the genetics society

Check for updates

ARTICLE

Differential effects of steroid hormones on levels of broad-sense heritability in a wild bird: possible mechanism of environment × genetic variance interaction?

Szymon M. Drobniak[1,2,3 ✉], Joanna Sudyka[1], Mariusz Cichoń[1], Aneta Arct[1,2,4], Lars Gustafsson[2], Dorota Lutyk[1,5] and Katarzyna Janas[1,5]

Genetic variation is one of the key concepts in evolutionary biology and an important prerequisite of evolutionary change. However, we know very little about processes that modulate its levels in wild populations. In particular, we still are to understand why genetic variances often depend on environmental conditions. One of possible environment-sensitive modulators of observed levels of genetic variance are maternal effects. In this study we attempt to experimentally test the hypothesis that maternally transmitted agents (e.g. hormones) may influence the expression of genetic variance in quantitative traits in the offspring. We manipulated the levels of steroid hormones (testosterone and corticosterone) in eggs laid by blue tits in a wild population. Our experimental setup allowed for full crossing of genetic and rearing effects with the experimental manipulation. We observed that birds treated with corticosterone exhibited a significant decrease in broad-sense genetic variance of tarsus length, and an increase in this component in body mass on the 2nd day post-hatching. Our study indicates, that maternally transmitted substances such as hormones may have measurable impact on the levels of genetic variance and hence, on the evolutionary potential of quantitative traits.

INTRODUCTION

Evolutionary change relies on the existence of genetic variance in phenotypic traits (Fisher 1930; Lande and Arnold 1983; Lande and Shannon 1996). According to the general theorem of selection, evolutionary change in a phenotypic trait is equal to the genetic covariance between the trait and fitness (Price 1970; Lynch and Walsh 1998; Teplitsky et al. 2014; Walsh and Lynch 2018). Most of the available evidence for the role of genetic variance in trait evolution comes from laboratory populations and planned breeding studies (including agricultural and artificial selection for specific, desirable properties of organisms) (Lynch and Walsh 1998; Drobniak and Cichoń 2016), which may bias genetic parameters by exposing organisms to conditions unlike those experienced in nature. Far less estimates come from natural populations, even in spite of an increase in numbers of heritability estimates from wild populations (Postma 2014).

Genetic variation arises through mutations and gene interactions (Lynch and Walsh 1998; Lai et al. 2019), and large portions of it remain cryptic (Masel 2006). However, under certain conditions specific components of this variance may increase or decrease. Levels of genetic variance observed in nature may vary with respect to a number of factors (Merilä and Fry 1998; Charmantier and Garant 2005; Oltman et al. 2005; Brommer et al. 2008; Pitala et al. 2009; Galloway et al. 2009; Gunay et al. 2011; Schroeder et al. 2012; Rowiński and Rogell 2017). For instance, some studies demonstrated that stressful environments may induce decreases in the observed levels of genetic variance, sometimes even to the point of completely removing estimable genetic variance in them (Hoffmann and Merilä 1999; Teplitsky et al. 2014). Sexes also can differ in heritabilities of sex-specific variants of certain traits (Jensen et al. 2003; Foerster et al. 2007; Poissant et al. 2010; Wyman and Rowe 2014). Also, other studies suggested that local environment generated by parents could also influence observed heritabilities. Such "environment" may not always reflect typical notions of external habitats and can for example encompass characteristics of parents (e.g. parental age, a characteristic demonstrated to affect genetic variance in the offspring (Kim et al. 2011; Drobniak et al. 2015)). As variation in quantitative genetic parameters is one of the factors proposed to contribute to maintaining genetic variance in the wild (Hoffmann and Merilä 1999; Gienapp and Brommer 2014; Teplitsky et al. 2014), studying it is one of the most important avenues in evolutionary research.

Mechanisms underlying the abovementioned modifications of genetic variances are largely unknown. Irrespectively of the underlying factor influencing genetic variance, several hypotheses were brought up to explain observed patterns of heritability variation. It is possible that external/local/internal environments experienced by individuals modulate expression levels of genes at

[1]Institute of Environmental Sciences, Jagiellonian University, Kraków, Poland. [2]Department of Animal Ecology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden. [3]Evolution & Ecology Research Centre, School of Biological, Environmental and Earth Sciences, University of New South Wales, Sydney, NSW, Australia. [4]Institute of Systematics and Evolution of Animals, Polish Academy of Sciences, Kraków, Poland. [5]These authors contributed equally: Dorota Lutyk, Katarzyna Janas. Associate editor Sara Knott.
✉email: szymek.drobniak@gmail.com

the molecular level, resulting in the observed patterns in quantitative genetic parameters (Jensen et al. 2003; Fox and Wolf 2006). Mutation accumulation, specific to certain environments or individual characteristics, could also be responsible for such patterns (such explanation was so far considered mainly in the context of senescence and genetic variance increase with age) (Wilson et al. 2007; Charmantier et al. 2014). Finally, rich work on molecular mechanisms able to release existing cryptic genetic variation suggests that mechanisms such as heat-shock proteins (HSPs), prions or alternative splicing events may change conformational states of involved proteins, and effectively add variation on top of this present due to the existence of different allelic variants of genes (Queitsch et al. 2002; Bergman and Siegal 2003; Gibson and Dworkin 2004; Masel 2006; Berger 2011). To date all studies approaching the condition-dependent expression of genetic variance either focused on quantitative genetic estimates of heritabilities under varying rearing conditions (Hoffmann and Merilä 1999; Charmantier and Garant 2005) or explored evolutionary capacitance by creating genetic variants with alterations in specific molecular systems (e.g. prions, True and Lindquist 2000; HSPs, Queitsch et al. 2002). However, such approaches are largely correlative (e.g. assaying individual traits under existing conditions) or use approaches (genetic knockouts or induced mutants) that are unlikely to reflect responsiveness of physiological systems to natural environments. Literature lacks proper experimental tests of the impact of well-known environment-sensing physiological mechanisms (e.g. hormonal milieu, oxidative status, neuronal signalling) on the expression of quantitative genetic variance, and hence – traits' evolutionary potential.

In this study, we aimed at filling this gap by directly following one of the possible mechanisms. One of the ways environments can impact the development and fitness of individuals is through parental effects, i.e. consistent effects of parents on their offspring phenotype which, in certain cases, can correlate with environmental variability (Groothuis and Schwabl 2008). Maternal effects received more attention, perhaps because maternal phenotypes can influence—or are perceived as such—offspring traits in more ways (e.g. through maternal transfer of resources and biologically active compounds to the offspring at the stage of eggs or developing embryos (Groothuis and Schwabl 2008; Wolf and Wade 2009; Coslovsky et al. 2012)). Condition-dependent expression of genetic variance, also known as G × E (genotype-by-environment interaction, which essentially is a form of plasticity in genetic variances), was repeatedly shown across different types of environmental gradients (reviewed in Charmantier and Garant (2005) & Rowiński and Rogell (2017)), e.g. between favourable and resource-limited rearing conditions (Gebhardt-Henrich and Noordwijk 1991; Merilä 1997; Merilä and Fry 1998; Garant et al. 2005). At the same time, environmental heterogeneity that can drive such resource-related rearing environment changes has been demonstrated to modulate the amounts of hormones transmitted to eggs (Groothuis et al. 2005; Hegyi et al. 2011; Remeš 2011; Coslovsky et al. 2012). It is therefore likely that hormonally mediated maternal effect may be one of the factors mediating the observed patterns of G × E in wild populations. Steroid hormones are particularly interesting in this context: they have profound effects on offspring development (Hayward and Wingfield 2004; Tschirren et al. 2005 2009; Tobler and Sandell 2007; Coslovsky et al. 2012; Ruuskanen et al. 2012; Schweitzer et al. 2013; Lutyk et al. 2017) and they can directly impact the expression of genes by acting as transcription modulators in the nuclei of cells after binding to their specific receptors (Kawata 1995; Baker 1997; Podmokła et al. 2018). As such, steroid hormones are well documented as mediators of maternal effects, and one reason for this is the ease of manipulating their levels in eggs of wild birds (Groothuis and Schwabl 2008).

Here we employed a direct manipulation of levels of two steroid hormones (testosterone, TESTO henceforth, and corticosterone, CORT henceforth) in eggs of the blue tit (*Cyanistes caeruleus*)—a model species in evolutionary ecology. Our experiment involved cross-fostering (Kruuk and Hadfield 2007) at the egg stage and a fully crossed, factorial design ensuring that genetic and environmental sources of trait variation in these birds would be fully interacted with the hormone level manipulation. The choice of hormones was deliberate: testosterone and corticosterone are well studied, both in a laboratory and wild population contexts (Podmokła et al. 2018) and where repeatedly shown to serve as mediators of environmental cues (Groothuis et al. 2005; Groothuis and Schwabl 2008; Lessells et al. 2016). They differ in physiological impact and often mediate different kinds of information (corticosterone being a well-established mediator of stress responses, whereas testosterone being involved in primary sexual characters development and reproductive investment regulation (Groothuis and Schwabl 2008)), although it should be noted that corticosterone is far less studied in evolutionary ecology contexts than testosterone. We predicted that the levels of genetic variation in certain phenotypic traits would be affected by this manipulation: assuming that by supplementing hormones we would simulate a stressful (CORT) or male-like sex-specific (TESTO) reaction, we expected a decrease in genetic variance in hormone-treated birds, compared to control birds receiving a sham manipulation (Merilä and Fry 1998; Jensen et al. 2003). The traits we chose comprised a set of body size descriptors (weights at different ages and tarsus length) and a frequently measured proxy of immunological state of individuals, the phytohaemagglutinin hyperreactivity response. Earlier studies suggested the existence of G×E in all these traits in relation to a number of environmental factors (Merilä and Fry 1998; Ruuskanen et al. 2012; Drobniak et al. 2015).

METHODS
General field methods
The experiment was performed in a wild population of blue tits, studied since 2002 on the Baltic island of Gotland, Sweden (57°01′ N; 18°16′ E) in three breeding seasons (2014–2016). In this population blue tits breed in wooden nest-boxes distributed uniformly across 23 study plots of varying size; density of breeding pairs is uniform across plots of different size (unpublished data). Most plots are covered by oak (*Quercus robur*), ash (*Fraxinus excelsior*) and poplar (*Populus sp.*) forests, with dense common hazel undergrowth (*Corylus avellana*). Some plots lack the undergrowth and are covered by bright, loose oak forests with wet, rich meadows abundant in orchids. In the studied population, tits lay almost exclusively one clutch per year. Females lay on average 11 eggs (range: 5–17) and incubate them for 13 days; chicks fledge at the age of 17–20 days.

