle, D. 1992. Comparing evolvability and variability of quantitative traits. Genetics 130:195–
468 204.
469 Houle, D., B. Morikawa, and M. Lynch. 1996. Comparing mutational variabilities. Genetics 143.
470 Jones, J. S. 1987. The heritability of fitness: Bad news for 'good genes'? Trends Ecol. Evol.
471 2:35–38.
472 Kruuk, L. E. B. 2004. Estimating genetic parameters in natural populations using the "animal
473 model." Philos. Trans. R. Soc. B Biol. Sci. 359:873–890.
474 Kruuk, L. E. B., T. H. Clutton-Brock, J. Slate, J. M. Pemberton, S. Brotherstone, and F. E.
475 Guinness. 2000. Heritability of fitness in a wild mammal population. Proc. Natl. Acad. Sci.
476 97:698–703.
477 McCleery, R. H., R. A. Pettifor, P. Armbruster, K. Meyer, B. C. Sheldon, and C. M. Perrins.
478 2004. Components of Variance Underlying Fitness in a Natural Population of the Great Tit
479 Parus major. Am. Nat. 164:E62–E72.
480 McFarlane, S. E., J. C. Gorrell, D. W. Coltman, M. M. Humphries, S. Boutin, and A. G.
481 Mcadam. 2014. Very low levels of direct additive genetic variance in fitness and fitness
482 components in a red squirrel population. Ecol. Evol. 4:1729–1738.
483 McFarlane, S. E., J. C. Gorrell, D. W. Coltman, M. M. Humphries, S. Boutin, and A. G.
484 McAdam. 2015. The nature of nurture in a wild mammal's fitness. Proc. R. Soc. B Biol.
485 Sci. 282:20142422.

Merilä, J., and B. C. Sheldon. 1999. Genetic architecture of fitness and nonfitness traits: Empirical patterns and development of ideas.

Merilä, J., and B. C. Sheldon. 2000. Lifetime reproductive success and heritability in nature. Am. Nat. 155:307–310.

Morrissey, M. B., and T. Bonnet. 2019. Analogues of the fundamental and secondary theorems of selection, assuming a log-normal distribution of expected fitness. J. Hered. 110:396–402.

Morrissey, M. B., D. J. Parker, P. Korsten, J. M. Pemberton, L. E. B. Kruuk, and A. J. Wilson. 2012. The prediction of adaptive evolution: Empirical application of the secondary theorem of selection and comparison to the breeder's equation. Evolution (N. Y). 66.

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

Queller, D. C. 2017. Fundamental Theorems of Evolution. Am. Nat. 189:345–353.

R Core Team. 2019. A language and environment for statistical computing. R Found. Stat. Comput. Vienna:Austria.

Reid, J. M., P. Arcese, P. Nietlisbach, M. E. Wolak, S. Muff, L. Dickel, and L. F. Keller. 2021. Immigration counter-acts local micro-evolution of a major fitness component: Migration-selection balance in free-living song sparrows. Evol. Lett. 5:48–60.

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

Shaw, R. G., and F. H. Shaw. 2014. Quantitative genetic study of the adaptive process. Heredity (Edinb). 112:13–20.

Szostek, K. L., and P. H. Becker. 2012. Terns in trouble: demographic consequences of low breeding success and recruitment on a common tern population in the German Wadden Sea. J. Ornithol. 153:313–326.

