Supporting Information for:

Sex-specific additive genetic variances and correlations for fitness in a song sparrow (*Melospiza melodia*) population subject to natural immigration and inbreeding

Matthew E. Wolak[1,2], Peter Arcese[3], Lukas F. Keller[4,5], Pirmin Nietlisbach[4,6], and Jane M. Reid[1]

[1]School of Biological Sciences, University of Aberdeen, Aberdeen, Scotland

[2]Department of Biological Sciences, Auburn University, Auburn, AL, USA

[3]Department of Forest and Conservation Sciences, University of British Columbia, Vancouver, British Columbia, Canada

[4]Department of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

[5]Zoological Museum, University of Zurich, Karl-Schmid-Strasse 4, CH-8006 Zurich, Switzerland

[6]Department of Zoology, University of British Columbia, Vancouver, British Columbia, Canada

* Corresponding authors: matthew.wolak@auburn.edu and jane.reid@abdn.ac.uk

CONTENTS

APPENDIX S1: LITERATURE SUMMARY

Studies that estimated additive genetic variance (or heritability) in some measure of fitness in a free-living population (including humans) were identified through a literature review conducted in December 2017 (Table S1). Full quantitative details of all estimates are provided in Hendry et al. (2018).

Table S1. (A) Summary of studies that estimated additive genetic variance (V_A, or heritability) in sex-specific fitness measured approximately from zygote to zygote, and **(B)** examples of further studies that estimated V_A in fitness measured from adult to adult in females and males. 'X' identifies studies that estimated V_A in female or male fitness. Superscript 'S' denotes a study reporting estimates of V_A that differed from zero. Superscripts '0' and '(0)' denote studies with estimates that were either zero or were not different from zero, respectively (where a difference from zero was assessed using either a likelihood ratio test or by inspecting whether a lower 95% credible interval or confidence interval (CI) limit converged towards zero). Absence of a superscript indicates studies that did not report results of a likelihood ratio test or a 95%CI. Superscripts summarize multiple estimates when a study reported more than one estimate of V_A in a given sex, consequently multiple superscripts are possible. Studies that estimated the cross-sex genetic correlation (r_A) are highlighted (grey shading). Estimated values of r_A (with standard errors or 95%CIs where available) are noted. In (A), studies are ordered by publication date within plants (1-4), non-human animals (5-14) and humans (15-17).

		Species & system	V_A in female fitness	V_A in male fitness	r_A	Paper
A	1	*Ipomoea purpurea* Durham, USA	X			Simms 1989
	2	*Ipomopsis aggregata*	$X^{(0)}$			Campbell 1997

	Gunnison, USA				
3	*Erigeron annus*	X^{0}			Stratton 1992
	Weld Preserve, USA				
4	*Chamaecrista fasciculata*	$X^{S(0)}$			Etterson 2004
	3 populations, USA				
5	*Cervus elaphus*	X^{0}	$X^{(0)}$		Kruuk et al.
	Rum, UK				2000
6	*Ficedula albicollis*	$X^{(0)}$	$X^{(0)}$		Merilä &
	Gotland, Sweden				Sheldon 2000
7	*Ovis canadensis*	$X^{S(0)}$			Réale & Festa-
	2 populations, Canada				Bianchet 2000
8	*Parus major*	$X^{(0)}$	$X^{(0)}$		McCleery et al.
	Wytham Wood, UK				2004
9	*Macaca mulatta*	X^{S}			Blomquist 2009
	Cayo Santiago, Puerto Rico				
10	*Macaca mulatta*	X^{S}			Blomquist 2010
	Cayo Santiago, Puerto Rico				
11	*Cervus elaphus*	X^{0}			Stopher et al.
	Rum, UK				2012
12	*Tamiasciurus hudsonicus*	$X^{(0)}$	$X^{(0)}$	-0.95	McFarlane et al.
	Yukon, Canada			(-0.99,0.78)	2014[1]
13	*Passerculus sandwichensis*	X	X		Wheelwright et
	Kent Island, Canada				al. 2014
14	*Tamiasciurus hudsonicus*	$X^{(0)}$			McFarlane et al.
	Yukon, Canada				2015
15	*Homo sapiens*	$X^{S(0)}$	$X^{(0)}$		Imaizumi et al.
	Uto, Japan				1970
16	*Homo sapiens*	X^{S}	X^{S}	0.34	Zietsch et al.
	Swedish Twin Registry				2014
17	*Homo sapiens*	$X^{(0)}$	X^{S}		Gavrus-Ion et

		Hallstatt, Austria				al. 2017
B	1	*Ovis canadensis*	X^0	X^0		Coltman et al.
		Ram Mountain, Canada				2005
	2	*Homo sapiens*	X^S	X^0		Pettay et al.
		Finland				2005
	3	*Ficedula albicollis*	X	X	-0.85	Brommer et al.
		Gotland, Sweden			(0.59)	2007
	4	*Cervus elaphus*	$X^{(0)}$	$X^{(0)}$	-0.48	Foerster et al.
		Rum, UK			(0.44)	2007
	5	*Laurus novaehollandiae*	$X^{(0)}$	X^0		Teplitsky et al.
		Kaikoura, New Zealand				2009

[1]McFarlane et al. (2014) also provided a second estimate of r_A of 0.96 (95%CI -0.84, 0.99), but reported that this model did not converge.

Literature cited in Table S1

Blomquist, G. E. 2009. Fitness-related patterns of genetic variation in rhesus macaques. Genetica 135:209-219.

Blomquist, G. E. 2010. Heritability of fitness in female macaques. Evol. Ecol. 24:657-669.

Brommer, J. E., M. Kirkpatrick, A. Qvarnström, and L. Gustafsson, L. 2007. The intersexual genetic correlation for lifetime fitness in the wild and its implications for sexual selection. PLoS One 2:e744.

Campbell, D. R. 1997. Genetic and environmental variation in life-history traits of a monocarpic perennial: a decade-long field experiment. Evolution 51:373-382.

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

Etterson, J. R. 2004. Evolutionary potential of *Chamaecrista fasciculata* in relation to climate change. II. Genetic architecture of three populations reciprocally planted along an environmental gradient in the Great Plains. Evolution 58:1459-1471.

Foerster, K., T. Coulson, B. C. Sheldon, J. M. Pemberton, T. H. Clutton-Brock, and L. E. B. Kruuk. 2007. Sexually antagonistic genetic variation for fitness in red deer. Nature 447:1107-1111.

Gavrus-Ion, A., T. Sjøvold, M. Hernández, R. González-José, and M. E. E. Torné. 2017. Measuring fitness heritability: Life history traits versus morphological traits in humans. Am. J. Phys. Anthropol. 2017:1-10.

Imaizumi Y, M. Nei, and T. Furusho. 1970. Variability and heritability of human fertility. Ann. Hum. Genet. 33:251–59.

Kruuk, L. E. B, T. H. Clutton-Brock, J. Slate, J. M. Pemberton, S. Brotherstone, and F. E. Guinness. 2000. Heritability of fitness in a wild mammal population. Proc. Natl. Acad. Sci. 97:698-703.

McCleery, R. H., R. A. Pettifor, P. Armbruster, K. Meyer, B. C. Sheldon, and C. M. Perrins. 2004. Components of variance underlying fitness in a natural population of the great tit *Parus major*. Am. Nat. 164:E62-E72.

McFarlane, S. E., J. C. Gorrell, D. W. Coltman, M. M. Humphries, S. Boutin, and A. G. McAdam. 2014. Very low levels of direct additive genetic variance in fitness and fitness components in a red squirrel population. Ecol. & Evol. 4:1729-1738.

McFarlane, S. E., J. C. Gorrell, D. W. Coltman, M. M. Humphries, S. Boutin, and A. G. McAdam. 2015. The nature of nurture in a wild mammal's fitness. Proc. R. Soc. B 282:20142422.

Merilä, J., and B. C. Sheldon. 2000. Lifetime reproductive success and heritability in nature. Am. Nat. 155:301-310.

Pettay, J., L. E. B. Kruuk, J. Jokela, and V. Lummaa. 2005. Heritability and genetic constraints of life-history trait evolution in preindustrial humans. Proc. Natl. Acad. Sci. 102:2838-2843.

Réale, D., and M. Festa-Bianchet. 2000. Quantitative genetics of life-history traits in a long-lived wild mammal. Heredity 85:593-603.

Simms, E. L., and M. D. Rausher. 1989. The evolution of resistance to herbivory in *Ipomoea purpurea*. II. Natural selection by insects and costs of resistance. Evolution 43:573–85.

Stopher, K. V., C. A. Walling, A. Morris, F. E. Guinness, T. H. Clutton-Brock, J. M. Pemberton, and D. H. Nussey. 2012. Shared spatial effects on quantitative genetic parameters: accounting for spatial autocorrelation and home range overlap reduces estimates of heritability in wild red deer. Evolution 66:2411-2426.

Stratton, D. A. 1992. Life-cycle components of selection in *Erigeron annuus*: II. Genetic variation. Evolution 46:107–20.

Teplitsky, C., J. A. Mills, J. W. Yarrall, and J. Merilä. 2009. Heritability of fitness components in a wild bird population. Evolution 63:716-726.

Wheelwright, N. T., L. F. Keller, and E. Postma. 2014. The effect of trait type and strength of selection on heritability and evolvability in an island bird population. Evolution 68:3325-3336.

Zietsch, B. P., R. Kuja-Halkola, H. Walum, and K. J. H. Verweij. 2014. Perfect genetic correlation between number of offspring and grandoffspring in an industrialized human population. Proc. Natl. Acad. Sci. 111:1032-1036.

APPENDIX S2: OVERALL APPROACH AND DATA SPECIFICATIONS

S2.1 Metrics of fitness

Numerous metrics of fitness can be defined, that consider relative versus absolute fitness; consider time-dependent versus time-independent reproduction and/or contributions on an annual rather than lifetime basis; consider reproductive value of offspring reflecting population age-structure; and that assign fitness to different individuals across generations (e.g. Brommer 2000; Coulson et al. 2006; Metcalf and Pavard 2007; Orr 2009; Sæther and Engen 2015). We restricted our current analyses to fitness measured as the number of chicks produced by each chick as an approximation to absolute fitness measured from zygote-to-zygote, thereby retaining links with fundamental evolutionary theory including in the emerging context of 'evolutionary rescue' (e.g. Gomulkiewicz and Shaw 2013; Hendry et al. 2018). Further, because most individuals in our dataset had zero fitness, our time-independent metric is tightly correlated with time-dependent metrics such as individual lambda across the population (e.g. Brommer et al. 2002; Lebigre et al. 2012). We could not measure fitness as the number of eggs produced by each egg because paternity is unknown for eggs that did not hatch (due to individual failure or, more commonly, due to whole clutch predation). Further, because song sparrows can lay multiple replacement clutches, total observed fecundity does not relate simply to fitness because females that lay most eggs are those whose breeding attempts repeatedly fail.

S2.2 Overall approach to quantifying V_A in fitness

Our overall approach was to build a hierarchy of models that considered sex-specific fitness (defined as above), then decomposed fitness into biologically relevant subcomponents; namely juvenile survival and adult lifetime reproductive success (LRS), followed by adult annual reproductive success (ARS) and annual survival.