All breeding attempts were regularly inspected by visiting all available nest-boxes every 4–5 days, recording nest construction/egg laying stage, and determining species occupying each nest-box (apart from blue tits, the population is also home to great tits (*Parus major*) and collared flycatchers (*Ficedula albicollis*)). Selected nests were assigned to experimental triplets (see Fig. 1 and the next section for a more detailed description of the experiment). Figure 1 provides a summary of procedures and measurements performed in each nest. Parents in each nest were caught on the 14th day post-hatching with mist-nets, ringed with aluminium bands (if not having one already) and measured for tarsus length, wing & tail length and body weight. Age of adults was determined based on the presence of moult limits in the tail and between primary and secondary wing coverts (Demongin 2016). Sex was determined by the presence of a brood patch in females.

Experimental nestlings were also injected (in two of the three seasons) with a small dose of phytohaemagglutinin (PHA, Sigma-Aldrich, Germany) to determine their cell-mediated immune response (Sarv and Horak 2009). Briefly, 0.2 mg of PHA suspended in 400 ul of buffered saline was injected into the right wing-web of each nestling on the 11th day post-hatching. The thickness of the web prior to the injection, and 24 h afterwards was determined with three measurements using a pressure-sensitive spessimeter (Mitutoyo, Japan model 7313) to the nearest 0.01 mm. The difference between averaged triplets of "after" and "before" measurements quantifies the amount of swelling resulting from PHA hypersensitivity
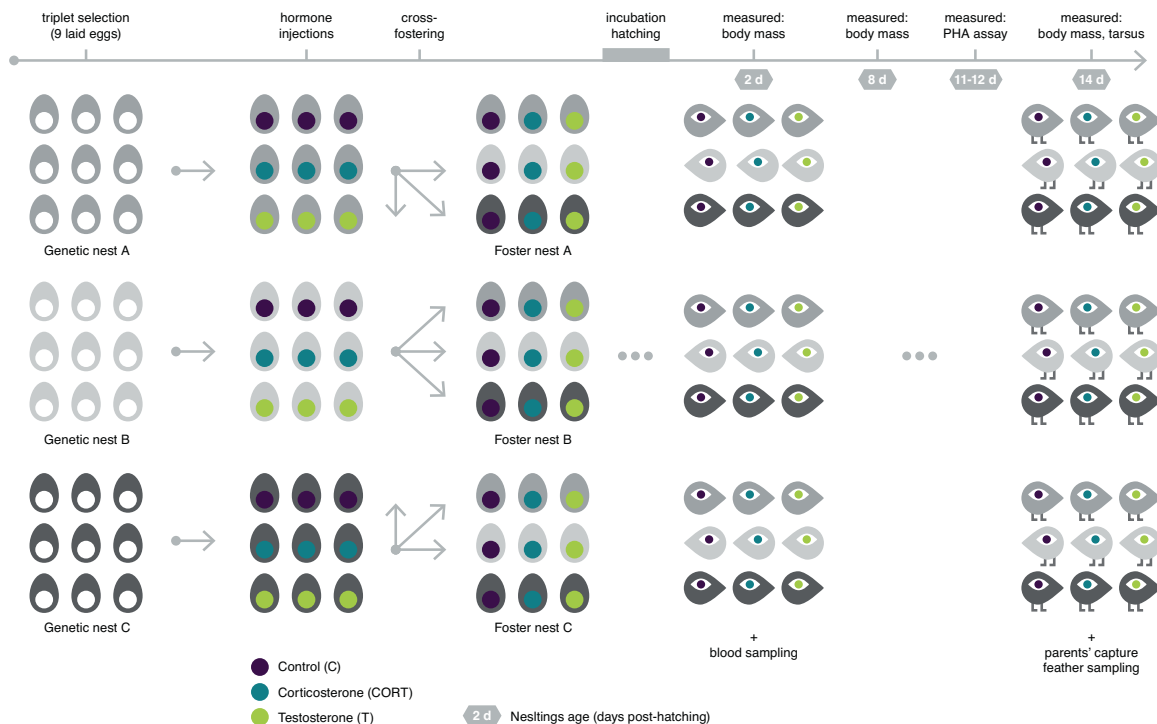
Fig. 1 Schematic summary of the experiment and different measurements performed during its course. The top time axis is not to scale. Represented egg/chick numbers may differ between nests; also, hatching failure or nestling mortality may lead to some individuals dropping out

reaction and is treated as a proxy of cell-mediated immune response. The three initial and three post-injection measurements were highly repeatable (technical repeatability $r^2 > 0.96$ ($p < 0.01$) in all cases). Wing web thickness is positively correlated with body size; to account for this, all assayed nestlings were also weighed on the 12th day. The PHA assay was performed only in years 2015 and 2016. Nonetheless, our analyses still are robust and valid: PHA treatment (or lack thereof) was always applied to all three hormonal groups (i.e. it could not generate observed differences between hormone-treated groups) and all nests not treated with PHA are grouped in one year (i.e. possible effects of not receiving PHA injections on other measured variables, however small, are linked to the year effect and fully explained by the year factor).

Blood samples retrieved from nestlings were used to determine the sex of each chick, using a well-established protocol described (Griffiths et al. 1998). Briefly, after isolating bird DNA a PCR was used to amplify a fragment of the chromohelicase (CHD) gene located on sex chromosomes and exhibiting a sex-specific length dimorphism, scored after separating the PCR products on an agarose gel. In some nestlings (Table S3) the sex could not be assigned due to technical reasons (not enough genetic material for reliable PCR, failure of the PCR reaction or ambiguous result with the gel bands markedly differing in intensity).

Field procedures conformed with the legal requirements of Sweden (permit from Jordbruksverket to LG; Swedish ringing licence RC712 to SMD).

Steroid-injection experiment

Nests inspected during egg-laying were grouped into triplets based on their equal laying dates. In each triplet, at the stage of 9 laid eggs, a hormone injection manipulation was performed. The eggs were taken out of their nests and safely transported to the field laboratory. Prior to collection, the eggs where candled using a battery-powered torch to make sure no signs of early incubations were visible. For the time of manipulation, females were left with an equal number of dummy plastic eggs. After transporting to the lab, the eggs were weighed, photographed, and labelled. Each egg was assigned by random to one of three experimental groups: testosterone group (TESTO), corticosterone group (CORT), and control group (C). Each egg was individually marked with a non-toxic marker and then injected with 3 ul of experimental solution. In group C this was pure sesame oil. In group TESTO each dose contained 1.7 ng of testosterone (17β-hydroxy-4-androsten-3-on; Sigma-Aldrich, Germany); in

the CORT group each dose contained 0.6 ng of corticosterone (11β,21-dihydroxyprogesterone; Sigma-Aldrich, Germany). Testosterone was dissolved directly in the sesame oil, whereas corticosterone (due to its poorer solubility in oil) was first dissolved in absolute ethyl alcohol (Gam et al. 2011), and then 10 μl of such stock dissolved in the sesame oil. To make the groups fully comparable, oil in the C and TESTO groups was also spiked with 10 μl of 100% ethanol. Nonetheless, resulting solutions quickly evaporate the residual ethyl alcohol, which anyway would be present in a concentration of 0.5% and less.

The doses of hormones were determined following a hormone assay on randomly chosen unmanipulated eggs from the studied population, sampled in preceding seasons (TESTO concentration ± SD: 2.13 ± 0.81 ng/yolk; CORT concentration: 0.61 ± 0.26 ng/yolk; $N = 10$). It should be noted that these values represent a single snapshot of hormonal concentrations—which may vary across seasons and individuals but should nevertheless provide a useful baseline. These values are close to published estimates from the blue tit and the closely related great tit (Tschirren et al. 2004; Vedder et al. 2007; Kingma et al. 2009; Lessells et al. 2016). Final doses were calculated as 2 SDs rounded up to the nearest 0.1 ng, therefore ensuring that the distribution of hormone concentrations in manipulated eggs would be shifted by 2 SD of their natural values (likely the shift would be slightly smaller due to downward bias of variance estimation based on small sample of wild eggs, but it still would be substantial).

Injections were performed using a 25 μl 702RN Hamilton (Hamilton, Nevada, USA) micro-syringe with type-4 26 s removable needle. Each experimental group had its own syringe; we also used several replacement needles kept in 96% ethanol to keep them clean and sterile. Prior to each injection the egg was gently swabbed with a small amount of ethanol to disinfect a portion of its shell. Then, a disposable sterile needle was used to make a small hole in the shell, and the Hamilton syringe was inserted through it, under the visual control thanks to a flashlight illuminating the egg from beneath it. To make sure that the content of the syringe was injected into the yolk, we performed several trial injections on eggs from deserted nests, using a food dye as the injected liquid. After freezing, these eggs were cut open to verify that the injection procedure delivered the liquid into the yolk and only there, which indeed was the case in 100% cases. Unfortunately, this verification would not indicate in how many cases the injection would compromise the integrity of the yolk membrane (freezing preserves the yolk shape, and even if severely damaged,

Table 1. Structure of random-effects models in successively simpler GLMMs fitted to the data

Model structure	Interpretation in terms of variance components
Gh; Nh; Rh	Models allowing for heterogenous genetic (nest-of-origin; G)/foster (nest-of-rearing; N)/residual (R) variances across treatment groups
G; N; R	Models assuming homogenous variances across experimental groups (i.e. only one variance estimated per random term)
X_r[a]	Cross-treatment correlations estimated
$X_{r=1}$; $X_{r=0}$	Cross-treatment correlations constrained at a value (zero or unity)
$Xh_{C \equiv T}$	Some variances or correlations constrained to be equal top reduce model complexity (here – variances for the C and TESTO groups)

"Model structure" describes each model symbolically and is later used in result tables. The following column describes the form of each random term for each model considered

E.g. $Gh_{r=1} + R + E$ implies heterogenous genetic variances (different V_G in treatment groups), correlations constrained to be identical and equal to unity, nest-of-rearing and residual variances homogenous

[a]These models are reported only in the Supplementary Materials in expanded model selection tables

its content would not leak out due to albumen pressure inside an egg). After injection the hole in egg's shell was closed with a drop of Vetbond (3M, Minnesota, USA), a tissue adhesive used in surgical procedures. All egg manipulations were performed on a clean table frequently swiped with ethyl alcohol to minimise risk of egg contamination.