509 Teplitsky, C., J. A. Mills, J. W. Yarrall, and J. Merilä. 2009. HERITABILITY OF FITNESS
510 COMPONENTS IN A WILD BIRD POPULATION. Evolution (N. Y). 63:716–726.
511 Vedder, O., and S. Bouwhuis. 2018. Heterogeneity in individual quality in birds: overall patterns
512 and insights from a study on common terns. Oikos 127:719–727.
513 Vedder, O., I. Pen, and S. Bouwhuis. 2021. How fitness consequences of early-life conditions
514 vary with age in a long-lived seabird: A Bayesian multivariate analysis of age-specific
515 reproductive values. J. Anim. Ecol. 1365-2656.13471.
516 Vedder, O., S. Verhulst, C. Bauch, and S. Bouwhuis. 2017. Telomere attrition and growth: a life-
517 history framework and case study in common terns. J. Evol. Biol. 30:1409–1419.
518 Vedder, O., H. Zhang, A. Dänhardt, and S. Bouwhuis. 2019. Age-Specific Offspring Mortality
519 Economically Tracks Food Abundance in a Piscivorous Seabird. Am. Nat. 193:588–597.
520 Walsh, B., and M. W. Blows. 2009. Abundant Genetic Variation + Strong Selection =
521 Multivariate Genetic Constraints: A Geometric View of Adaptation. Annu. Rev. Ecol. Evol.
522 Syst. 40:41–59.
523 Wheelwright, N. T., L. F. Keller, and E. Postma. 2014. The effect of trait type and strength of
524 selection on heritability and evolvability in an island bird population. Evolution (N. Y).
525 68:3325–3336.
526 Wolak, M. E., P. Arcese, L. F. Keller, P. Nietlisbach, and J. M. Reid. 2018. Sex-specific additive
527 genetic variances and correlations for fitness in a song sparrow (Melospiza melodia)
528 population subject to natural immigration and inbreeding. Evolution (N. Y). 72:2057–2075.
529 Zhang, H., O. Vedder, P. H. Becker, and S. Bouwhuis. 2015. Age-dependent trait variation: the
530 relative contribution of within-individual change, selective appearance and disappearance in
531 a long-lived seabird. J. Anim. Ecol. 84:797–807.

FIGURES

Figure 1. Phenotypic distributions of A) lifetime fitness measured as the total number of fledglings a locally-hatched fledgling produced in its lifetime, and B) annual reproductive success, measured as the number of fledglings an adult breeder produced in a year.