Alternative QGGLMM formulations exist that analytically incorporate characteristics of typical fitness distributions, such as zero-inflated or hurdle Poisson models. However, such models would yield different biological interpretations. For example, a hurdle Poisson model would partition variance into separate processes that create zero fitness and the remaining zero-truncated Poisson distribution, and thereby quantify V_A in binary fitness and in the magnitude of reproductive success conditional on non-zero success. Instead, we explicitly defined and analyzed the known biologically meaningful traits of juvenile survival and adult LRS which, together, generate the observed distribution of fitness. Aster models can directly model the complex distribution of fitness resulting from sequential expression of fitness components (Geyer et al. 2007; Shaw et al. 2008; Shaw and Etterson 2012; Shaw and Shaw 2014), but cannot yet incorporate relatedness measured from non-standard (e.g. wild population) pedigree structures.

In general, major fitness components are expected to contribute positively to overall fitness (Lande and Arnold 1983; Falconer 1989, p. 331-334; Walling et al. 2014). In principle, estimates of V_A and COV_A in and among juvenile survival and adult LRS mathematically combine to yield a complete accounting of V_A in fitness. However, when V_As and COV_As are estimated from models using different link functions, latent scale estimates cannot be simply combined.

In our current analyses, V_A for fitness (as opposed to fitness components) could not be reliably transformed back to the observed data scale integrating over fixed effects (e.g. de Villemereuil et al. 2016, see Appendix S4). This was because posterior simulation from the fitted QGGLMM yielded poor predicted phenotypes as a consequence of fitting a Poisson distribution to overdispersed sex-specific fitness distributions containing large proportions of zeroes and extreme right tails (Fig. 1A). Consequently, we could not quantitatively compare estimates of V_A in fitness and fitness components, as estimated by different QGGLMMs. Similarly, we were unable to fit an adequate pentavariate QGGLMM comprising sex-specific fitness and LRS and juvenile survival, which would directly estimate the total V_A in fitness attributable to fitness components (e.g. Walsh

and Blows 2009). However, qualitative interpretation of the patterns of V_A, COV_A and r_A among the fitness components allows heuristic accounting for V_A in overall fitness.

S2.3 Data restrictions

Fitness: Of the total of 2838 song sparrow chicks banded on Mandarte between 1993 and 2012, 16 (0.56%) were excluded from analyses of fitness because their sex was unknown. In addition, five individuals hatched during 2010-2012 were still alive in April 2016, meaning that their lifetime fitness was incomplete. However, since the fitness values for all five individuals up to April 2016 rank within the top four in their respective cohorts, and song sparrows aged five years or greater typically accrue little further fitness, records for these five individuals were retained in analyses. Since these five individuals comprise 0.18% of the dataset their retention is unlikely to cause bias, and facilitated appropriate estimation of cohort variance. The same five individuals were also retained in analyses of adult LRS.

Juvenile survival: To maximize use of all available data, analyses of juvenile survival utilized phenotypic data for two additional cohorts of individuals (hatched in 2013 and 2014) compared to the analyses of fitness. Of the 3125 total chicks banded on Mandarte between 1993 and 2014, 21 (0.67%) were excluded from analyses of juvenile survival because their sex was unknown (n=18), or because the date on which the first egg was laid in their natal nest was unknown (n=3).

APPENDIX S3: PEDIGREE STRUCTURE AND GENETIC GROUPS

S3.1 Pedigree structure

The song sparrow study system and dataset have multiple properties that facilitate QGGLMM estimation of V_A in key traits beyond simply the availability of complete life-history, fitness and pedigree data. First, the frequent extra-pair paternity, and social repairing following mate death or divorce, generates numerous maternal and paternal half-sibs reared in the same and different environments (Germain et al. 2018). Second, the multi-annual life-history, with considerable variation in lifespan (Fig. 5) means that close relatives can be observed in multiple different years. Third, dispersal is panmictic within Mandarte (total study area approximately 6 hectares) such that there is no detectable spatial structure in relatedness (Germain et al. 2016). Together, these properties facilitate separation of additive genetic and spatio-temporal environmental effects on phenotypes, and hence associated variance components.

Overall, the mean coefficient of kinship (k) among all individuals whose phenotypic fitness was observed and included in the bivariate QGGLMM was 0.072 ± 0.034SD (median 0.066, interquartile range 0.052-0.084, range 0.003-0.466). These descriptive statistics were virtually identical across individuals whose phenotypes were included in each of the other three QGGLMMs.

S3.2 Genetic groups

Standard QGGLMMs estimate V_A and COV_A in a default baseline population that comprises 'phantom parents' of all pedigreed individuals with unknown parents (Kruuk 2004; Wolak and Reid 2017). Due to the comprehensive long-term fieldwork, there are few non-founder (i.e. post 1975) individuals in the Mandarte song sparrow pedigree with unknown dam and sire identities. The primary exceptions concern some chicks hatched in 1980 (when reduced fieldwork meant that

surviving adults were identified but chicks were not fully assigned to parents, Keller 1998; Smith et al. 2006), and subsequent immigrants (Marr et al. 2002; Reid et al. 2006).

Each unknown (i.e. ‘phantom’) parent identity in the pedigree was assigned to one of two genetic groups defined as: 'founder' (all adults present in 1975 and hence hatched before the study began plus unknown parents of chicks hatched in 1980), and 'immigrant' (all parents of individuals identified as immigrants to Mandarte; see main Methods; Marr et al. 2002). We pooled unknown parents from 1980 into the founder genetic group because, since mean song sparrow generation time exceeds two years and several adults alive in 1980 and 1981 were the offspring of 1975 founders, there is unlikely to have been a substantial intervening change in mean breeding value for fitness or fitness components. Exploratory analyses supported this expectation. Specifically, additional analyses that split unknown parents from 1980 into their own genetic group estimated no difference between mean additive genetic value for this group and the 1975 founders.

Meanwhile, our defined ‘founder’ genetic group likely contains some pre-1975 immigrants and their descendants. Indeed, within a meta-population where demes are connected by some dispersal and resulting gene flow, single demes are by definition not genetically ‘pure’. Our defined founder population simply constitutes a baseline in which to estimate V_A and COV_A, concurring with the default assumption of all pedigree-based QGGLMMs and hence all previous wild population analyses (Kruuk 2004; Wolak and Reid 2017), and against which to compare the mean additive genetic values of subsequent immigrants.

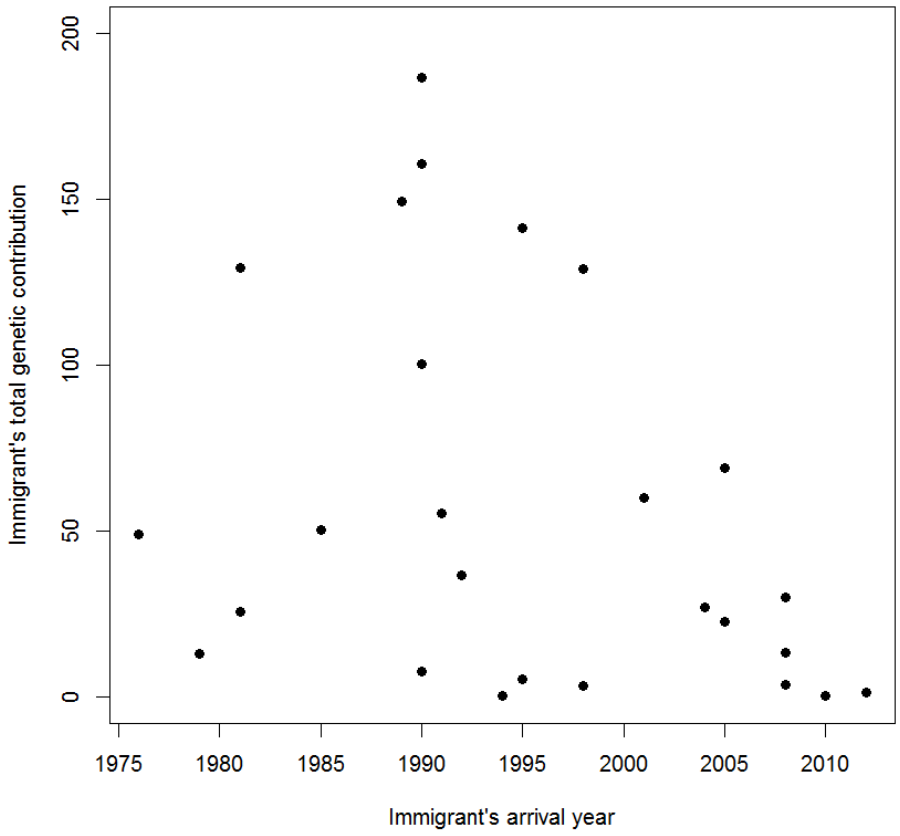
To calculate the contribution of each genetic group to each individual song sparrow in the dataset, we compiled a pedigree with all phantom parents assigned to the founder or immigrant genetic groups, and applied the ggcontrib() function in the nadiv R package (v2.14.3.2, Wolak 2012; Wolak and Reid 2017). Since non-zero values of immigrant genetic group (*IGG*) coefficients were ubiquitous among focal phenotyped individuals (see Results), and immigrants’ own phenotypes were not included in analyses, these analyses allow additive effects of immigrants’

genes to be separated from environmental effects on immigrants' own phenotypes. Such models assume that V_A is homogeneous across the defined genetic groups (Wolak and Reid 2017).

Immigrants that did not produce any banded offspring on Mandarte (i.e. had zero fitness) do not contribute to the pedigree and hence are completely excluded from our analyses. Estimated genetic group effects (i.e. β_{IGG}) should consequently be interpreted as the relative mean genetic values of immigrants versus defined founders given introgression, not as the mean absolute genetic values for fitness of all immigrants that arrived on Mandarte.

Overall, 26 immigrants made a non-zero total genetic contribution to the set of individuals included in the quantitative genetic analysis of fitness (i.e. those hatched during 1993-2012), where total genetic contribution is defined as the sum of twice the coefficient of kinship (k) between each individual immigrant and all focal phenotyped individuals. These immigrants arrived in a range of years (Fig. S1). However, the largest individual contributions came from some immigrants that arrived between ca. 1989 and 2000 (Fig. S1). This might be expected, since these individuals have the opportunity to be ancestors of many focal phenotyped individuals. In contrast, immigrants that arrived more recently (e.g. during 2001-2011) cannot be ancestors of focal phenotyped individuals that hatched prior to their arrival. Overall, immigrants that arrived since 1989 accounted for 82% of the total immigrant genetic contribution, implying that estimated genetic group effects are primarily attributable to these individuals. With more years of data, future analyses could potentially estimate additive genetic effects of different groups of immigrants that arrived in different sets of years (Wolak and Reid 2017).

Figure S1. Relationship between an individual immigrant's year of arrival on Mandarte and its total genetic contribution to the set of individuals whose phenotypic fitness was measured. Only the 26 immigrants whose total genetic contribution exceeded zero, and hence that contributed to estimation of additive genetic effects, are shown.



APPENDIX S4: DETAILS OF MODEL SPECIFICATIONS AND IMPLEMENTATION

All R code, including model and prior statements using the R package MCMCglmm (Hadfield 2010), is available on GitHub: https://github.com/matthewwolak/Wolak_etal_SongSparrowFitnessQG and the Dryad Digital Repository: https://doi.org/10.5061/dryad.p7p1jb3 (Wolak et al. 2018). However, to facilitate interpretation of the main text, we provide simplified model equations below. This complements information provided in the main Methods text, while leaving further detailed explanation to other sources devoted to the QGGLMM framework and the relationship between observed non-Gaussian phenotypes and latent variables (e.g. de Villemereuil et al. 2016, their equations 3a,b,c).