After injection eggs were cross-fostered by randomly assigning each egg to one of the triplet nests. Egg randomisation was ensured by their blinded sampling from transportation container, just before marking each egg with a unique code. Afterwards the codes assigned to each rearing nest were matched, ensuring that one random egg from each origin nest by treatment combination ended up in a given rearing nest. Hence, this cross-fostering protocol ensured that all combinations of the experimental treatment and nest-of-origin were present in each nest-of-rearing. After cross-fostering the eggs were returned to their assigned nests and left there for incubation. On the following days any additional eggs were treated similarly (transported to the laboratory, injected with a randomly chosen sham/CORT/TESTO solution, and returned to a nest); this protocol was stopped once on a given day incubation commencement was noted (i.e. eggs were not covered with nest material and warm). 1–2 days before the expected hatching date (11–12 days after the incubation start) all experimental nests were visited again. After verifying the development stage of eggs (by egg candling), all injected eggs were again gently collected and transported in a warmed box to the lab (leaving females with dummy eggs to ensure they would not desert their nests). There, they were placed in individual paper containers and put in an incubator set to 38 degrees and 70% relative humidity for hatching. From that moment the eggs were checked every hour. All chicks hatched between hours 0500 and 2000 were weighed, marked individually by nail clipping and taken back to their foster nests within 1 h of hatching. Chicks hatched after 2000 were left in the incubator until 0500 and brought to their nests the following day. Hatching the nestlings in the incubator allowed us to assign the experimental group to each chick upon hatching.

In 14 cases codes assigned to eggs were not fully legible on egg collection or were modified due to errors in marking the eggs. In all such cases the identity of nestlings (i.e. their assignment to the experimental group) was successfully recovered before recording the relevant data in the database. However, to remain conservative, we checked if omitting these eggs would have any impact on the main effects seen in our study. This sensitivity analysis indicated no such bias.

Statistical analyses

Analyses included between 689 (2-days old chicks) and 621 (14-days old chicks) individuals (but substantially less in case of PHA response). Individuals came from 156 genetic nests and were reared in 143 nests (the latter number being lower due to brood desertions).

To determine patterns of genetic variances in the offspring traits, we applied linear mixed models, fitting them to measured phenotypic traits: body weight measurements at days 2, 8 and 14, tarsus length (on the 14th day post-hatching) and PHA hypersensitivity response. In all models, response variables were assumed to be normally distributed (assumption checked by visually inspecting model residuals plotted against fitted values). In all models we also visually verified the distribution of estimated BLUPs to make sure they are approximately normal.

Each model contained fixed categorical effects of sex (males, females, and unknown sex; females as intercept reference group), study year (2014–2016; 2014 as intercept reference group), experimental treatment (control, CORT, TESTO; control as intercept reference group) and sex by treatment interaction. When interpreting fixed-effect results, non-significant sex-treatment interactions were removed. In addition, the PHA model also included body mass on day 12 as PHA-response is known to be partially correlated with body mass. Random effects included: nest-of-origin (genetic family, **G** or **Gh** – see Table 1 for details of abbreviations) effect, nest-of-rearing (foster nest, **N** or **Nh**) effect and residual error effect (**R** or **Rh**). Preliminary analyses included a random term of the nest triad. It was subsequently removed from final models to simplify analysis as it consistently explained non-significant amounts of variance and its omission had no impact on other estimates.

To estimate the effect sizes (i.e. the differences in genetic variances between treatment groups) and their sampling variances we used parametric bootstrapping. Bootstrapping was applied to the simplest model with separate treatment-specific genetic variances. We generated 1000 random samples from each model expressed as: $\mathbf{y} = \mathbf{Xb} + \mathbf{Z_G u_G} + \mathbf{Z_N u_N} + \mathbf{e}$ (where $\mathbf{X}$ – appropriate design matrix of fixed effects; $\mathbf{b}$ – vector of fixed effects estimates; $\mathbf{Z_G}$ – design matrix of the nest-of-origin effect; $\mathbf{u_G} \sim N(\mathbf{0}, \mathbf{G})$ – vector o genetic effects with $\mathbf{G} = \mathbf{I} \otimes \mathbf{V_G}$, $\mathbf{V_G}$ being the estimated 3×3 covariance matrix for treatment groups; $\mathbf{Z_N}$ and $\mathbf{u_N} \sim N(\mathbf{0}, \mathbf{N})$, $\mathbf{N} = \mathbf{I} \otimes \mathbf{V_N}$ – design matrix and vector of nest-of-rearing effects; $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R})$, $\mathbf{R} = \mathbf{I} \otimes \mathbf{V_R}$ – vector of random error deviations sampled from; $\mathbf{V_R}$ and $\mathbf{V_N}$ were either scalar variances or 3×3 covariance matrices, depending on the form of the best model for a given response. Each sample was then re-analysed with an appropriate model to generate 1000 estimates for treatment-specific genetic variances, from which the distributions of differences in genetic variance between treatment groups were extracted. Summary statistics were calculated as means of bootstrap samples, their 95% confidence intervals obtained as 2.5 and 97.5 quantiles, and respective p values calculated as proportions of bootstrap samples above (for negative estimates)/below (for positive estimates) zero. We have also performed a parallel bootstrapping analysing by resampling only the residuals of each model (i.e. keeping the random effects fixed across all samplings)—but this generated qualitatively identical results and is not presented. At the stage of refitting the models with the resampled data, we discarded all models that failed to converge (i.e. in reality we generated >1000 samples, to have a final collection of at least 1000 estimates).

The choice of models eventually analysed using bootstrapping was based on a likelihood-driven simplification of random term structures (i.e. we aimed at models that maximised likelihood and model parsimony). All random effects were structured to allow for hormone-specific estimates of relevant variances. In all cases the variance structures were set to allow for heterogenous variances among the three experimental groups (3×3 covariance matrices; **Xh** models – Table 1). In the end we have fitted, for each response variable, a set of decreasingly complex models, testing various aspects of model variance structures, starting with the most complex model (that included all possible treatment-specific variances) and simplifying it to remove redundant terms. The order of tests made sure that factors possibly confounding the genetic variance (i.e. the treatment-specific residual and rearing variances) were tested first. The residual correlations between treatment groups are not identifiable in our experimental design and were not estimated (fixed at $r = 0$).

To reduce the complexity of models and simplify presentation of results we have also fixed the cross-treatment correlations (which anyway require considerably greater power to estimate, compared to variances) for genetic and rearing effects at $r = 1$ (their expected value). Table 1 provides an overview of all types of fitted models, and (co)variance constraints involved. In Supplementary Materials we provide an expanded sequence of tested models, including the intermediate stages of testing whether the cross-treatment correlations are lower than $r = 1$, as well as correlations estimates. In each successive step the more complex model was tested against the simplified one using a likelihood-ratio test. Logged ratios of model likelihoods $d = 2\log(\ell_{Ha}/\ell_{Ho})$ were assumed to be distributed as a mixture of χ^2 variates with $k \in \{s, 1, 2, \ldots, s+q\}$ degrees of freedom (Self and Liang 1987; Stram et al. 1994). The asymptotic distribution against which each likelihood-ratio statistic is tested is

$$d \sim \sum_{k=q}^{s+q} \binom{q}{k-s} 2^{-q} \chi_k^2$$

where s – the number of tested (co)variance parameters that lie inside the parameter space (e.g. correlations/covariances, where the tested hypothesis is $\theta = 0$), q – number of tested (co)variance parameters restricted by null-hypothesis at the boundary of their parameter space (e.g. $\theta = 0$ for variances, $\theta = 1$ or -1 for correlations) (Self and Liang 1987). For a simple $df = 1$ test of one variance component (H_0: $\sigma^2 = 0$) this simplifies to rescaling the p value of the resulting test by 0.5: $p = 0.5[1 - P(\chi_{df=1}^2 \leq d)]$.

In addition to single-trait models that focused on differences between experimental groups we have also fitted a multivariate model that included tarsus length and body weights at days 2, 8 and 14, to estimate between-trait genetic correlations across three experimental groups. In this model all (co)variance matrices were assumed to have a block-diagonal structure (i.e. did not allow for correlations between different traits measured in different experimental groups), and estimated all cross-trait correlations at all random effects' levels. We present these results in the Supplementary Materials section and discuss their power considerations in the Discussion.

Estimated variances were used to calculate heritabilities within experimental groups/traits (ratios of genetic variance to the sum of other variance components). Standard errors of were derived using the delta method. Similar method was applied to calculations of genetic correlations presented in the Supplement. Since the model we employed assumes the chicks in one nest of origin are full-siblings, the nest-of-origin effect estimates ½ of the total genetic variance in traits (which includes additive genetic variance, dominance variance if present, as well as variance generated by early maternal effects) and so heritability estimates we derive here are *sensu lato*. Dominance is assumed to be negligible in similar studies (Class and Brommer 2020; Tolvanen et al. 2020). Error variance is also composed of multiple components (it includes, apart from pure residual variance, also ½ of the additive genetic variance). However, interpreting relative contributions of residual and additive genetic variances is difficult as the exact sources of purely residual variance are unidentifiable. Maternal effects may contribute to variation between families but their effect should dissipate with time and hence their influence of traits in older nestlings should be small (Thomson et al. 2017), a pattern that we can clearly see in case of body weights between day 2 and day 8 in our data (see Results and Discussion). The full-siblings assumption is also partly violated by small but significant proportion of extra-pair young detected in the studied population in selected breeding seasons (Arct et al. 2013), resulting in some of the offspring being actually maternal half-siblings. However, resulting bias in genetic variance estimates should be and small and negligible (Firth et al. 2015) as contributions of this error to additive genetic variance and early maternal effects (both included in the nest-of-origin effect) would likely cancel out. We also assumed random distribution of within-pair and extra-pair offspring across treatments and so this issue should not bias genetic parameter estimates systematically.

For all models, we performed three types of sensitivity analyses. To explore the possibility of correcting for prenatal maternal effects by including egg mass in the models, we refitted a subset of data from 2015 to 2016 adding standardised egg mass as a fixed predictor. Secondly, to check the inclusion of nests where significant proportion of chicks failed to hatch could bias the estimates of genetic parameters in any way, we refitted the models only including genetic nests where at least four chicks successful hatched. Finally, we also checked whether reduction in brood size resulting from hatching failure may contribute to the observed patterns—we did this by refitting models with the final brood size as an additional predictor.