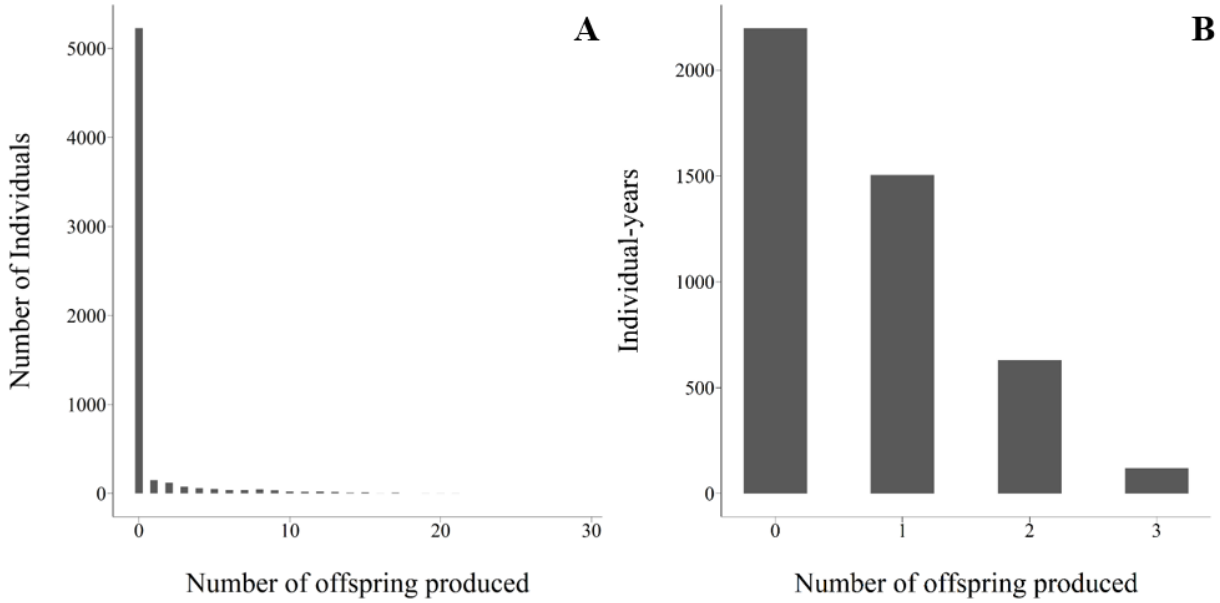


Figure 2. Posterior MCMC samples (bars), kernel density estimation (solid black line), posterior mean (red dotted line), and 95% Credible Intervals (black dashed lines) for the A) additive genetic variance (V_A), B) heritability (h^2) and C) evolvability (I_A) of the Poisson component of lifetime fitness, and the D) additive genetic variance (V_A), E) heritability (h^2) and F) evolvability (I_A) of the zero–inflated component of lifetime fitness. Distributions are reported on the observed data scale.