To most easily describe the models and consequent results, we focus on modeling latent variables ($\boldsymbol{l}$). Throughout the model statements below, we use $\mathbf{X}$ and $\mathbf{Z}$ to indicate fixed and random design matrices, respectively, that associate the correct fixed ($\mathbf{b}$) and random effects to the latent variables in ($\boldsymbol{l}$). Latent variables ($\boldsymbol{l}$) are normally distributed values that, through the non-linear model structure, correspond to the phenotypes on the observed scale (z). Each observed phenotype (z) has an expected value in the statistical model, $\boldsymbol{\eta}$. The difference between the expected value $\boldsymbol{\eta}$ and the observed value $\boldsymbol{z}$ is due to a statistical error process, which assigns values according to the QGGLMM's assumed error distribution (e.g., in our case, binomial or Poisson errors) and any parameters that describe that distribution (e.g., a scale parameter for a Poisson distribution). The expected values $\boldsymbol{\eta}$ are transformed into latent scale variables $\boldsymbol{l}$ according to a specified link function (e.g., in our case, logarithmic or logit link functions) and system to explain additional variation on the latent scale, in our case additive overdispersion. We use the variable $\mathbf{o}$ to indicate the overdispersion vector (which is fixed to a constant variance of one for binary models with logit link, such as juvenile survival and adult survival).

S4.1 Model specifications

We include here the remaining information about model details that are in addition to the presentation in the main text for the choice of error distributions, link functions, and handling of overdispersion unique to each QGGLMM.

Fitness: Our bivariate QGGLMM for fitness was of the general form:

$$\boldsymbol{l} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Zc} + \mathbf{o},$$

where $\boldsymbol{l}$ contains female and male latent variables $[\boldsymbol{l}_{female}, \boldsymbol{l}_{male}]^T$ and ^{T} denotes the transpose. The vectors **a**, **c**, and **o** contain female and male additive genetic $[\mathbf{a}_{female}, \mathbf{a}_{male}]^T$ and cohort $[\mathbf{c}_{female}, \mathbf{c}_{male}]^T$ random effects and sex-specific overdispersion $[\mathbf{o}_{female}, \mathbf{o}_{male}]^T$.

We used a multivariate extension of the parameter expanded prior distribution for the covariance matrices of the additive genetic and hatch-year cohort variance components, and an inverse Wishart prior distribution for the sex-specific residual variances (see also Appendix S6). The model was run for 12,505,000 iterations (burn-in 5,000; thinning interval 2,500).

Juvenile survival and adult lifetime reproductive success: To estimate and account for common environmental effects stemming from parental care and natal conditions, effects of the identities of each chick's mother, social father, social parent pair, and brood were additionally fitted for juvenile survival. While these four effects may be somewhat confounded, our aim was not to precisely estimate associated variances, but simply to minimize possible bias in V_A (e.g. Kruuk and Hadfield 2007; Reid et al. 2014).

Analogous common environmental effects were not fitted to female and male adult LRS, because most observed adults had no same-sex siblings that survived to adulthood (69.7% of females and 65.5% of males respectively), or no same-sex broodmates (88.0% of females and 84.1% of males respectively). The median, mean, and maximum adults produced per unique parent pair or brood are:

Adult females/pair: 1, 1.54, 6

Adult males/pair: 1, 1.54, 6

Adult females/brood: 1, 1.13, 3

Adult males/brood: 1, 1.17, 3

Further, exploratory analyses and previous results did not reveal substantial parental effects on adult life-history traits in our system (e.g. Reid et al. 2011, 2014).

Consequently, our trivariate QGGLMM for juvenile survival and female and male adult lifetime reproductive success (LRS) was of the general form:

$$\boldsymbol{l} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Zy} + \mathbf{Zdam} + \mathbf{Zsoc.sire} + \mathbf{Zbrood} + \mathbf{Zpair} + \mathbf{o},$$

where $\boldsymbol{l}$ contains juvenile survival, female LRS, and male LRS latent variables $[\boldsymbol{l}_{juvenile.survival}, \boldsymbol{l}_{female.LRS}, \boldsymbol{l}_{male.LRS}]^{T}$ and ^{T} denotes the transpose. The vector **a** contains additive genetic effects for juvenile survival, female LRS, and male LRS $[\boldsymbol{a}_{juvenile.survival}, \mathbf{a}_{female.LRS}, \mathbf{a}_{male.LRS}]^{T}$. The vector **y** contains natal-year cohort effects on juvenile survival, female LRS, and male LRS $[\boldsymbol{y}_{juvenile.survival}, \boldsymbol{y}_{female.LRS}, \boldsymbol{y}_{male.LRS}]^{T}$. The effects **dam**, **soc.sire**, **brood**, and **pair** contain social dam, social sire, brood, and social parent pair for juvenile survival only (through the design matrix **Z**). The vector **o** contains fixed variance of one for all juvenile survival records and estimates overdispersion for both female LRS and male LRS $[\boldsymbol{1}_{juvenile.survival}, \mathbf{o}_{female}, \mathbf{o}_{male}]^{T}$.

A trivariate extension of the parameter expanded prior was used for the additive genetic and hatch-year cohort covariance components, where the scale parameter was specified as a diagonal matrix with the elements of the leading diagonal equaling 1000, 1000, and 10 for female LRS, male LRS, and juvenile survival, respectively. An inverse Wishart prior distribution was used for the residual variances in female and male LRS. For the fixed regression on first egg lay date, we standardized each lay date by subtracting mean lay date and dividing by the standard deviation of all observations. The model was run for 40,025,000 iterations (burn-in 25,000; thinning interval 8,000).

Adult annual reproductive success: Our bivariate QGGLMM of annual reproductive success (ARS) was of the general form:

$$\boldsymbol{l} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Zp} + \mathbf{Zc} + \mathbf{o},$$

where $\boldsymbol{l}$ contains female and male latent variables $[\boldsymbol{l}_{female}, \boldsymbol{l}_{male}]^{T}$ and ^{T} denotes the transpose. The vectors **a**, **p**, **c**, and **o** contain female and male additive genetic $[\mathbf{a}_{female}, \mathbf{a}_{male}]^{T}$, permanent individual $[\mathbf{p}_{female}, \mathbf{p}_{male}]^{T}$, and cohort $[\mathbf{c}_{female}, \mathbf{c}_{male}]^{T}$ random effects and sex-specific overdispersion $[\mathbf{o}_{female}, \mathbf{o}_{male}]^{T}$.

We used a multivariate extension of the parameter expanded prior distribution for the covariance matrices of the additive genetic and year variance components. A slightly different extension of the parameter expanded prior distribution was used for the diagonal matrix for female and male permanent individual variance components, with the covariance fixed to zero. An inverse Wishart prior distribution was used for the sex-specific residual variances. The model was run for 25,015,000 iterations (burn-in 15,000; thinning interval 5,000).

Adult annual survival: Our univariate QGGLMM of adult annual survival was setup as an event history survival model of the general form:

$$\boldsymbol{l} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Zp} + \mathbf{Zy} + \mathbf{o},$$

where $\boldsymbol{l}$ contains latent variables for both sexes combined. The vectors **a**, **p**, **y**, and **o** contain additive genetic, permanent individual, survival year random effects, and a overdispersion with a fixed variance of one. For the event history survival model, the permanent individual variance estimates overdispersion relative to the assumed geometric distribution of age-specific survival events, where each individual has the same overdispersion effect (o) among all survival records. This is in contrast to the overdispersion represented by the 'units' (or residual) term of the MCMCglmm software, which represents a unique and independent deviation from the expectation (o) for each record of each individual.

The scaled non-central F-distribution prior, obtained via parameter expansion, and a scale parameter of 10 was used for all variance components (with residual variance fixed to one). Models were run for 10,05,000 iterations (burn-in 5,000; thinning interval 2,000). For both fixed age effects and random individual effects (modeling additive overdispersion) to be identifiable in event history survival models, the number of age effects can be no more than the total number of ages modeled minus two. This condition is fulfilled by our specification of four age categories (ages 1, 2, 3-5 or ≥6 years).

S4.2 Parallel MCMC chains and convergence criteria

Individual QGGLMMs were split into 20 chains and run simultaneously either on the Maxwell High Performance Computing cluster, University of Aberdeen, UK or the Hopper cluster, Auburn University, USA. Starting values of latent variables and variance components for each sub-chain were drawn from overdispersed distributions compared to the posteriors. The burn-in and thinning interval for each model were set to ensure MCMC chain convergence. Final MCMC samples were obtained by combining the retained samples from each chain only when the following conditions were met by each chain: Gelman and Rubin's convergence diagnostic (potential scale reduction value) was <1.1 (Gelman and Rubin 1992); no detected under/overflow of the link function (i.e. the maximum absolute value for the latent variable in each chain was less than approximately 35); the average absolute autocorrelation values for the fixed effect solutions and variance components were all <0.1. The final combined samples were accepted only if the absolute autocorrelation values across the entire combined chains were <0.1 and visual inspection of the MCMC posterior traces also indicated convergence.

S4.3 Posterior and prior visualization

Describing uncertainty in quantitative genetic parameter estimates can be difficult when relying on statistics describing sampling distributions (e.g., standard errors) of parameters close to their

boundaries (e.g., variances near zero or correlations close to ±1) and require Gaussian approximations (e.g. Wolak and Reid 2017). To facilitate interpretation of results, for key metrics we depict full marginal posterior distributions alongside prior distributions and distributional descriptive statistics (e.g., posterior mean and 95% highest posterior density credible intervals, Figures 2-5 and S2-S11). To illustrate the distribution of posterior samples, we plotted histograms (with manually-specified bin widths and unit total area) alongside the posterior kernel density estimates (with bandwidths calculated during density estimation). Minor discrepancies between peaks of visualized histograms and posterior modes of the kernel densities simply reflect these different visualization methods, and should not be further interpreted.

To visually inspect the influence of priors on posteriors we plotted an approximate prior density on top of each plot of posterior samples. For independent variance components, prior densities were derived by evaluating the density function for the *F*-distribution at each of the posterior histogram mid-points. These densities were then scaled to give a total area under the curve (over the range depicted by the histogram bins) equal to one.

For covariances and associated variances, approximate prior densities were obtained from 10,000 independent samples from the prior distribution. For heritabilities, approximate prior distributions were obtained by performing the same calculations as for the posterior with each sample of the simulated prior (see below). Kernel density estimation was then applied to the samples from the prior. To 'smooth' the kernel density estimation of the relatively flat prior densities for correlations, we simply calculated the mean density across the majority of the range between -1 and 1 and plotted this density (0.5) across the entire range (giving an area of one under the curve).

Plot areas where posterior samples have a different shape from the prior distribution (e.g. more/less density in a tail, or less density under the mode) indicate that the posterior distribution is substantively influenced by the data rather than solely the prior. Posterior and prior densities

represent relative probability densities and so can be interpreted such that parameter values along the x-axis that have greater density have a greater probability under the model, given the data.