When reporting heritabilities and genetic correlations, the reported variance components values are obtained from the best possible model (according to sequential LR tests) but, for illustrative purposes, with the reported component (e.g. genetic variances when reporting heritabilities, and genetic variances and correlations, when reporting genetic correlations) left unconstrained. This way of reporting ensures providing meaningful numbers instead of e.g. three identical values in models where no heterogeneity in genetic variances was detected. Thus, heritabilities may be reported as different (although non-significantly) when in the simplest model the components are constrained to be equal to simplify the model.

Egg hatching success was analysed using a generalised linear mixed model, with egg mass, year and experimental group as fixed predictors, and nest-of-origin and nest-of-incubation as a random predictor. Hatching success was modelled as a binary (0 – failed; 1 – hatched) variable, the models used a logit link function. Since egg mass data is available only for years 2015 and 2016, differences between years and experimental groups were also validated using a reduced model without egg mass (i.e. one that includes the year 2014), but no significant differences in observed patterns were observed.

Mixed models were run in *ASReml-R* v. 4.1 (Butler 2019), all analyses were run in the R computational environment v. 4.0.2 (R Core Team and R Core Team 2014). Before analysis we have removed from the data all individuals where the initial assignments to experimental groups were lost for some reason, e.g. because of premature hatching (and consequent failure to assign chicks to their respective experimental groups).

Hatching success

Despite keeping all procedures as precise and aseptic as possible, our manipulation had a measurable impact on the chicks' hatchability (likely resulting from water loss resulting from incompletely closed eggs, introduction of microorganisms interfering with embryos' development, or bursting of particularly small egg yolks after delivering additional 3 ul of liquid), which is usual in similar studies. While natural hatchability in the studied population reaches 98.0% (mean based on a random sample of 57 non-manipulated nests in 2014), the hatchability in nests manipulated by egg injections dropped to 52.0% on average. Experimental groups did not differ in hatchability (control group: 51.1%, CORT: 49.0%, TESTO: 55.3%; binomial GLMM with control eggs as reference, estimates with SE: $\beta_{CORT} = -0.03 \pm 0.14$, $\beta_{TESTO} = 0.23 \pm 0.14$, $p = 0.11$). Experimental years did differ, with the year 2015 having slightly higher hatchability (binomial GLMM: $p = 0.038$). The median brood size in experimental nests was 5 hatchlings (IQR: 4–6).

When hatchability data are filtered by removing cases of egg dehydration (observation of sticky, thickened or completely dried egg content upon hatch checks), and cases of egg infection (egg content rotten and showing clear signs of bacterial infection) the hatchability levels are slightly higher: 63% overall (control: 60%; CORT: 61%; TESTO: 67%). They still do not differ significantly between treatment groups (binomial GLMM, estimates with SE: $\beta_{CORT} = 0.05 \pm 0.15$, $\beta_{TESTO} = 0.31 \pm 0.15$, $p = 0.09$).

In line with our expectation that smaller eggs would fail to hatch more frequently, we detected a significant positive relationship between egg size and probability of hatching (binomial GLMM, estimates ± SE: $\beta = 2.96 \pm 0.85$, $p < 0.001$ for all hatching failures; $\beta = 2.64 \pm 0.92$, $p = 0.004$ after excluding egg drying and egg infection). There was no significant interaction between experimental treatment and egg mass ($p = 0.72$ for filtered hatching failures, $p = 0.76$ for all unhatched eggs included). Several in ovo studies using species with small eggs with egg size similar to blue tits report similar and lower (down to ~50%) hatchabilities (Winter et al. 2013; Marri and Richner 2014). Small-egged species may be more sensitive to similar manipulations as their yolks are smaller (hence more prone to irreversible damage, or to accidental damage to the germinal disc itself), and their eggs contain smaller amounts of water (making them more prone to dehydration if shell puncture is not sealed completely). It is also possible that, apart from small size of blue tit eggs, transporting the eggs may have also negatively impacted hatching success, compared to studies where eggs were not cross-fostered between nests (majority of studies) and/or transported to incubators before hatching. Finally, within a random subsample of eggs that were collected during hatching checks and opened to determine the embryo stage, a vast majority of hatching failures was due to embryonic development stopping early, during the first stages of growth. We have defined 5 stages (0 = no signs of fertilisation/only germinal disc visible; 1 = visible vascularisation but embryo <2 mm in size, no dark eye pigmentation; 2 = first signs of dark eye pigmentation,

embryo up to 5 mm in size; 3 = embryo with well-developed beak and toes, approx. 8 mm in size, not filling the entire egg; 4 = fully formed nestling with little/no residual albumen); according to this staging, over 55% of failed eggs stopped developing at stages 0 or 1; Fig. S1).

RESULTS
Fixed effects

Means and standard deviations of raw data, together with sample sizes, are provided in Table S3. Body mass and tarsus length were sexually dimorphic (with males being heavier and larger, Table 2). Sex effect on body mass was not observed in 2nd day nestlings, although males still tended to be larger in this age group. Hormonal manipulations exerted no statistically significant effect on most of the measured variables, although day 14 body mass of nestlings tended to be lowest in the CORT-manipulated group (Table 2). However, body mass on the 2nd day was significantly influenced by an interaction between sex and treatment (Table 2). Males tended to be heavier than females in the control group, and in the other hormone-treated groups males had similar weights on the second day. In particular, in testosterone group, a masculinising effect was observed with females being on average heavier than males (although this difference was not statistically significant; Fig. S2). We detected no confounding effect of the person measuring the tarsus length.

Random effects

In all variables we observed statistically significant levels of genetic variance (Table 3), resulting in broad-sense heritabilities consistent with those reported elsewhere in the literature (Merilä and Fry 1998; Hadfield et al. 2007; Drobniak et al. 2015; Perrier et al. 2018). For tarsus length, bootstrapping indicated the presence of clear differences in nest-of-origin variances between he experimental groups. The difference between both C and TESTO group—and the CORT group—was substantial and positive ($V_{G(C)} - V_{G(CORT)} = 0.061$, 95%CI: [0.001; 0.126], $p = 0.026$; $V_{G(TESTO)} - V_{G(CORT)} = 0.050$, 95%CI: [−0.001; 0.110], $p = 0.048$; between control and TESTO groups: $V_{G(C)} - V_{G(TESTO)} = 0.012$, 95%CI: [−0.060; 0.088], $p = 0.632$; Fig. 2). This pattern was confirmed by model likelihoods: the best supported model was the one showing a significant contribution of both nest-of-rearing and nest-of-origin effects, and a significant drop in genetic variance in the CORT-treated group, compared to the other two treatments (TESTO and C; model $\mathbf{Gh_{C \equiv TESTO} + R + E}$; Table 3 and Table S1, Fig. 3A). When estimated separately for each experimental group, the genetic variance was highest in TESTO and C, resulting in highest *sensu lato* heritabilities ($h^2 \pm$ SE: 0.38 ± 0.13 and 0.35 ± 0.12, respectively; Fig. 3A). Heritability in the CORT group was significantly lower (0.12 ± 0.07; Fig. 3B).

In day 14 and day 8 body masses, both the nest-of-rearing and nest-of-origin explained considerable amounts of variation, but there was no sign of any treatment-specific effect on the estimated genetic (nest-of-origin) variances (Fig. 3; preferred models **G** over **Gh**). The effect-sizes of variance differences between treatments were relatively small (the largest effect: for body mass on day 8, $V_{G(C)} - V_{G(CORT)} = -0.130$, 95%CI: [−0.329; 0.080], $p = 0.110$; Fig. S3). Body mass heritabilities had similar magnitudes across treatment groups, both at day 8 (C: 0.48 ± 0.12, CORT: 0.59 ± 0.14, TESTO: 0.64 ± 0.12) and at day 14 (C: 0.35 ± 0.13, CORT: 0.43 ± 0.13, TESTO: 0.37 ± 0.12). Treatment effects on the PHA hypersensitivity response were mostly visible in a marked drop in the control group genetic variance, compared to hormone-treated groups ($V_{G(C)} - V_{G(CORT)} = -0.007$, 95%CI: [−0.015; 0.001], $p = 0.041$; $V_{G(C)} - V_{G(TESTO)} = -0.006$, 95% CI: [−0.014; 0.002], $p = 0.070$; between control and TESTO groups: $V_{G(TESTO)} - V_{G(CORT)} = -0.001$, 95%CI: [−0.012; 0.010], $p = 0.390$; Fig. S3). Accordingly, heritabilities in steroid-treated groups were higher (CORT: 0.45 ± 0.19, TESTO: 0.47 ± 0.19) compared to the markedly smaller and statistically

Table 2. Fixed effects estimates for all response variables

Fixed term	Estimate	SE	Z	p
Tarsus length				
Intercept	16.96	0.53	32.02	<0.001
Measurer ID (2)	−0.14	0.13	−1.12	0.213
Measurer ID (3)	−0.49	0.52	−0.94	0.257
Sex (male)	0.36	0.04	8.34	<0.001
Sex (unsexed)	0.23	0.09	2.57	0.015
Year (2015)	−0.62	0.53	−1.18	0.199
Year (2016)	−0.51	0.54	−0.94	0.256
Treatment (CORT)	0.08	0.05	1.55	0.121
Treatment (TESTO)	0.04	0.05	0.84	0.281
Body mass (day 14)				
Intercept	11.06	0.12	90.77	<0.001
Sex (male)	0.46	0.06	8.16	<0.001
Sex (unsexed)	0.38	0.12	3.22	0.002
Year (2015)	−0.23	0.14	−1.61	0.108
Year (2016)	−0.24	0.14	−1.69	0.095
Treatment (CORT)	−0.08	0.07	−1.27	0.178
Treatment (TESTO)	−0.003	0.06	−0.05	0.398
Body mass (day 8)				
Intercept	8.00	0.16	50.71	<0.001
Sex (male)	0.32	0.07	4.53	<0.001
Sex (unsexed)	0.16	0.14	1.17	0.200
Year (2015)	−0.23	0.19	−1.25	0.182
Year (2016)	−0.03	0.19	−0.18	0.392
Treatment (CORT)	0.03	0.08	0.43	0.364
Treatment (TESTO)	−0.01	0.08	−0.17	0.393
Body mass (day 2)				
Intercept	2.03	0.07	27.39	<0.001
Sex (male)	0.07	0.03	2.01	0.053
Sex (unsexed)	−0.02	0.06	−0.36	0.374
Year (2015)	−0.17	0.09	−1.88	0.068
Year (2016)	−0.18	0.09	−2.01	0.053
Treatment (CORT)	0.04	0.04	1.00	0.243
Treatment (TESTO)	0.00	0.04	−0.12	0.396
Treatment:Sex (TESTO:male)	−0.22	0.08	−2.80	0.011
Treatment:Sex (TESTO: unsexed)	−0.31	0.11	−2.88	0.010
Treatment:Sex (CORT:male)	−0.13	0.08	−1.72	0.101
Treatment:Sex (CORT: unsexed)	−0.10	0.12	−0.87	0.211
PHA response				
Intercept	−0.07	0.09	−0.84	0.280
Body mass (day 12)	0.05	0.01	6.36	<0.001
Sex (male)	0.04	0.01	2.62	0.013
Sex (unsexed)	−0.01	0.03	−0.35	0.375
Year (2016)	0.26	0.02	12.29	<0.001
Treatment (CORT)	−0.01	0.02	−0.65	0.322
Treatment (TESTO)	0.01	0.02	0.60	0.333

The table presents: regression coefficient values, their standard errors and associated Z statistics, and p values

indistinguishable from zero heritability in the control group (C: 0.11 ± 0.11). Nevertheless, these effects should be treated with greater caution (due to reduced sample size in the PHA variable).