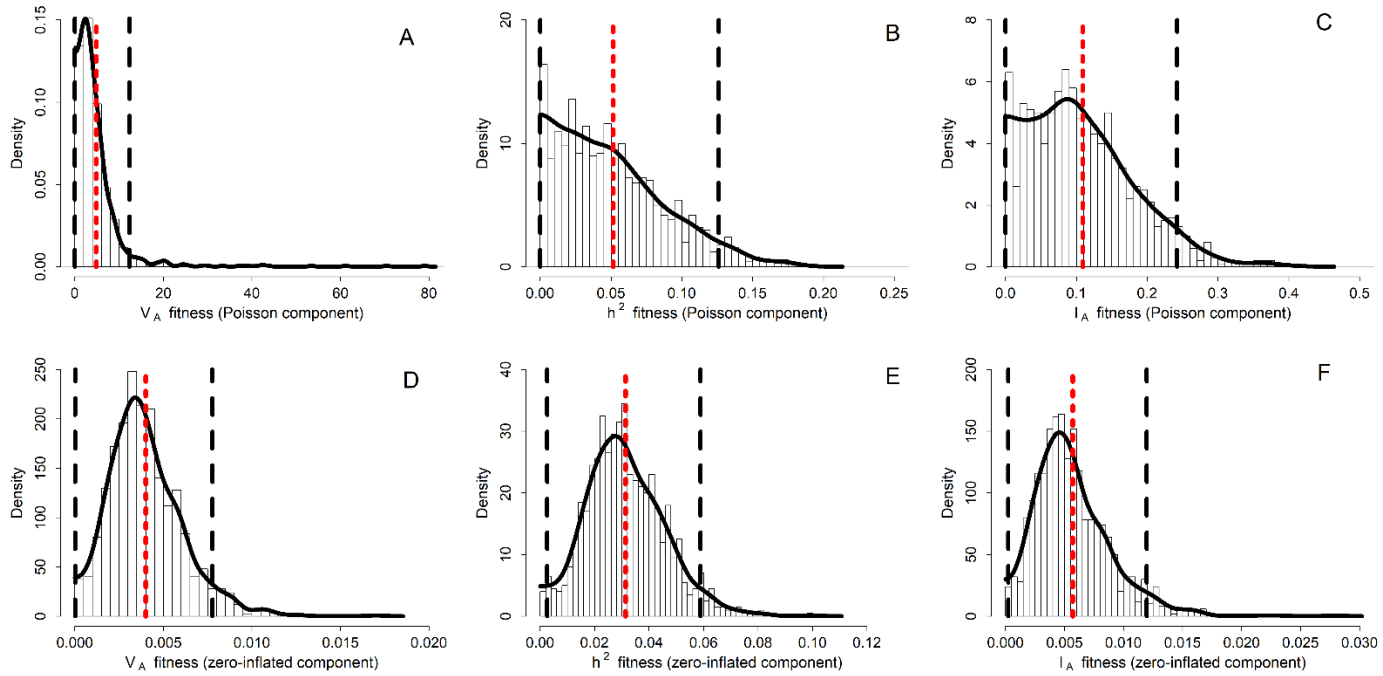
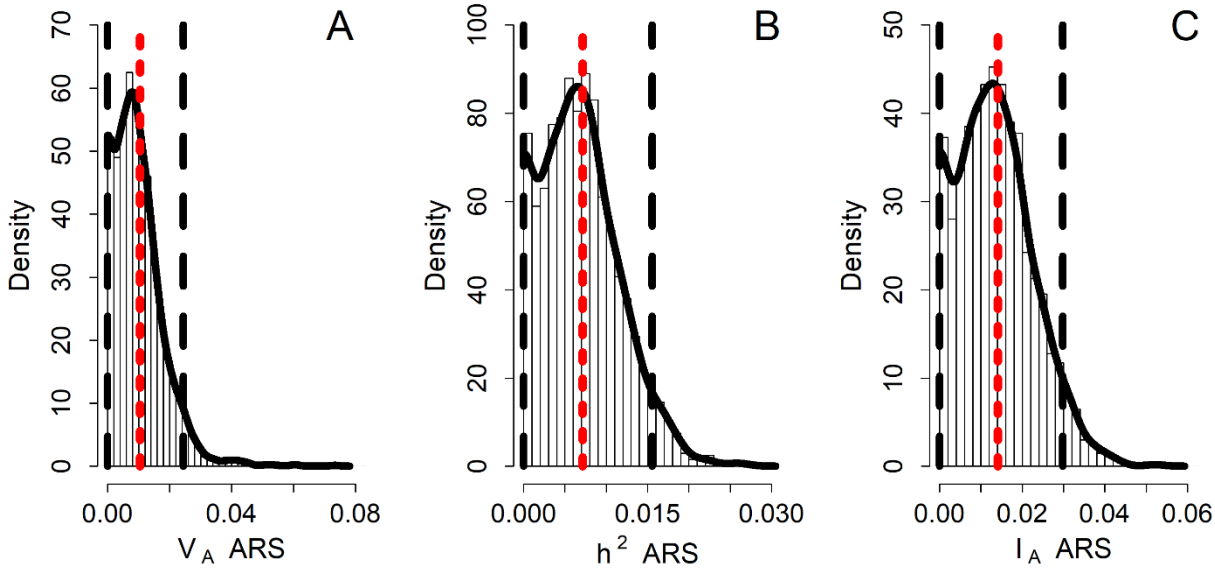
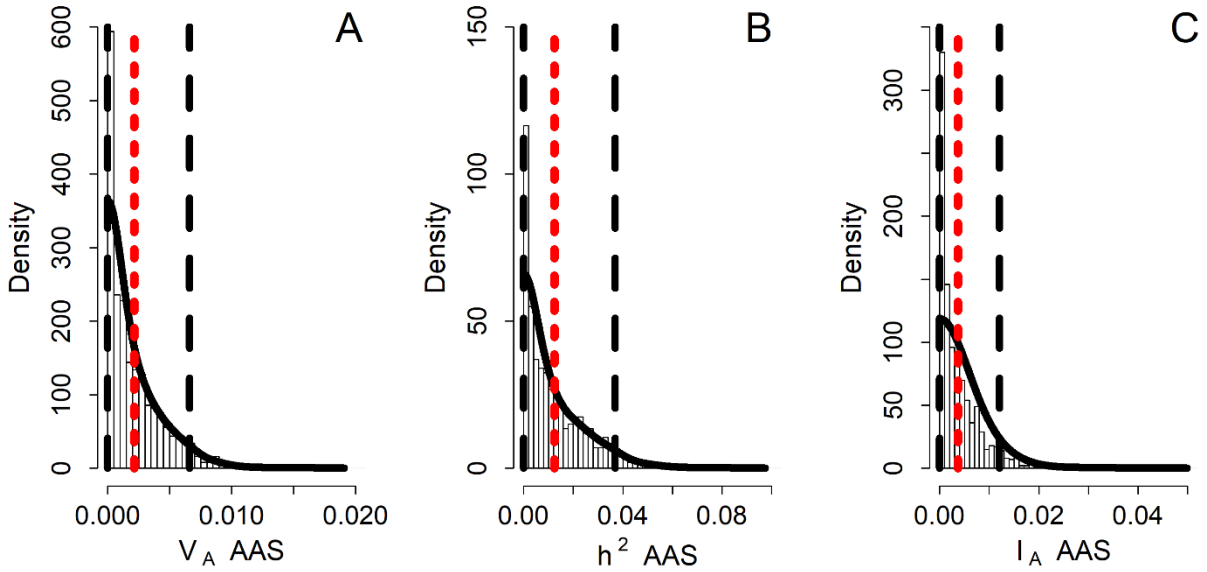