S4.4 Equations for derived quantities

Posterior means, modes, and 95% credible intervals for additive genetic correlations (r_A), heritabilities (h^2), evolvabilities (I_A), and coefficients of additive genetic variance (CV_A) were extracted from the posterior distributions of these metrics. These posterior distributions were computed by applying the following formulae to all retained samples of the posterior distributions of underlying variance components:

Additive genetic correlation:

$$r_A = COV_{A1,2} / (V_{A1} \times V_{A2})^{0.5}$$

where V_{A1} and V_{A2} are the latent-scale additive genetic variances in two focal traits and $COV_{A1,2}$ is the latent-scale additive genetic covariance. The same formula structure was used to calculate correlations between other random effects (e.g. cohort effects).

Heritability:

Latent-scale heritability (h^2_{latent}), conditional on fitted fixed effects, was calculated as:

$$h^2_{latent} = V_A / V_P$$

where:

$$V_P = (V_A + V_{RE} + V_O)$$

where V_A and V_P are the latent-scale additive genetic and phenotypic variances, respectively, V_{RE} is the sum of all other estimated latent-scale variances associated with random effects in a given model, and V_O is the latent-scale overdispersion variance. V_O was estimated in models that assumed

overdispersed Poisson distributions with log link function (i.e. for fitness and adult LRS and ARS), and fixed to one (by convention) in models for binary traits with logit link functions (i.e. for juvenile and adult annual survival).

Observed-scale heritability ($h^2_{observed}$) was calculated as:

$$h^2_{observed} = V_{A\text{-}observed} / V_{P\text{-}observed}$$

where $V_{A\text{-}observed}$ is the additive genetic variance in the focal trait on the observed phenotypic scale, obtained by back-transforming V_A estimated on the QGGLMM latent scale. $V_{P\text{-}observed}$ is the total phenotypic variance on the observed scale, obtained by back-transforming V_P estimated on the QGGLMM latent scale.

Evolvability:

Observed-scale evolvability ($I_{A\text{-}observed}$), which is a mean standardized additive genetic variance (Houle 1992), was calculated as:

$$I_{A\text{-}observed} = V_{A\text{-}observed} / \bar{X}^2$$

where $\bar{X}$ is the phenotypic trait mean, estimated by back-transforming the mean QGLMM latent variable onto the observed scale. Note that $I_{A\text{-}observed}$ can be easily transformed into an observed-scale coefficient of additive genetic variance ($CV_{A\text{-}observed}$) by:

$$I_{A\text{-}observed} = (CV_{A\text{-}observed} / 100)^2$$

since, by definition, $CV_{A\text{-}observed} = 100 \times (V_{A\text{-}observed})^{0.5} / \bar{X}$

Back-transformations to obtain means and variances on the observed phenotypic scale, integrating over fixed effects, were implemented in the R package QGglmm (v0.6.0, de Villemereuil et al. 2016). We integrated over fixed effects by calculating trait-specific means of latent-scale predictions: we set the coefficients of *f* and *IGG* for all individuals to zero (reflecting defined founder population values), computed latent-scale predictions while marginalizing random effects, then calculated the latent-scale mean of each trait in each model. These latent-scale means

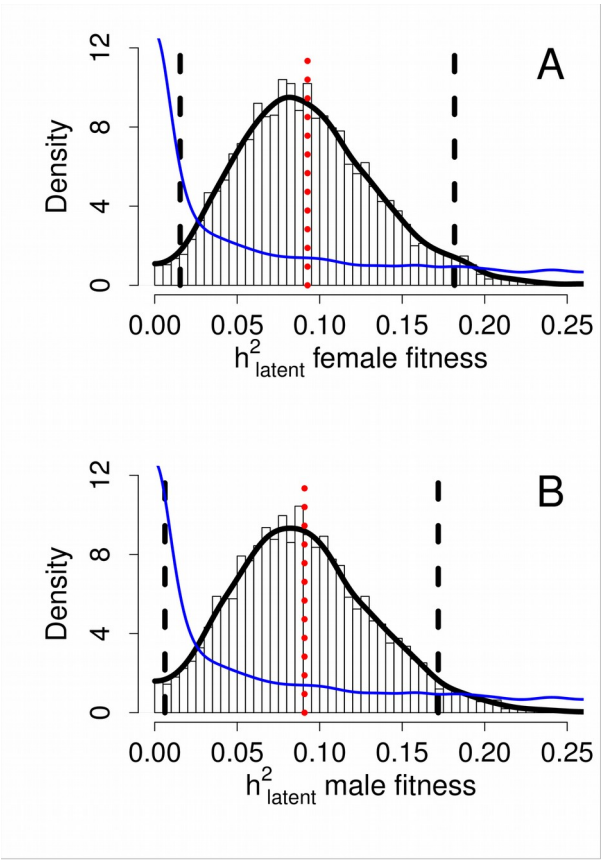
were then supplied to functions in the QGglmm package that were used to back-transform each sample of the posterior distributions of latent-scale variances.

APPENDIX S5: ADDITIONAL DETAILS OF RESULTS

S5.1 Sex-specific fitness

Posterior mean fitness estimated on the latent scale (i.e. the model intercept) was -1.01 (posterior mode: -1.23; 95%CI: -4.21, 2.22) for females and -0.77 (posterior mode: -1.07; 95%CI: -3.88, 2.33) for males. Posterior distributions of the latent-scale heritabilities (h^2_{latent}) are shown in Fig. S2.

Figure S2. Marginal posterior MCMC samples (bars), kernel density estimation (solid black line), posterior mean (red dotted line), 95% credible interval limits (black dashed lines) and prior (solid blue line) for the latent-scale heritability (h^2_{latent}) in female (A) and male (B) fitness in song sparrows. The priors are depicted over the range of the posterior distributions, but extend to values of one.



S5.2 Juvenile survival and adult lifetime reproductive success

Table S2 summarizes additional parameters estimated by the trivariate QGGLMM for juvenile survival and adult female and male lifetime reproductive success beyond those reported in Table 2. Figures S3-S5 depict posterior distributions of key derived parameters.

Table S2. Marginal posterior means, modes (in square brackets), and 95% credible intervals (in parentheses) for additional latent-scale estimates from the trivariate QGGLMM for juvenile survival and adult female and male lifetime reproductive success (LRS). Intercepts are the model intercepts for each trait (i.e. posterior mean fitness estimated on the latent scale). V_{mother}, V_{father}, V_{pair} and V_{brood} are the latent-scale variances in juvenile survival attributable to mothers, social fathers, parent pairs and broods respectively. $\beta_{lay.date}$ is the regression on first egg lay date in each individual's natal nest. The juvenile survival intercepts are estimated for females and males separately, all other parameters are estimated for juveniles of both sexes combined. All other model estimates are in table 2 (main text). Lower 95% CI limits that converged towards zero are reported as <0.001.

	Intercept	V_{mother}	V_{father}	V_{pair}	V_{brood}	$\beta_{lay.date}$
Female juvenile survival	-0.41 [-0.55] (-1.65, 0.80)	0.07 [0.002] (<0.001, 0.22)	0.08 [0.001] (<0.001, 0.23)	0.24 [0.003] (<0.001, 0.60)	0.13 [0.002] (<0.001, 0.37)	-0.22 [-0.20] (-0.34, -0.09)
Male juvenile survival	0.01 [0.14] (-1.22, 1.20)					
Female LRS	1.96 [2.09] (1.16, 2.73)					
Male LRS	1.57 [1.73] (-0.33, 3.68)					

Figure S3. Marginal posterior MCMC samples (bars), kernel density estimation (solid black line), posterior mean (red dotted line), 95% credible interval limits (black dashed lines), and prior (solid blue line) for the latent-scale heritability (h^2_{latent}, A) and observed-scale heritability ($h^2_{observed}$, B) in juvenile survival. The prior for h^2_{latent} is depicted over the range of the posterior distribution, but extends to values of one. Note that x-axis scales differ between panels.

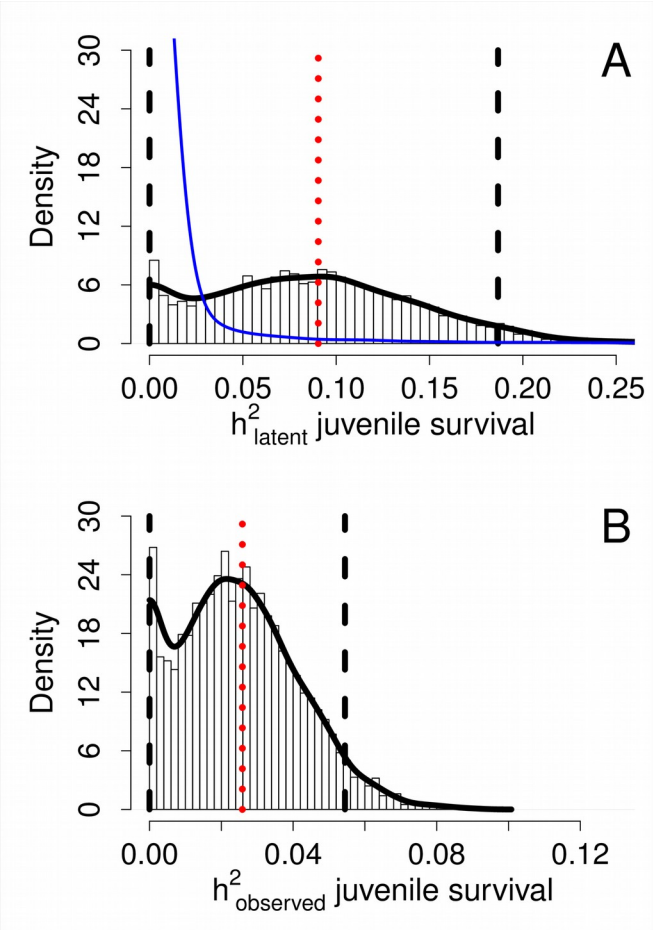


Figure S4. Marginal posterior MCMC samples (bars), kernel density estimation (solid black line), posterior mean (red dotted line), 95% credible interval limits (black dashed lines), and prior (solid blue line) for (A and B) latent-scale heritability (h^2_{latent}) and (C and D) observed-scale heritability ($h^2_{observed}$) in adult lifetime reproductive success (LRS), in (A and C) females and (B and D) males. The priors for h^2_{latent} are depicted over the range of the posterior distribution, but extend to values of one. Note that axis scales differ among panels.