Genetic variances in the day 2 body mass were also heterogenous between experimental groups, with the CORT group having markedly larger origin-related variance then C and TESTO groups, a

Table 3. Detailed sets of models considered for each variable, with interpretation of all performed model comparisons

Model 1	ID	log(ℓ)	N_{var}	Model 2	LRT	df	p	Interpretation
Tarsus length								
Gh, Nh, Rh	1	36.74	9	2	0.21	2	0.54	No heterogeneity in V_R
Gh, Nh, R	2	36.64	7	3	0.65	2	0.39	No support for heterogenous V_R
Gh, N, R	3	36.31	5	5	3.07	2	0.09	Weak support for complete heterogeneity in V_G
Gh$_{C \equiv TESTO}$, N, R	**4.1**	**36.27**	**4**	**5**	**3**	**1**	**0.04**	CORT V_G lower than other
Gh$_{TESTO \equiv CORT}$, N, R	4.2	35.14	4	5	0.73	1	0.19	
Gh$_{C \equiv CORT}$, N, R	4.3	34.86	4	5	0.18	1	0.33	
G, N, R	5	34.77	3	6	34.24	1	<0.001	Significant V_G
—	—	—	—	7	91.78	1	<0.001	Significant V_N
N, R	6	17.65	2					
G, R	7	−11.12	2					
Body mass (14)								
Gh, Nh, Rh	1	−123.37	9	2	1.40	2	0.25	No heterogeneity in V_R
Gh, Nh, R	2	−124.07	7	3	1.30	2	0.48	No support for heterogenous V_N
Gh, N, R	3	−124.11	5	5	0.09	2	0.91	No support for complete heterogeneity in V_G
Gh$_{C \equiv TESTO}$, N, R	4.1	−124.20	4	5	0.18	1	0.34	No support for variance differences in simpler models
Gh$_{TESTO \equiv CORT}$, N, R	4.2	−124.16	4	5	0.08	1	0.40	
Gh$_{C \equiv CORT}$, N, R	4.3	−124.20	4	5	0.01	1	0.47	
G, N, R	**5**	**−124.20**	**3**	**6**	**77.64**	**1**	**<0.001**	Significant V_G
—	—	—	—	7	31.02	1	<0.001	Significant V_N
N, R	6	−163.02	2					
G, R	7	−139.71	2					
Body mass (8)								
Gh, Nh, Rh	1	−280.39	9	2	3.90	2	0.08	No heterogeneity in V_R
Gh, Nh, R	2	−282.34	7	3	0.18	2	0.46	No support for heterogenous V_R
Gh, N, E	3	−282.43	5	5	1.63	2	0.44	No support for complete heterogeneity in V_G
Gh$_{C \equiv TESTO}$, N, R	4.1	−283.25	4	5	0.01	1	0.48	No support for variance differences in simpler models
Gh$_{TESTO \equiv CORT}$, N, R	4.2	−282.56	4	5	1.37	1	0.12	
Gh$_{C \equiv CORT}$, N, R	4.3	−282.74	4	5	1.01	1	0.15	
G, N, R	**5**	**−283.25**	**3**	**6**	**91.16**	**1**	**<0.001**	Significant V_G
—	—	—	—	7	97.32	1	<0.001	Significant V_N
N, R	6	−328.83	2					
G, R	7	−331.91	2					
Body mass (2)								
Gh, Nh, Rh	1	206.88	9	2	5.04	2	0.03	Residual variance is heterogenous
—	—	—	—	3	0.41	2	0.46	No support for heterogenous V_R
Gh, Nh, R	2	204.35	7					
Gh, N, Rh	3	206.67	5	5	4.46	2	0.06	Weak support for complete heterogeneity in V_G
Gh$_{C \equiv TESTO}$, N, Rh	**4.1**	**206.25**	**4**	**5**	**3.62**	**1**	**0.03**	CORT V_G higher than other
Gh$_{TESTO \equiv CORT}$, N, Rh	4.2	205.69	4	5	2.49	1	0.06	
Gh$_{C \equiv CORT}$, N, Rh	4.3	204.47	4	5	0.07	1	0.40	
G, N, Rh	5	204.44	3	6	70.14	1	<0.001	Significant V_G
—	—	—	—	7	115.54	1_	<0.001	Significant V_N
N, Rh	6	169.37	2					
G, Rh	**7**	146.67	**2**					
PHA response								
Gh, Nh, Rh	1	611.36	9	2	4.43	2_v	0.05	Heterogenous V_R
—	—	—	—	3	8.49	2_v	0.01	Heterogenous V_N
—	—	—	—	5	2.98	2_v	0.09	Weak support for heterogenous V_G
Gh, Nh, R	2	609.14	7					

Table 3. continued

Model 1	ID	log(ℓ)	N_{var}	Model 2	LRT	df	p	Interpretation
Gh, N, Rh	3	607.12	5					
$Gh_{C \equiv TESTO}$, Nh, Rh	4.1	609.91	4	**5**	3.62	1_v	0.38	
$Gh_{TESTO \equiv CORT}$, Nh, Rh	**4.2**	**611.27**	**4**	**5**	**2.78**	**1_v**	**0.05**	Support for C having lower genetic variance
$Gh_{C \equiv CORT}$, Nh, Rh	4.3	610.50	4	5	1.25	1_v	0.16	
G, Nh, Rh	5	609.87	3	6	8.40	3_v	<0.001	Significant V_G
—	—	—	—	7	13.42	1_v	<0.001	Significant V_N
Nh, Rh	6	605.67	2					
G, Rh	7	603.16	2					

Model naming corresponds to Table 1. *Model IDs are used to specify which model comparisons were performed at each step. Number of variance components in the model – N_{var}. LRT statistics are provided in each for the comparison of a given model against a simpler one (specified by "Model 2"). Difference in the number of variance parameters between compared models – df. Logged model likelihood – log(ℓ). LRT = 2 log(ℓ_Ha/ℓ_Ho) = 2[log(ℓ_Ha) – log(ℓ_Ho)]. V_R, V_G, V_N – residual, genetic and rearing variance, respectively*
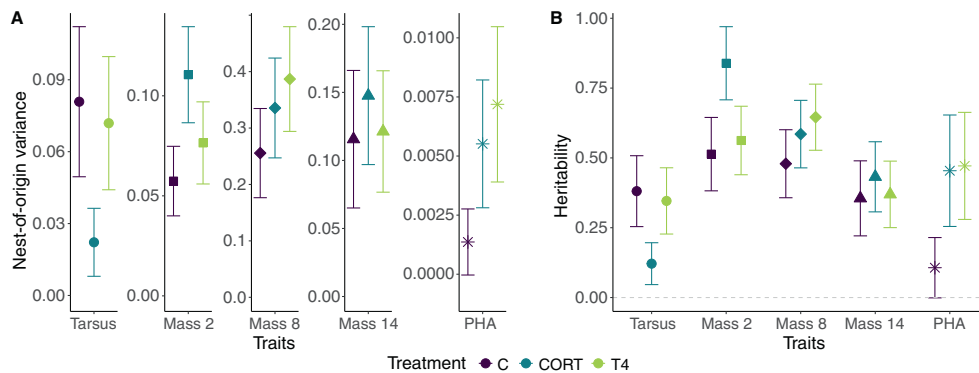
Fig. 2 Heritabilities and genetic variances of the studied traits (symbol-coded) split by the treatment group (colour-coded). Genetic variances (**A**) are presented on trait-specific scales. Heritabilities (**B**) come from models with fully heterogenous genetic variance structures. Whiskers represent SE (estimated via the delta method for heritability or estimated by the model for variances)
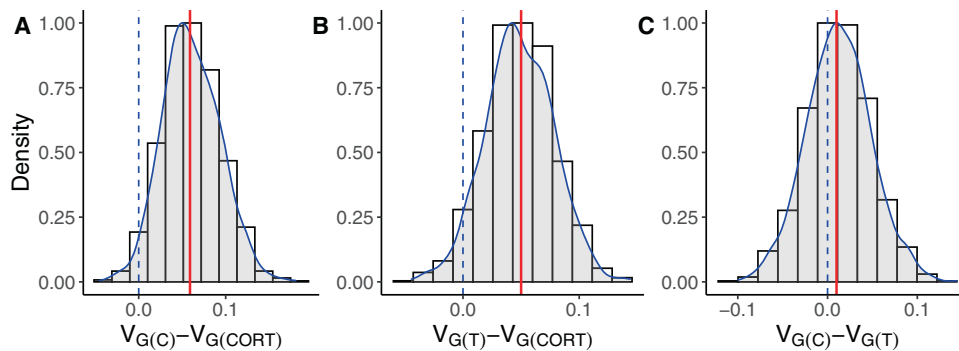
Fig. 3 Results of bootstrapping simulations of the tarsus length variable mixed models. Histograms represent simulated distributions of differences in genetic variances calculated between treatment groups indicated on the x-axes. Overlaid on the histograms are: kernel density estimators (blue solid line), zero lines (blue dotted) and differences in variances from original mixed models (red solid lines). Panels represent differences in genetic variance between control and CORT groups (**A**), TESTO and CORT groups (**B**), and control and TESTO groups (**C**)

pattern clearly visible in the effect sizes of variance differences between groups ($V_{G(C)} - V_{G(CORT)} = -0.048$, 95%CI: [−0.100; 0.005], $p = 0.039$; $V_{G(TESTO)} - V_{G(CORT)} = -0.030$, 95%CI: [−0.083; 0.020], $p = 0.130$; $V_{G(C)} - V_{G(TESTO)} = -0.019$, 95%CI: [−0.064; 0.030], $p = 0.218$; between control and TESTO groups; Fig. S3). This translated into differences in treatment-specific heritabilities (C: 0.51 ± 0.13; CORT: 0.83 ± 0.13; TESTO: 0.57 ± 0.12). In addition, two traits (body mass on day 2 and PHA response) exhibited significant, albeit relatively small, heterogeneity in residual variances (in both cases accounted for in the models used for bootstrapping; Table S1). Detailed estimates of all variance components from models selected as the best supported are presented in Table S1.