Figure 3. Posterior MCMC samples (bars), kernel density estimation (solid black line), posterior mean (red dotted line), and 95% Credible Intervals (black dashed lines) for the A) additive genetic variance (V_A), B) heritability (h^2) and C) evolvability (I_A) of annual reproductive success (ARS). Distributions are reported on the observed data scale.



548 **Figure 4.** Posterior MCMC samples (bars), kernel density estimation (solid black line), posterior
549 mean (red dotted line), and 95% Credible Intervals (black dashed lines) for the A) additive genetic
550 variance (V_A), B) heritability (h^2) and C) evolvability (I_A) of annual adult survival (AAS).
551 Distributions are reported on the observed data scale.



552
553
554
555
556
557
558
559
560
561

TABLES

Table 1. Posterior modes, means and 95% Credible Intervals (in brackets) for observed data-scale variance estimates from the quantitative genetic analyses of lifetime fitness.

Model component	Pop. Mean	V_P	V_A	h^2	I_A
Zero-inflated	0.854, 0.848 (0.777,0.908)	0.119, 0.127 (0.083,0.173)	0.00401, 0.00399 (0,0.008)	0.031, 0.0314 (0.003,0.059)	0.006, 0.00571 (0,0.012)
Poisson	5.71, 6.45 (3.86,10.2)	17.2, 480 (20.4,549)	2.29, 4.82 (0.002,12.3)	0.0232, 0.0516 (0,0.126)	0.0878, 0.109 (0,0.242)

The results are shown for the Zero-inflated and Poisson components of the model. All statistics (Pop. Mean, population mean; V_P, phenotypic variance; V_A, additive genetic variance; h^2, heritability; I_A, evolvability) presented in the table are reported on the observed data-scale. Estimates are reported as follows: posterior mode, posterior mean, and (95% Credible Interval).

568 **Table 2.** Posterior modes, means and 95% Credible Intervals (in brackets) for observed data-scale variance estimates from the
569 quantitative genetic analyses of annual reproductive success (ARS) and annual adult survival (AAS).

Fitness component	Sample size	Number of individuals	Pop. Mean	V_P	V_A	h^2	I_A
ARS	4453	793	0.84, 0.851 (0.581,1.16)	1.03, 1.5 (0.717,2.9)	0.00728, 0.0104 (0,0.024)	0.00598, 0.00717 (0,0.016)	0.013, 0.014 (0,0.03)
ASS	5290	793	0.799, 0.779 (0.672,0.873)	0.16, 0.169 (0.112,0.221)	7.69×10-06, 0.00211 (0,0.007)	4.65×10-05, 0.0124 (0,0.037)	4.4×10-05, 0.00371 (0,0.0121)

570 All statistics (Pop. Mean, population mean; V_P, phenotypic variance; V_A, additive genetic variance; h^2, heritability; I_A, evolvability)
571 presented in the table are reported on the observed data scale. Estimates are reported as follows: posterior mode, posterior mean, and
572 (95% Credible Interval).