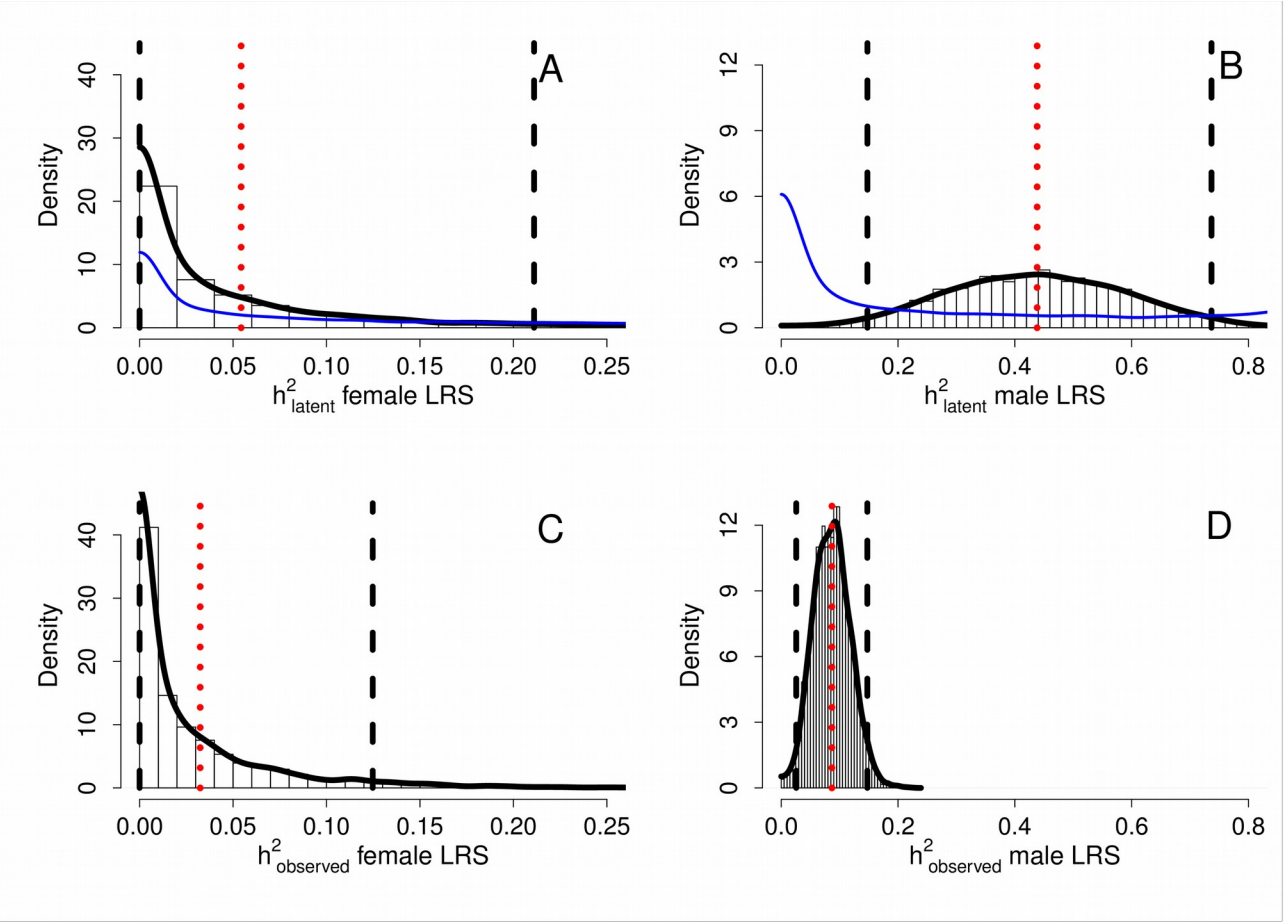
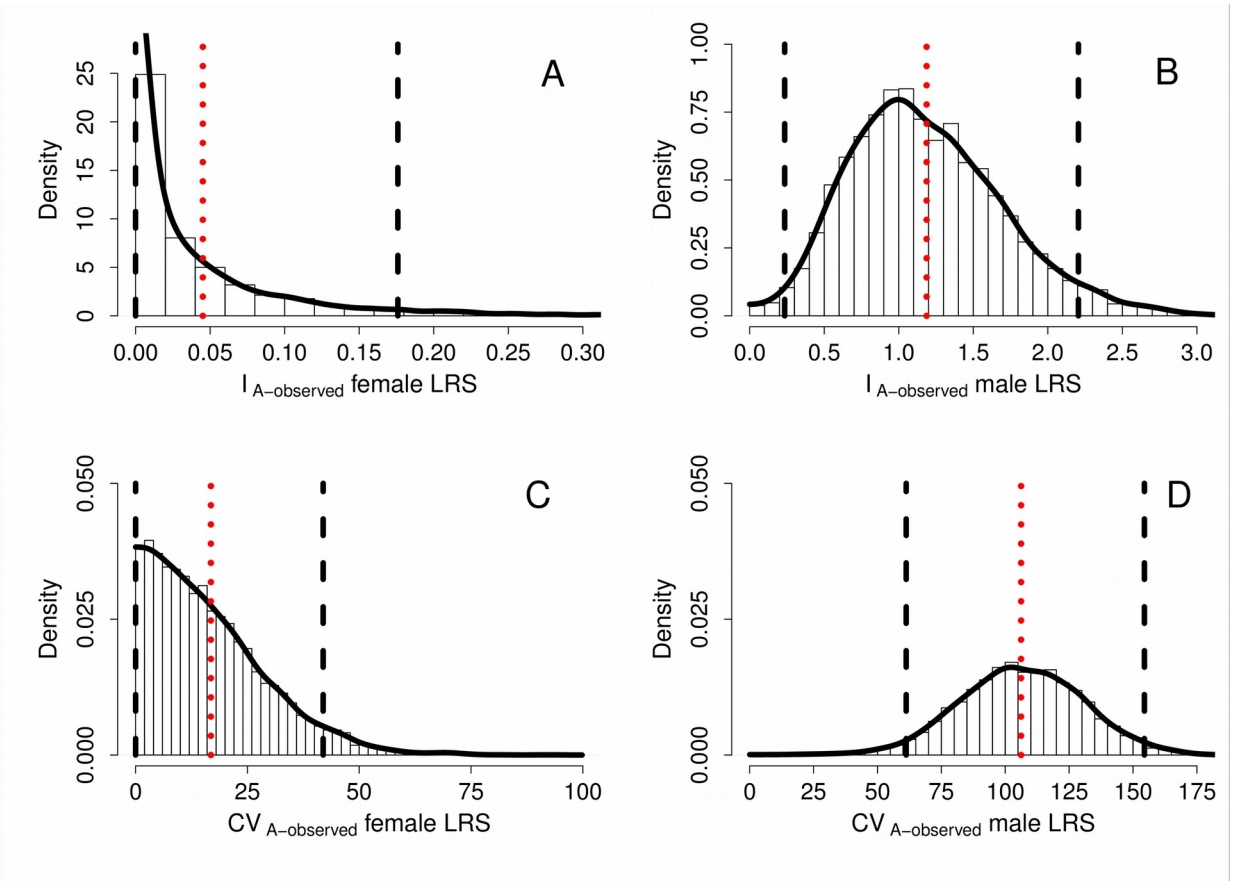


Figure S5. Marginal posterior MCMC samples (bars), kernel density estimation (solid black line), posterior mean (red dotted line), and 95% credible interval limits (black dashed lines) for the observed-scale (A and B) evolvability ($I_{A\text{-observed}}$) and (C and D) coefficient of additive genetic variance ($CV_{A\text{-observed}}$) in adult lifetime reproductive success (LRS), in (A and C) females and (B and D) males. Note that axis scales differ among panels.



S5.3 Annual reproductive success

To account for known age effects on ARS, fixed effects of four age categories were modeled (ages 1, 2, 3-5 and ≥6 years). The numbers of observations of ARS within each age category were:

Age 1: females 254, males 331

Age 2: females 125, males 187

Ages 3-5: females 129, males 218

Age ≥6: females 18, males 37 (maximum observed age was nine years)

ARS at age one year was fitted as a sex-specific intercept, and sex-specific effects of the other three age categories were estimated as contrasts to each sex-specific intercept (Table S3).

Figures S6 and S7 depict posterior distributions of key derived parameters for adult female and male annual reproductive success (ARS).

Table S3. Marginal posterior means, modes (in square brackets), and 95% credible intervals (in parentheses) for age effects estimated from the bivariate QGGLMM for adult female and male annual reproductive success (ARS). Estimates for ages 2, 3-5 and ≥6 are contrasts from the intercept at age 1 year. All other model estimates are in table 2 (main text).

	Age 1	Age 2	Ages 3-5	Age ≥6
Female ARS	1.31	0.32	0.28	-0.07
	[1.34]	[0.32]	[0.27]	[-0.07]
	(0.90, 1.76)	(0.20, 0.44)	(0.16, 0.39)	(-0.39, 0.22)
Male ARS	0.18	1.00	1.13	0.64
	[0.23]	[1.00]	[1.15]	[0.68]
	(-0.80, 1.22)	(0.82, 1.20)	(0.95, 1.34)	(0.27, 1.02)

Figure S6. Marginal posterior MCMC samples (bars), kernel density estimation (solid black line), posterior mean (red dotted line), 95% credible interval limits (black dashed lines), and prior (solid blue line) for (A and B) latent-scale heritability (h^2_{latent}) and (C and D) observed-scale heritability ($h^2_{observed}$) in adult annual reproductive success (ARS), in (A and C) females and (B and D) males. The priors for h^2_{latent} are depicted over the range of the posterior distribution, but extend to values of one.

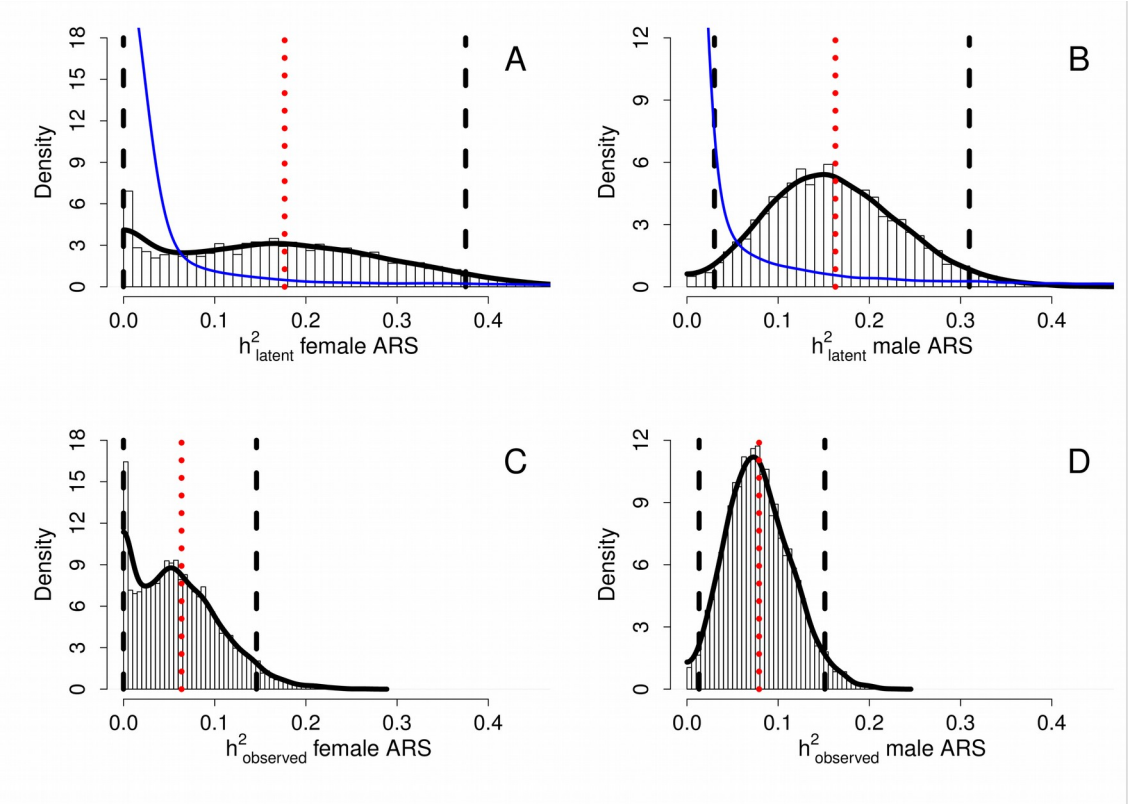
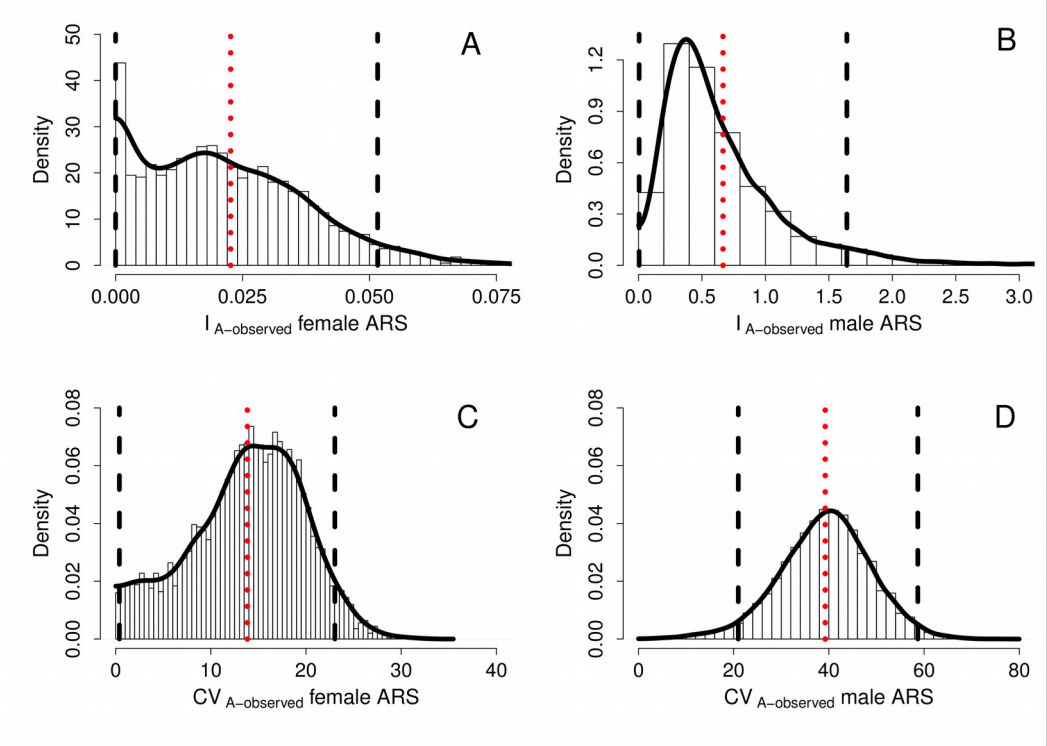