Body weight of nestlings on the day of hatching (day 0) was analysed separately and it did not show any differences due to experimental hormonal treatment (see Table S3). The nest-of-origin and nest-of-rearing effects showed no hormone-specific structuring. As expected, the nest-of-origin effect explained the majority of variance in this trait (55.7%); nest-of-rearing explained only 7.4% (which is expected as the only rearing component at this stage may result from incubation-related factors).

To account for some maternal effects that may correlate with natural maternally transmitted yolk hormones, we refitted all the above models with an additional fixed predictor of egg mass, and its interaction with experimental treatment. Explanatory power of

these models was lower as egg masses were available only for 2 years (2015 and 2016). Nevertheless, inclusion of egg mass neither did affect the sequence of preferred models, nor change final conclusions about genetic variance patterns in tarsus length, PHA response and day 2 and 8 body masses. Models for body mass on day 14 showed lower estimates of nest-of-origin variance (Table S5) in the CORT group, but this effect was due to restricting the analysis to two out of three years. Nevertheless, when comparing compatible models (i.e. one including egg mass and one excluding it, both based on years 2015–2016) the inclusion of egg mass did lead to decrease in nest-of-origin effects that was (in terms of its magnitude) the largest among all random terms. Egg mass effect on tarsus length was substantial (standardised effect with SE: $\beta = 0.14 \pm 0.05$, $p < 0.001$), similarly for mass on day 14 ($\beta = 0.20 \pm 0.06$, $p < 0.001$), mass on day 8 ($\beta = 0.20 \pm 0.08$, $p < 0.001$) and mass on day 2 ($\beta = 0.07 \pm 0.03$, $p = 0.054$), but it did not vary between experimental groups in any model (see also Supplementary Table S5). The effect of egg mass on PHA response was negligible ($\beta = 0.008 \pm 0.013$, $p = 0.55$).

A sensitivity analysis performed to check the robustness of our analysis to the presence of genetic nests with overall low survival (performed by removing from the data all nests-of-origin with <4 chicks surviving to trait measurement, the subset retained 107 out of 159 genetic nests) returned the same sequences of preferred models and qualitatively identical results regarding genetic variances. Since chick hatchability and mortality affected final brood sizes in experimental triads generating some differences in brood sizes, we have also checked if variation in intra-brood competition could contribute to the observed differences in variance components. In models including final brood size as a covariate, we observed no changes in final conclusions. Identical conclusion was true when the brood size of each nestling's genetic nest was included. In line with this, the brood sizes continuous predictors were in both cases statistically non-significant in all analysed traits.

DISCUSSION

Differences in heritabilities measured in different biological contexts are not uncommon in natural populations. Apart from population specificity of heritabilities (which contributes to marked variation in heritabilities even within one species (Lynch and Walsh 1998)), a great deal of attention has been paid to changes in trait genetic variances observed under varying biological conditions. Genetic variance in a trait is one of the most fundamental ingredients of phenotypic evolution (Lande and Shannon 1996; Walsh and Lynch 2018). As such, changes to genetic variance "expressed"—or visible to natural selection—induced, e.g. by individual characteristics or environmental variability should play an important role in modulating the course of evolutionary change and conserving the levels of genetic variance (Lynch and Walsh 1998). This issue is especially interesting as several physiological systems (e.g. heat-shock proteins) can act as "evolutionary capacitors" releasing, under certain conditions, cryptic genetic variance (Gibson and Dworkin 2004).

Here, we have observed that in ovo maternal hormones constitute one of the cues that may modulate observed levels of genetic variance (and consequently heritability). By altering the hormonal environment of developing embryos, in conjunction with a cross-fostering experiment that allows for separation of phenotypic variance contributors, we were able to show that steroid hormones acting early in individual development can alter the observed levels of genetic variance (see also the Supplement for additional results and discussion on cross-trait genetic correlations). Of all studied traits, we have observed an effect of hormonal manipulation on genetic variance in the tarsus length, a trait typically exhibiting moderate to high levels of heritability in wild populations. The heritability of tarsus length in our study

(taking the control group as reference) agreed with other published estimates, including in the blue tit (Bonneaud et al. 2009; Teplitsky et al. 2009; Nilsson et al. 2009; Delahaie et al. 2017; Perrier et al. 2018), and with a more general set of published estimates of body size heritabilities from wild populations (Postma 2014) and yet, in the corticosterone-treated nestlings, heritability dropped to statistically indistinguishable from zero. These genetic variance differences (very low variance under one set of conditions, typical levels under two others) suggest an existence of hormone-mediated G × E interaction. An opposite pattern was visible in the body weight on the 2nd day after hatching; here, individuals in the corticosterone-treated group had markedly higher heritabilities than the control and testosterone groups. The effects we observed in tarsus length were robust to including egg mass in the model. It indicates that the observed effect is robust to maternal effects correlated with egg size. However, similar analysis performed on day 2 body mass (i.e. trait that should reflect early maternal effects to a greater extent compared to day 14 body mass or tarsus length, data not presented) showed no change to genetic variance after accounting for egg mass. Therefore, whether this correction reflected pre-treatment variation in some maternal effects magnitude is speculative and requires more studies looking at the relationships between egg size and deposition of maternally derived compounds into eggs.

One reason explaining the low heritability of tarsus length (and the matching trend on day 14 body weight) in the corticosterone group may be the physiological function of this steroid. This hormone is usually considered to mediate stress response and an organism's mobilisation after experiencing stress (Schoech et al. 2011). In birds corticosterone has been shown to mimic stress induced body weight variations and supressed immune response (Roberts et al. 2007), trigger development of stress-like phenotypes (Roulin et al. 2008), amplify behavioural differences along the shy-bold personality axis (Baugh et al. 2012), and impair parental care (Angelier et al. 2009) or learning behaviour (Kitaysky et al. 2003). If corticosterone exposure is regarded as mimicking stress exposure, its effects on genetic variance in offspring traits may mimic those observed in genotype-by-environment interactions, when individuals are exposed to stressful or unfavourable conditions. Although the impact of such conditions may differ from species to species (Rowiński and Rogell 2017), the often observed pattern is reduction in observed levels of genetic variance (Hoffmann and Merilä 1999), supported by a meta-analysis of such results (Charmantier and Garant 2005). Hoffmann and Merilä (1999) argued that one possible mechanism of such reduction is stopping of offspring growth, caused by stress, before inter-individual variance in achieved body size can fully develop. In our study this explanation doesn't seem to be valid: we haven't observed any systematic differences in body size between the three treatment groups. Other mechanism that can be invoked in cases of condition-varying genetic variances involves changes in gene expression at a molecular level (Hodgins-Davis and Townsend 2009). Hormonal influence on gene expression patterns could lead to modifications of breeding values underlying phenotypes, and eventually to changes in the levels of genetic variance. Some studies suggests that in ovo corticosterone can induce gene expression changes in birds (Ahmed et al. 2016), but such evidence is not unambiguous (Lutyk et al. 2017). Function of corticosteroid receptors as transcription factors directly modulating expression of certain genes is known (Kawata 1995; Baker 1997). Nonetheless, more work is needed – especially at the very basic, molecular level and in early developing embryos.

It is also interesting that the observed effect is to certain extent complementary to the action of so called "evolutionary capacitors"—actors that can release cryptic genetic variation under certain conditions (Queitsch et al. 2002; Gibson and Dworkin 2004). In our setup corticosterone seems to reduce genetic

variance—which, under stressful conditions, may allow for the maintenance of certain genetic variants beyond the nestling period. Context-dependent changes in the expressed genetic variance can be expected as one of the properties of evolving, adapting systems (Masel 2006) —and hormones with relatively short time windows of activity may constitute an important component of the general system regulating genetic variation "visible" to natural selection.

In contrast to corticosterone manipulation, testosterone-exposed nestlings showed much subtler and less convincing changes in genetic variances of their traits. Physiologically, testosterone is traditionally associated with sex-specific effects and is assumed to mediate trade-offs between body maintenance and resource use in production of secondary sexual characters (Peters 2007; Kingma et al. 2009). Exposure to testosterone early in development was shown to stimulate development of sexually selected traits, bias sex-ratios towards males and increase dominance and competitive behaviours (Podmokła et al. 2018). Unfortunately, due to sample size we couldn't robustly test for sex-by-treatment interaction in genetic parameters. Nevertheless, in line with reported cases of sex-specific genetic variances in a number of traits (Wyman and Rowe 2014) we expected to see significant impact of hormonal manipulation in our study. One reason for not seeing such effect may be the use of testosterone alone, in contrast to many similar studies using both testosterone and androstenedione as two major sex-linked hormones (Podmokła et al. 2018). Our observation is similar to other studies where no significant in ovo testosterone impact was noted (Tschirren et al. 2005; Podmokła et al. 2018) and suggests that future studies should look more closely on complexes of similarly acting hormones (or even functionally different compounds applied together—e.g. Giraudeau et al. (2017)), applying them together to better mimic biological reality.

Of all other studied traits, only the PHA hypersensitivity response showed some heterogeneity in genetic variances. This was also the only trait where some heterogeneity was visible in the variation of nest-of rearing effects. As for the PHA response, the heterogeneity of genetic variances was visible in the control group having four-fold lower heritability than hormone-treated groups. Steroid hormones are known modulators of immune response and the observed effect may reflect this link, with evidence of both immunosuppressive and stimulating influence (Casto et al. 2001; Andersson et al. 2004; Rubolini et al. 2005; Navara et al. 2006; Roberts et al. 2007). However, this trait was analysed using a limited dataset (2 out of 3 years) and so this pattern should be treated with caution.

Lack of nest-of-rearing vs. treatment interaction in other traits emphasises that the effects of post-hatching parental provisioning and variation associated with the foster nest (including habitat variation) do not depend on the amounts of steroid hormones transferred to eggs. Even in the remaining measured traits our design may still be not powerful enough to detect such interactive effects. It is of course possible that nest-of-rearing variation in other traits not considered in our study, or nest-related variation in traits measured well after leaving the nest, could depend on maternal effects. Natal environment and rearing effects can be detected in phenotypic traits long after fledging (Evans and Sheldon 2012), but gradual disappearance of nest-of-origin (i.e. including maternal) effects even within the nesting period (Thomson et al. 2017) would make observation of such rearing vs. maternal hormones interactions difficult.

In terms of trait means, in most cases we observed no impact of hormonal treatment. The only exception was body mass on the 2nd day: here, female offspring tended to be influenced by both testosterone and corticosterone, with their masses drawn closer to the values of males. Masculinising effect of testosterone on morphology in not uncommon (Podmokła et al. 2018), evidence (albeit scarcer) also exists for such effects exerted by

corticosterone (Roberts et al. 1997; Mankiewicz et al. 2013). By the time of fledging, all phenotypic effects of hormones on trait averages seem to dissipate, which also in line with generally weak evidence for strong mean effects of in ovo hormonal manipulations (Podmokła et al. 2018). All observed changes in patterns of genetic variances are therefore pure $G \times E$s, without a measurable component of phenotypic plasticity (as trait means do not vary between treatments). In principle, $G \times E$ does not require the concurrent change in trait means (it is defined as variation in reaction norms of individual genotypes, which on average can cancel out rendering net zero change in means; Anholt and Mackay 2004; Saltz et al. 2018; Wang et al. 2019; Huang et al. 2020); still—by modifying the levels of genetic variance—such genotype-by-environment interactions can affect evolutionary trajectories of traits (Lynch and Walsh 1998).

On the methodological side, our study may suffer from several shortcomings. Our experiment was associated with considerable reduction in eggs hatching success. Smaller embryos generally hatch less successfully (Krist 2011), which was also the case in our data and could negatively select for hatchlings' size and eventually reduce body size variation. However, we detected no differences in hatchability between treatment groups, i.e. lowered hatchability should not contribute significantly to the observed differences in the broad-sense genetic variation. Most hatching failures occurred at very early stages of embryonic development, suggesting random causes such as permanent disruption of the yolk or damage to the germinal disc. This would also explain increased hatching failure of smaller eggs: smaller yolks would be more prone to rupture upon solution injection or damaging of the delicate germinal disc. It should also be noted that any treatment-specific mortality of embryos could also be attributed to a potential $G \times E$—if mortality would occur in a treatment- and genotype-specific way. Our study used a substantial increase in the in ovo hormone concentrations (up to 2 SDs over natural levels). It is unlikely that such concentrations could be toxic to embryos (partly because in such way toxicity would start at levels dangerously close to those seen in nature, and partly because embryos are expected to have some compensation mechanisms potentially shielding them from such detrimental effects of hormones (Groothuis and Schwabl 2008)).

One of the consequences of lower hatchability may be lowered statistical power, resulting from smaller numbers of individuals per nest-of-origin entering the analyses. After Klein (1974), we can very roughly estimate power to detect heritability differences similar to ours, with our numbers of genetic families and offspring per family, as varying between 47 and 54% (depending on the assumed numbers of offspring contributing to estimates). Klein's method assumes simple comparison of heritabilities—multi-level mixed model and likelihood-ratio tests should provide higher power still. Lowered power would explain why in some cases the observed differences border the nominal significance threshold. It would especially affect the more challenging estimation of genetic covariances—hence we decided to fix the cross-treatment correlations at their respective values, and do not focus on the estimates of cross-trait genetic correlations (see Supplementary Materials for brief outline of those estimates). Nonetheless, focusing on effect-sizes rather than arbitrarily set statistical significance can provide valid insights and reduce biases in published estimates of variance components. Future studies should focus on building more powerful designs to provide more robust tests to genetic covariance parameters in similar experiments.

We used cross-fostering to separate sources of phenotypic variation in measured traits. The estimated variation between full-sib families (nest-of-origin effect) should therefore contain ½ of a trait's additive genetic variance. However, if present, this effect may also contain pre-hatching maternal effects and ¼ of genetic dominance variance (the latter can be assumed to be negligible; Class and Brommer 2020). In birds, females seem to be repeatable
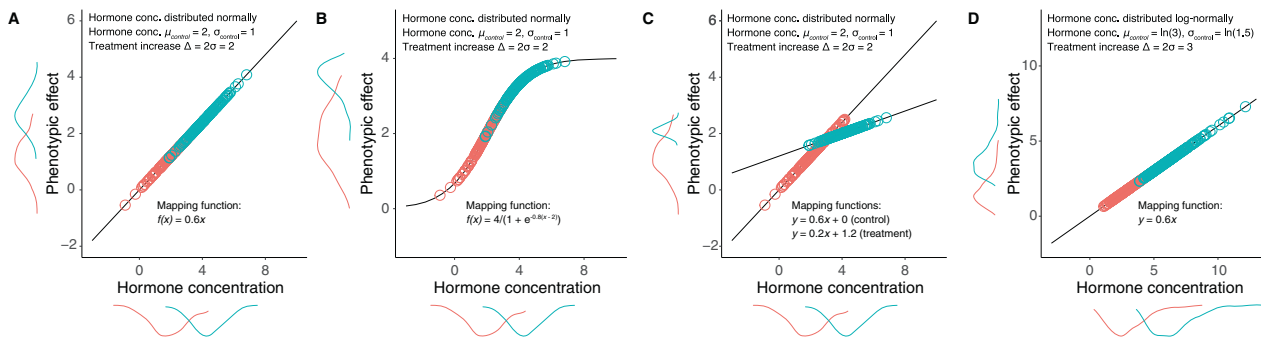
Fig. 4 Possible scenarios of hormonal treatment affecting the variance in maternal effects, depending on distribution of hormone concentrations and the function mapping hormone levels to phenotype values. Symmetrical distributions – with linear (**A**), sigmoidal (saturating; **B**) or heterogenous linear (concentration-dependent) mapping functions (**C**), and a linear mapping function with skewed (log-normal) distributions (**D**)

in the amount of hormones transferred to eggs (Tschirren et al. 2009; Ruuskanen et al. 2016)—therefore, if present, initial between-mothers variation in yolk hormones would be confounded with additive genetic variance in traits. Studies looking at the ontogenetic changes in nest-of-origin effects show that their impact decreases over the nesting period, with a parallel increase of the relative contribution of genetic effects (Pick et al. 2016; Thomson et al. 2017). Thus, our estimates for day 14 body mass and tarsus length should robustly reflect additive genetic variance contributions. Indeed, our heritability estimates in those two traits agree with other published values (Postma 2014). Early (day 0, 2 and 8) body masses, on the other hand, will likely be influenced to a varying degree by significant maternal effects—which is also reflected in the gradual drop of estimated broad-sense heritabilities seen in our study (Fig. 3).

The sensitivity analysis (inclusion of egg masses in the analysed models, based on the 2015–2016 subset of data) does to a certain extent explain the contribution of some maternal effects—but it cannot be seen as a completely robust solution. Therefore, the variation in nest-of-origin variances of the 2nd day body masses could be due to an interaction of hormonal treatment and genetic effects, but it can also be caused by an interaction between treatment and early maternal effects (especially taking unusually high estimates of day 2 body mass heritabilities). Interestingly, the inclusion of egg mass does seem to explain at least part of the nest-of-origin effects (Table S5). Its inclusion leads to changes in genetic (nest-of-origin) variances that are proportionally larger than changes in the nest-of-rearing and residual variances. Magnitude of these changes also is greater for a more labile and condition-sensitive trait (body mass, compared to tarsus length). Direct tests of the impact of egg mass on early nest-of-origin effects are not common. Some show measurable impact both on variance components and nestling growth patterns (Hadfield et al. 2013a), other suggest little to no influence (Hadfield et al. 2013b). More studies are needed, especially that the impact of egg characteristics may escape simple descriptors such as hormone concentrations (Valcu et al. 2019).

Our interpretation of a drop in genetic variance in corticosterone-treated tarsus length seems to also contradict the lack of concurrent drop in the trait's residual variance in this group (expected as within-family residual variance should contain a contribution of ½ of additive genetic variance). However, residual variances often interact with environmental factors on their own (Charmantier and Garant 2005). Within family residual variance is also expected to contain substantial amount of specific environmental variance, i.e. environment-generated variance arising at an individual level (e.g. due to developmental instabilities caused by environmental changes; Lynch and Walsh 1998). Thus, patterns of environment-specific changes in residual variance (or lack thereof) should be treated

in caution as usually we lack information to attribute them causally to specific factors.

In theory, a physiological phenomenon could be responsible for the observed differences in nest-of-origin variances if indeed they were driven by maternal and not genetic effects. Even if variation in maternal hormone concentrations would be unaffected between control and hormone-treated groups (Fig. 4, horizontal axes; the distribution is shifted by an experimental increase in hormone concentration, but its variance remains constant), hormone-linked phenotypes could exhibit treatment-specific reductions in variance if their values would depend on hormone levels in a non-linear way. This could occur, e.g. due to hormones having a saturating effect on the affected phenotype (Fig. 4B, occurring when manipulation would shift hormone concentrations to the plateau region of the curve), or due to the hormone-phenotype mapping having different functional form (Fig. 4C, e.g. if compensation mechanism would decrease trait sensitivity to the hormone under its elevated concentrations). The only scenario when a change in trait variance does not occur (vertical axes on Fig. 4) is a linear mapping between hormones and affected phenotypes (Fig. 4A, D; the same pattern is seen assuming a normal (a) and lognormal (d; Lessells et al. 2016) hormone concentration). However, all scenarios assuming some change in variance would also require a shift in trait mean, a necessary consequence of a shift in hormone concentration along the mapping curve. Since in our case trait means remained unchanged, variance changes observed in our study should reflect genuine effects of reduction in genetic variances. Nonetheless, better understanding of patterns demonstrated in this study is needed, e.g. using laboratory setups and planned breeding designs to clearly separate additive genetic variance from other variance components.