Figure S7. Marginal posterior MCMC samples (bars), kernel density estimation (solid black line), posterior mean (red dotted line), and 95% credible interval limits (black dashed lines) for the observed-scale (A and B) evolvability ($I_{A\text{-observed}}$) and (C and D) coefficient of additive genetic variance ($CV_{A\text{-observed}}$) in adult annual reproductive success (ARS), in (A and C) females and (B and D) males. Note that axis scales differ among panels.



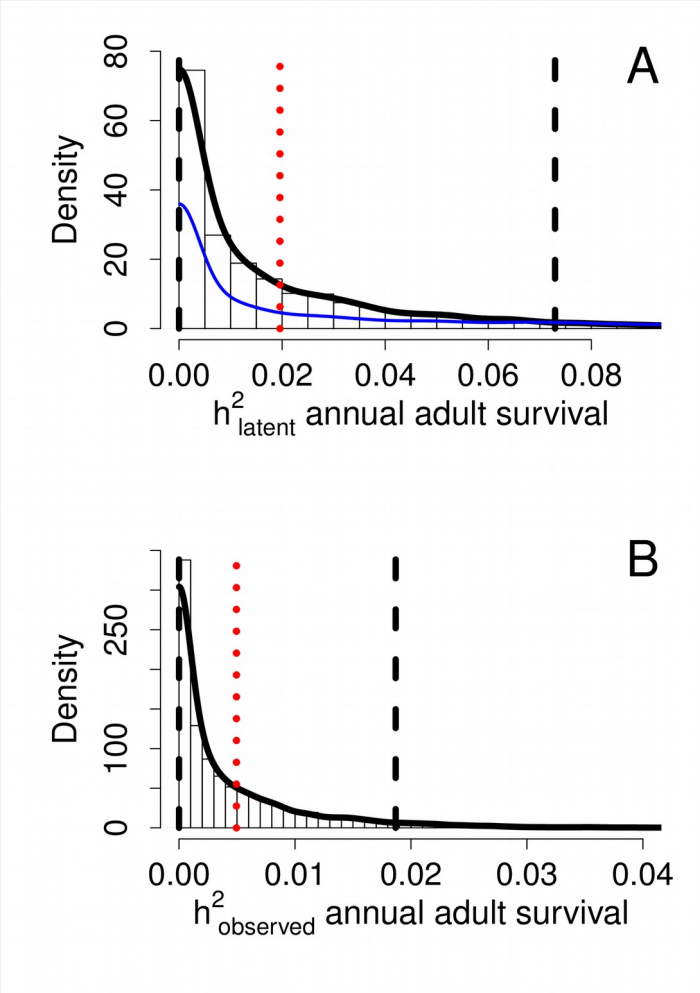
S5.4 Adult annual survival

Figure S8 depicts posterior distributions of heritabilities for adult annual survival. Adult survival at age one year was fitted as a sex-specific intercept, and effects of the other three age categories were estimated as combined-sex contrasts (Table S4).

Table S4. Marginal posterior means, modes (in square brackets), and 95% credible intervals (in parentheses) for age effects estimated from the univariate QGGLMM for adult female and male annual survival. Estimates for ages 2, 3-5 and ≥6 (for both sexes combined) are contrasts from the sex-specific intercept at age 1 year. All other model estimates are in table 3 (main text).

	Age 1	Age 2	Ages 3-5	Age ≥6
Female annual survival	0.05			
	[0.12]	0.23	-0.21	-1.03
	(-1.13, 1.29)	[0.29]	[6.4×10^{-4}]	[-0.84]
Male annual survival	0.26	(-0.24, 0.67)	(-0.98, 0.40)	(-2.54, 0.14)
	[0.31]			
	(-0.96, 1.54)			

Figure S8. Marginal posterior MCMC samples (bars), kernel density estimation (solid black line), posterior mean (red dotted line), 95% credible interval limits (black dashed lines), and prior (solid blue line) for the latent-scale heritability (h^2_{latent}, top panel) and observed-scale heritability ($h^2_{observed}$, bottom panel) in adult annual survival. The prior for h^2_{latent} is depicted over the range of the posterior distribution, but extends to values of one. Note that axis scales differ between panels.



APPENDIX S6: PRIOR SENSITIVITY ANALYSIS

To examine whether inferences from posterior distributions of our four primary QGGLMMs were substantially influenced by prior distributions we reran each QGGLMM with two alternative priors. The alternative parameter expanded priors were inverse wishart distributed, with different shape and scale parameters. For example, for the bivariate QGGLMM for female and male fitness (which estimated the additive genetic variance-covariance matrix G_A, the cohort variance-covariance matrix G_C, and independent sex-specific residual variances R), prior specifications for MCMCglmm (Hadfield 2010) were as follows:

Prior 1 (used in the main analysis):

G_A: V = diag(2)*0.02, nu = 3, alpha.mu = c(0,0), alpha.V = diag(2)*1000

G_C: V = diag(2)*0.02, nu = 3, alpha.mu = c(0,0), alpha.V = diag(2)*1000

R: V = diag(2), nu = 2

This formulation for G_A gives a relatively flat (i.e. uninformative) prior on the cross-sex genetic correlation (r_A).

Prior 2:

G_A: V = diag(2), nu = 2, alpha.mu = c(0,0), alpha.V = diag(2)*1000

G_C: V = diag(2), nu = 2, alpha.mu = c(0,0), alpha.V = diag(2)*1000

R: V = diag(2), nu = 2

This formulation for G_A gives a weakly 'U-shaped' prior on r_A, with greater density near -1 and 1.

Prior 3:

G_A: V = diag(2), nu = 1.002, alpha.mu = c(0,0), alpha.V = diag(2)*1000

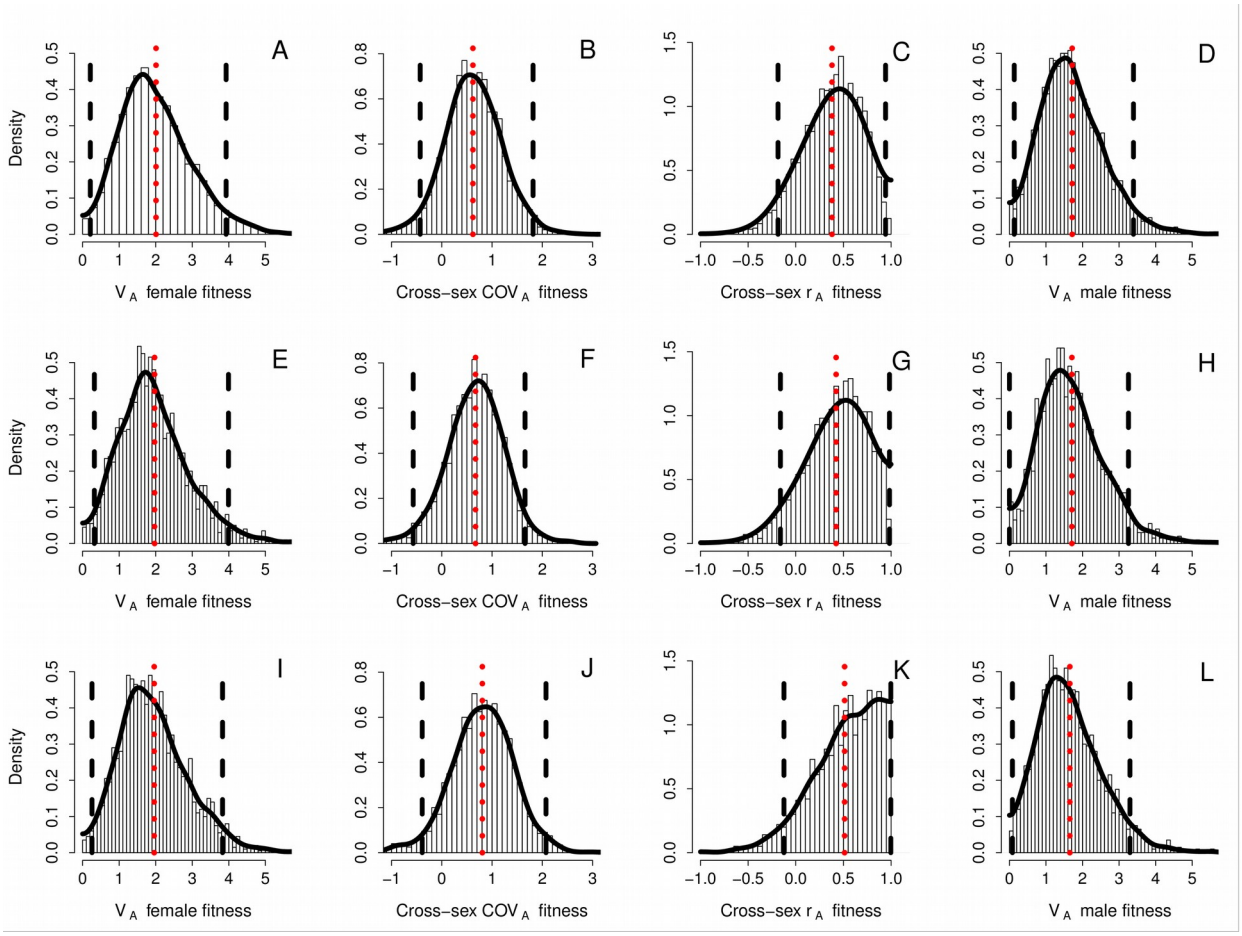
G_C: V = diag(2), nu = 1.002, alpha.mu = c(0,0), alpha.V = diag(2)*1000

R: V = diag(2), nu = 1.002

This formulation for G_A gives a more strongly 'U-shaped' prior on r_A, with greater density near -1 and 1.

Bivariate QGGLMMs for female and male fitness that used the three different prior formulations gave very similar posterior means, modes, and 95% credible intervals for the sex-specific additive genetic variances and the cross-sex additive genetic covariance and correlation (Figure S9), and for all other model components. Key conclusions were therefore robust to reasonable alternative priors. Further analyses showed that posterior distributions for the other three main QGGLMMs were also robust to similar sets of alternative priors (full details in R code).

Figure S9. Marginal posterior MCMC samples (bars), kernel density estimation (solid black line), posterior mean (red dotted line), and 95% credible interval limits (black dashed lines) for the additive genetic variance (V_A) in female fitness (left column), V_A in male fitness (right column), and the cross-sex additive genetic covariance (COV_A, center-left column) and correlation (r_A, center-right column). The top, middle, and bottom rows depict posterior distributions given prior specification 1, 2, and 3 respectively (see above).


A
B
C
D
E
F
G
H
I
J
K
L
Density
V A female fitness
Cross−sex COV A fitness
Cross−sex r A fitness
V A male fitness

APPENDIX S7: SEX-SPECIFIC JUVENILE SURVIVAL

We fitted an additional bivariate QGGLMM for sex-specific juvenile survival to estimate immigrant genetic group effects on sex-specific juvenile survival, and to further verify whether it was appropriate to model juvenile survival as a joint trait of both sexes. The model estimated sex-specific V_A in juvenile survival and the cross-sex COV_A, alongside sex-specific cohort variances and covariances and sex-specific intercepts and regressions on individual coefficient of inbreeding (*f*), immigrant genetic group coefficient (*IGG*), and first egg lay date (see main methods). Juvenile survival was modeled as a binary trait for each sex, with logit link function and sex-specific residual variances fixed to one. The posterior distribution of the cross-sex genetic correlation (r_A) was computed as per Appendix S4.

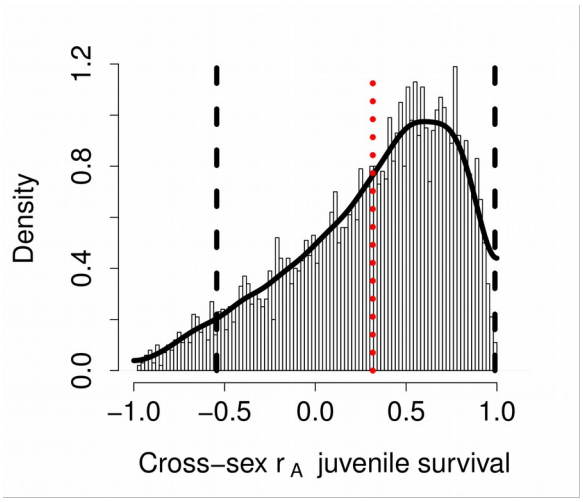
The posterior means for V_A in juvenile survival were moderate in both sexes, although the posterior modes were small and the lower 95%CI limits converged towards zero, probably due to relatively low power to estimate V_A for each sex separately (Table S5). Consequently, the cross-sex r_A was estimated with substantial uncertainty, but the posterior mean and mode were both moderately positive (Fig. S10, Table S5). This implies that there is overlap in genetic effects on juvenile survival in females and males, such that treating these traits as a joint trait yields a posterior distribution of V_A that is sufficiently similar to the posterior distributions of each sex-specific V_A to be used in the trivariate model with adult female and male LRS. Consequently, given currently available data, juvenile survival is best modeled as a single trait that is expressed by both sexes.

This analysis also showed that the slopes of the regressions of juvenile survival on *IGG* were similarly negative in females and males, with 95%CIs that did not overlap zero (Table S5). This implies that alleles originating in immigrants have similar negative effects on local juvenile survival in both sexes. There was also strong inbreeding depression, and negative effects of later lay dates, in juvenile survival in both sexes (Table S5).

Table S5. Marginal posterior means, modes (in square brackets), and 95% credible intervals (in parentheses) for key latent-scale estimates from a bivariate QGGLMM for sex-specific juvenile survival. Within the additive genetic matrix, variances are shown along the diagonal (bold) with covariances (COV) and correlations (r, italics) above and below the diagonal respectively. Sex-specific intercepts and slopes of regressions on individual coefficient of inbreeding (β_f), immigrant genetic group coefficient (β_{IGG}), and first egg lay date ($\beta_{lay.date}$) are also shown. Lower 95% CI limits that converged towards zero are reported as <0.001.

	Additive genetic matrix		**Intercept**	β_f	β_{IGG}	$\beta_{lay.date}$
	Female	Male				
Female	$\mathbf{V_A = 0.45}$	$COV_A = 0.12$	-0.43	-7.97	-2.75	-0.27
	[0.01]	[0.001]	[-0.34]	[-7.54]	[-2.48]	[-0.26]
	(<0.001, 1.03)	(-0.12, 0.44)	(-1.87, 1.14)	(-13.20, -2.91)	(-4.97, -0.60)	(-0.44, -0.08)
Male	*$r_A = 0.36$*	$\mathbf{V_A = 0.30}$	0.16	-10.50	-2.57	-0.19
	[0.63]	**[0.01]**	[0.06]	[-10.30]	[-2.47]	[-0.21]
	(-0.40, 0.99)	**(<0.001, 0.78)**	(-1.20, 1.54)	(-15.20, -5.62)	(-4.63, -0.62)	(-0.37, -0.01)

Figure S10. Marginal posterior MCMC samples (bars), kernel density estimation (solid black line), posterior mean (red dotted line), and 95% credible interval limits (black dashed lines) for the cross-sex additive genetic correlation (r_A) in juvenile survival.



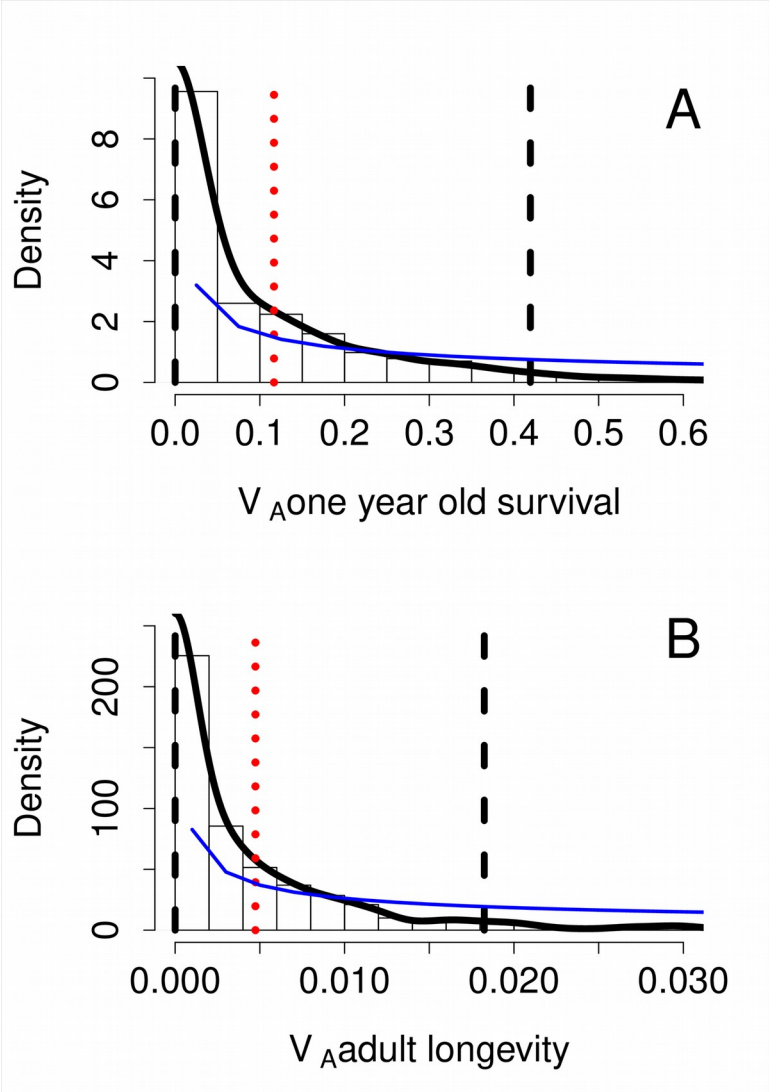
APPENDIX S8: ADDITIONAL ANALYSES OF ADULT SURVIVAL

Our fourth main QGGLMM, that treated adult survival as an annual binary trait for both sexes combined, suggested that there is very little V_A (Table 3). Further, QGGLMMs fitted to observations of survival or mortality for females and males separately returned similar conclusions. However, to examine whether V_A in adult annual survival is indeed small we fitted two QGGLMMs to two additional measures of adult survival.

The first model estimated V_A in survival through an individual's first year of adult life. The defined binary trait quantified whether or not each one-year old adult survived to age two (giving a single observation per adult). The model used a logit link function and additionally estimated year variance, with residual variance fixed to one (by convention). The second model estimated V_A in adult longevity, with longevity measured as age at death (again giving a single observation per adult). The model implemented a Poisson distribution with log link function and estimated cohort variance and residual variance assuming additive overdispersion.

These two models estimated V_A in adult first-year survival and longevity to be small (Fig. S11), reinforcing the conclusion drawn from the main analysis of adult annual survival (Fig. 5).

Figure S11. Marginal posterior MCMC samples (bars), kernel density estimation (solid black line), posterior mean (red dotted line), 95% credible interval limits (black dashed lines), and prior (solid blue line) for additive genetic variance (V_A) in (A) adult first-year survival and (B) adult longevity. Note that values of V_A in panels A and B are estimated on different latent scales and hence are not quantitatively comparable.



APPENDIX S9: ADDITIONAL ANALYSIS OF ADULT ANNUAL SURVIVAL AND ANNUAL REPRODUCTIVE SUCCESS

Our analyses of adult lifetime reproductive success (LRS) indicate that there are different magnitudes of V_A in female LRS versus male LRS (Table 2, Fig. 3). One potential explanation is that there are sex-specific trade-offs between the fitness components that generate adult LRS, namely adult annual survival and annual reproductive success (ARS). Specifically, ARS might be negatively genetically correlated with annual survival in females but positively genetically correlated in males, thereby reducing and increasing V_A in LRS respectively. One logical next step is to estimate the r_A between sex-specific adult annual survival and ARS.

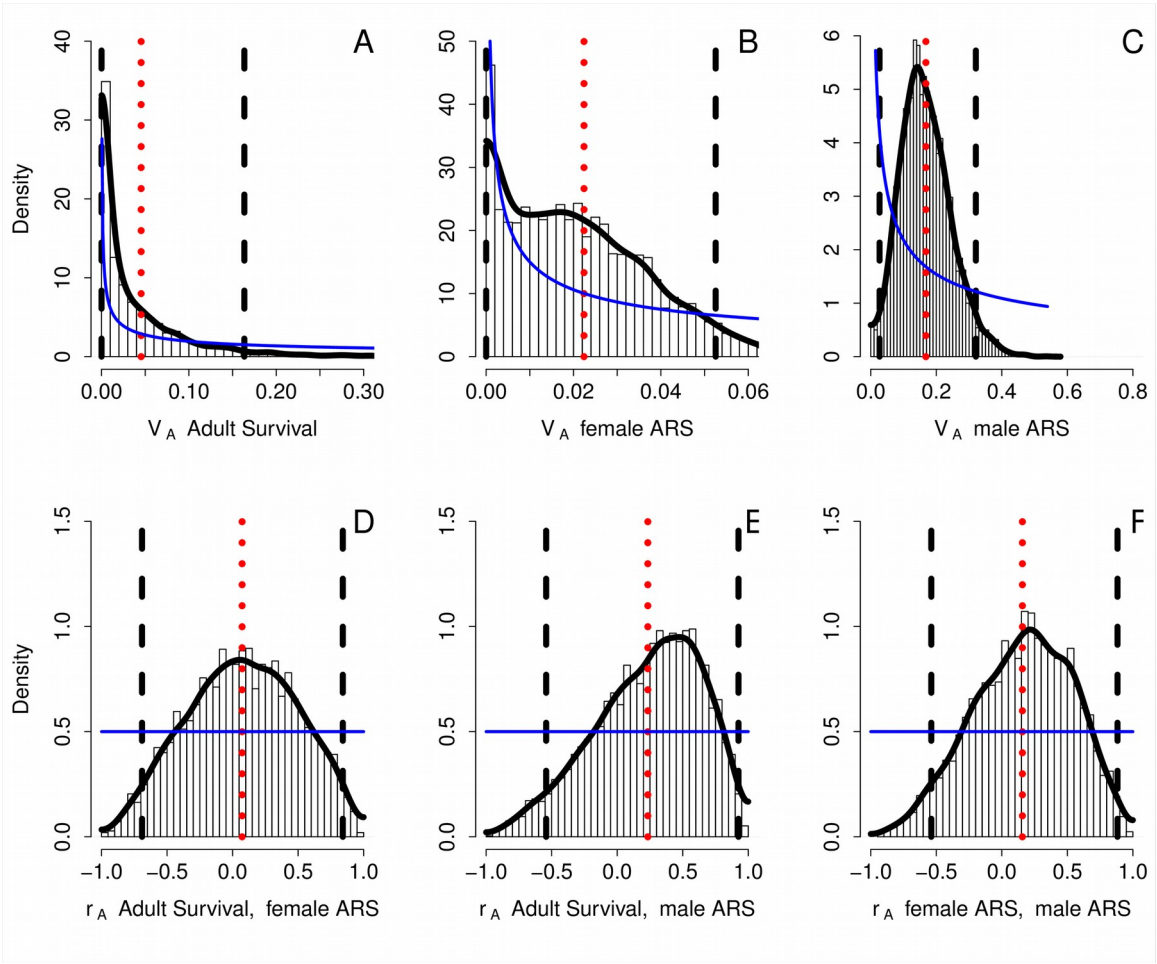
However, such analyses are problematic, and perhaps not required, given the song sparrow data. This is because the evidence suggests that there is very little V_A in adult annual survival, implying that genetic correlations with ARS cannot exist and/or are unlikely to be estimable with any useful precision. This depends on whether V_A in adult annual survival is truly zero or is non-zero but very small (and hence hard to estimate). Given the former, r_A is undefined. Given the latter, r_A can, in principle, still be estimated using a bivariate model. Indeed, V_A in adult annual survival may itself be estimated more precisely using a bivariate model if the absolute magnitude of the r_A with ARS is moderate or high, because of the increased information coming from correlated expression of ARS.

Therefore, to attempt to estimate the r_As between adult annual survival and female ARS and male ARS, we fitted a further trivariate QGGLMM. The overall model structure, justifications, and implementation were similar to the separate models of adult annual survival and ARS presented in the main Methods, with additional covariances.

As expected, this model returned marginal posterior distributions for V_A in female and male ARS that were extremely similar to those from the bivariate model for female and male ARS (Fig. S12B,C compared to Fig. 4A,B). The posterior distribution for V_A in adult annual survival was also very similar to that from the univariate model (Fig. S12A compared to Fig. 5C), further implying

that there is very little V_A in adult annual survival. Consequently, the posterior distribution of r_A between adult annual survival and female ARS (Fig. S12D) is very broad, indicating a posterior mean and mode around zero (mean=0.07, mode=0.10) but with considerable uncertainty (95%CI: -0.69, 0.84). However, the lower 95%CI limit did not converge towards -1 (Fig. 12D). Hence, there is no evidence that a trade-off between adult annual survival and female ARS acts as an absolute constraint on the magnitude of V_A in adult female LRS. Meanwhile, the posterior mean and mode for r_A between adult annual survival and male ARS were slightly positive (mean=0.23, mode=0.58), but again estimated with considerable uncertainty (95%CI: -0.54, 0.93; Fig. S12E).

Figure S12. Marginal posterior MCMC samples (bars), kernel density estimation (solid black line), posterior mean (red dotted line), 95% credible interval limits (black dashed lines), and prior (solid blue line) for additive genetic variance (V_A) in (A) adult annual survival, (B) female annual reproductive success (ARS), (C) male ARS, and additive genetic correlations (r_A) between (D) adult annual survival and female ARS, (E) adult annual survival and male ARS, and (F) female and male ARS.



LITERATURE CITED

Brommer, J. E. 2000. The evolution of fitness in life-history theory. Biol. Rev. 75:377–404.

Brommer, J. E., J. Merilä, and H. Kokko. 2002. Reproductive timing and individual fitness. Ecol. Let. 5:802-810.

Coulson, T., T. G. Benton, P. Lundberg, S. R. X. Dall, B. E. Kendall, and J. –M. Gaillard. 2006. Estimating individual contributions to population growth: evolutionary fitness in ecological time. Proc. R. Soc. B 273:547-555.

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

Falconer, D. S. 1989. Introduction to Quantitative Genetics. 3rd ed. John Wiley & Sons, Inc., New York.

Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. 7:457–472.

Germain, R., P. Arcese, and J. M. Reid. 2018. The consequences of polyandry for sibship structures, distributions of relationships and relatedness, and potential for inbreeding in a wild population. Am. Nat. in press.

Germain, R. R., M. E. Wolak, P. Arcese, S. Losdat, S., and J. M. Reid. 2016. Direct and indirect genetic and fine-scale location effects on breeding date in song sparrows. J. Anim. Ecol. 85:1613–1624.

Geyer, C. J., S. Wagenius, and R. G. Shaw. 2007. Aster models for life history analysis. Biometrika 94:415-426.

Gomulkiewicz, R., and R. G. Shaw. 2013. Evolutionary rescue beyond the models. Philos. Trans. R. Soc. B 368:20120093.

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

Hendry, A. P., D. J. Schoen, M. E. Wolak, and J. M. Reid. 2018. The contemporary evolution of fitness. Ann. Rev. Ecol. Evol. Syst. *In press*. https://doi.org/10.1146/annurev-ecolsys-110617-062358.

Houle, D. 1992. Comparing evolvability and variability of quantitative traits. Genetics 130:195–204.

Keller, L. F. 1998. Inbreeding and its fitness effects in an insular population of song sparrows (*Melospiza melodia*). Evolution 52:240–250.

Kruuk, L. E. B. 2004. Estimating genetic parameters in natural populations using the "animal model." Philos. Trans. R. Soc. B 359:873–890.

Kruuk, L. E. B., and J. D. Hadfield. 2007. How to separate genetic and environmental causes of similarity between relatives. J. Evol. Biol. 20:1890–1903.

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

Lebigre, C., P. Arcese, R. J. Sardell, L. F. Keller, and J. M. Reid. 2012. Extra-pair paternity and the variance in male fitness in song sparrows (*Melospiza melodia*). Evolution 66:3111–3129.

Marr, A. B., L. F. Keller, and P. Arcese. 2002. Heterosis and outbreeding depression in descendants of natural immigrants to an inbred population of song sparrows (*Melospiza melodia*). Evolution 56:131-142.

Metcalf, C. J. E., and S. Pavard. 2007. Why evolutionary biologists should be demographers. Trends Ecol. Evol. 22:205-212.

Orr, H. A. 2009. Fitness and its role in evolutionary genetics. Nat. Rev. Genet. 10:531–539.

Reid, J. M., P. Arcese, and L. F. Keller. 2006. Intrinsic parent-offspring correlation in inbreeding level in a population open to immigration. Am. Nat. 168:1-13.

Reid, J. M., P. Arcese, L. F. Keller, and S. Losdat. 2014. Female and male genetic effects on offspring paternity: additive genetic (co)variances in female extra-pair reproduction and male paternity success in song sparrows (*Melospiza melodia*). Evolution 68:2357–2370.

Reid, J. M., P. Arcese, R. J. Sardell, and L. F. Keller. 2011. Heritability of female extra-pair paternity rate in song sparrows (*Melospiza melodia*). Proc. R. Soc. B. 278:1114-1120.

Sæther, B.-E., and S. Engen. 2015. The concept of fitness in fluctuating environments. Trends Ecol. Evol. 30:273–281.

Shaw, R. G., C. J. Geyer, S. Wagenius, H. H. Hangelbroek, and J. R. Etterson. 2008. Unifying life-history analyses for inference of fitness and population growth. Am. Nat. 172:E35-E47.

Shaw, R. G., and J. R. Etterson. 2012. Rapid climate change and the rate of adaptation: insight from experimental quantitative genetics. New Phytol. 195:752-765.

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

Smith, J. N. M., L. F. Keller, A. B. Marr, and P. Arcese (eds). 2006. Conservation and Biology of Small Populations: The Song Sparrows of Mandarte Island. Oxford University Press, Oxford.

Walling, C. A., M. B. Morrissey, K. Foerster, T. H. Clutton-Brock, J. M. Pemberton, and L. E. B. Kruuk. 2014. A multivariate analysis of genetic constraints to life history evolution in a wild population of red deer. Genetics 198:1735–1749.

Walsh, B., and M. W. Blows. 2009. Abundant genetic variation + strong selection = multivariate genetic constraints: a geometric view of adaptation. Annu. Rev. Ecol. Evol. Syst. 40:41–59.

Wolak, M. E. 2012. nadiv: an R package to create relatedness matrices for estimating non-additive genetic variances in animal models. Methods Ecol. Evol. 3:792–796.

Wolak, M. E., P. Arcese, L. F. Keller, P. Nietlisbach, and J. M. Reid. 2018. Data from: Sex-specific additive genetic variances and correlations for fitness in a song sparrow (*Melospiza melodia*) population subject to natural immigration and inbreeding. Dryad Digital Repository. https://doi.org/10.5061/dryad.p7p1jb3.

Wolak, M. E., and J. M. Reid. 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.