In summary, our study provides the first experimental attempt at identifying mechanisms that may be responsible for the modulation of the expressed genetic variance in nature. Clearly, steroid hormones do offer an interesting study system and are capable of producing patterns mimicking those seen in G × E in wild systems. Complementary studies in wild populations, as well as more in-depth analyses looking at the molecular pathways involved, are needed to better understand and explain the mechanism of how steroid hormones may mediate the observed effects. Subsequent studies should also attempt to increase the statistical power of similar analyses, e.g. by using model systems with larger eggs, better suited for such studies through increased resilience to in ovo manipulations. Taking the widespread occurrence of environmental effects on the levels of genetic variance in nature, and the commonly accepted role of steroid hormones in mediating environment-induced influences during development, we believe that our study provides a refreshing and novel perspective on the issue.

DATA AVAILABILITY

REFERENCES

Ahmed AA, Sifaldin AZ, Musa HH, Musa TH, Fedail JS (2016) Prenatal corticosterone altered glucocorticoid receptor and glucocorticoid metabolic enzyme gene expression in chicken ovary. Asian Pac J Reprod 5:227–231

Andersson S, Uller T, Lõhmus M, Sundström F (2004) Effects of egg yolk testosterone on growth and immunity in a precocial bird. J Evol Biol 17:501–505

Angelier F, Clément-Chastel C, Welcker J, Gabrielsen GW, Chastel O (2009) How does corticosterone affect parental behaviour and reproductive success? A study of prolactin in black-legged kittiwakes. Funct Ecol 23:784–793

Anholt RRH, Mackay TFC (2004) Quantitative genetic analyses of complex behaviours in Drosophila. Nat Rev Genet 5:838–849

Arct A, Drobniak SM, Podmokła E, Gustafson L, Cichoń M (2013) Benefits of extra-pair mating may depend on environmental conditions - an experimental study in the blue tit (Cyanistes caeruleus). Behav Ecol Sociobiol 67:1809–1815

Baker ME (1997) Steroid receptor phylogeny and vertebrate origins. Mol Cell Endocrinol 135:101–107

Baugh AT, Schaper SV, Hau M, Cockrem JF, de Goede P, Oers Kvan (2012) Corticosterone responses differ between lines of great tits (Parus major) selected for divergent personalities. Gen Comp Endocrinol 175:488–494

Berger D (2011) High temperatures reveal cryptic genetic variation in a polymorphic female sperm storage organ. Evolution 65:2830–2842

Bergman A, Siegal ML (2003) Evolutionary capacitance as a general feature of complex gene networks. Nature 424:549–552

Bonneaud C, Sinsheimer JS, Richard M, Chastel O, Sorci G (2009) Mhc polymorphisms fail to explain the heritability of phytohaemagglutinin- induced skin swelling in a wild passerine. Biol Lett 5:784–787

Brommer JE, Rattiste K, Wilson AJ (2008) Exploring plasticity in the wild: laying date-temperature reaction norms in the common gull Larus canus. Proc R Soc B Biol Sci 275:687–693

Butler D (2019) asreml: Fits the Linear Mixed Model. R package version 4.1.0.110. https://www.vsni.co.uk

Casto JM, Nolan Jr V, Ketterson ED (2001) Steroid Hormones and Immune Function: Experimental Studies in Wild and Captive Dark-Eyed Juncos (Junco hyemalis). Am Nat 157:408–420

Charmantier A, Brommer JE, Nussey DH (2014) The quantitative genetics of senescence in wild animals. In: Charmantier A, Garant D, Kruuk LEB (eds) Quantitative Genetics in the Wild. Oxford University Press, Oxford, p 68–83

Charmantier A, Garant D (2005) Environmental quality and evolutionary potential: Lessons from wild populations. Proc R Soc B Biol Sci 272:1415–1425

Class B, Brommer JE (2020) Can dominance genetic variance be ignored in evolutionary quantitative genetic analyses of wild populations? Evolution 74:1540–1550

Coslovsky M, Groothuis T, de Vries B, Richner H (2012) Maternal steroids in egg yolk as a pathway to translate predation risk to offspring: Experiments with great tits. Gen Comp Endocrinol 176:211–214

Delahaie B, Charmantier A, Chantepie S, Garant D, Porlier M, Teplitsky C (2017) Conserved G-matrices of morphological and life-history traits among continental and island blue tit populations. Heredity 119:76–87

Demongin L (2016) Identification guide to birds in the hand. Laurent Demongin, Beauregard-Vendon

Drobniak SM, Cichoń M (2016). No evidence for the positive relationship between genetic correlations and heritabilities. bioRxiv: 039388.

Drobniak SM, Dubiec A, Gustafsson L, Cichoń M (2015) Maternal Age-Related Depletion of Offspring Genetic Variance in Immune Response to Phytohaemagglutinin in the Blue Tit (Cyanistes caeruleus). Evol Biol 42:88–94

Evans SR, Sheldon BC (2012) Quantitative genetics of a carotenoid-based color: heritability and persistent natal environmental effects in the great tit. Am Nat 179:79–94

Firth JA, Hadfield JD, Santure AW, Slate J, Sheldon BC (2015) The influence of non-random extra-pair paternity on heritability estimates derived from wild pedigrees. Evolution 69:1336–1344

Fisher RA (1930) The genetical theory of natural selection. Clarendon Press, Oxford

Foerster K, Coulson T, Sheldon BC, Pemberton JM, Clutton-Brock TH, Kruuk LEB (2007) Sexually antagonistic genetic variation for fitness in red deer. Nature 447:1107–1110

Fox CW, Wolf JB (2006) Evolutionary genetics. Concepts and case studies. Oxford University Press, New York, NY

Galloway LF, Etterson JR, McGlothlin JW (2009) Contribution of direct and maternal genetic effects to life-history evolution. N Phytol 183:826–838

Gam AE, Mendonça MT, Navara KJ (2011) Acute corticosterone treatment prior to ovulation biases offspring sex ratios towards males in zebra finches Taeniopygia guttata. J Avian Biol 42:253–258

Garant D, Kruuk LEB, Wilkin TA, McCleery RH, Sheldon BC (2005) Evolution driven by differential dispersal within a wild bird population. Nature 433:60–65

Gebhardt-Henrich SG, Noordwijk AJV (1991) Nestling growth in the Great Tit I. Heritability estimates under different environmental conditions. J Evol Biol 4:341–362

Gibson G, Dworkin I (2004) Uncovering cryptic genetic variation. Nat Rev Genet 5:681–690

Gienapp P, Brommer JE (2014) Evolutionary dynamics in response to climate change. In: Charmantier A, Garant D, Kruuk LEB (eds) Quantitative Genetics in the Wild. Oxford University Press, Oxford, p 254–274

Giraudeau M, Ziegler A-K, Pick JL, Ducatez S, Canale CI, Tschirren B (2017) Interactive effects of yolk testosterone and carotenoid on prenatal growth and offspring physiology in a precocial bird. Behav Ecol 28:31–38

Griffiths R, Double MC, Orr K, Dawson RJG (1998) A DNA test to sex most birds. Mol Ecol 7:1071–1075

Groothuis TGG, Müller W, Von Engelhardt N, Carere C, Eising C (2005) Maternal hormones as a tool to adjust offspring phenotype in avian species. Neurosci Biobehav Rev 29:329–352

Groothuis TGG, Schwabl H (2008) Hormone-mediated maternal effects in birds: Mechanisms matter but what do we know of them? Philos Trans R Soc B Biol Sci 363:1647–1661

Gunay F, Alten B, Ozsoy ED (2011) Narrow-sense heritability of body size and its response to different developmental temperatures in Culex quinquefasciatus (Say 1923). J Vector Ecol 36:348–354

Hadfield JD, Heap EA, Bayer F, Mittell EA, Crouch NMA (2013a) Disentangling Genetic and Prenatal Sources of Familial Resemblance Across Ontogeny in a Wild Passerine. Evolution 67:2701–2713

Hadfield JD, Heap EA, Bayer F, Mittell EA, Crouch NMA (2013b) Intraclutch Differences in Egg Characteristics Mitigate the Consequences of Age-Related Hierarchies in a Wild Passerine. Evolution 67:2688–2700

Hadfield JD, Nutall A, Osorio D, Owens IPF (2007) Testing the phenotypic gambit: Phenotypic, genetic and environmental correlations of colour. J Evol Biol 20:549–557

Hayward LS, Wingfield JC (2004) Maternal corticosterone is transferred to avian yolk and may alter offspring growth and adult phenotype. Gen Comp Endocrinol 135:365–371

Hegyi G, Herényi M, Szöllősi E, Rosivall B, Török J, Groothuis TGG (2011) Yolk androstenedione, but not testosterone, predicts offspring fate and reflects parental quality. Behav Ecol 22:29–38

Hodgins-Davis A, Townsend JP (2009) Evolving gene expression: from G to E to G × E. Trends Ecol Evol 24:649–658

Hoffmann AA, Merilä J (1999) Favourable and Unfavourable Conditions. Trends Ecol Evol 14:96–101

Huang W, Carbone MA, Lyman RF, Anholt RRH, Mackay TFC (2020) Genotype by environment interaction for gene expression in Drosophila melanogaster. Nat Commun 11:5451

Jensen H, Saether B-E, Ringsby TH, Tufto J, Griffith SC, Ellegren H (2003) Sexual variation in heritability and genetic correlations of morphological traits in house sparrow (Passer domesticus). J Evol Biol 16:1296–1307

Kawata M (1995) Roles of steroid hormones and their receptors in structural organization in the nervous system. Neurosci Res 24:1–46

Kim SY, Drummond H, Torres R, Velando A (2011) Evolvability of an avian life history trait declines with father's age. J Evol Biol 24:295–302

Kingma SA, Komdeur J, Vedder O, Von Engelhardt N, Korsten P, Groothuis TGG (2009) Manipulation of male attractiveness induces rapid changes in avian maternal yolk androgen deposition. Behav Ecol 20:172–179

Kitaysky AS, Kitaiskaia EV, Piatt JF, Wingfield JC (2003) Benefits and costs of increased levels of corticosterone in seabird chicks. Horm Behav 43:140–149

Klein TW (1974) Heritability and genetic correlation: statistical power, population comparisons, and sample size. Behav Genet 4:171–189

Krist M (2011) Egg size and offspring quality: a meta-analysis in birds. Biol Rev 86:692–716

Kruuk LEB, Hadfield JD (2007) How to separate genetic and environmental causes of similarity between relatives. J Evol Biol 20:1890–1903

Lai Y-T, Yeung CKL, Omland KE, Pang E-L, Hao Y, Liao B-Y et al. (2019) Standing genetic variation as the predominant source for adaptation of a songbird. Proc Natl Acad Sci 116:2152–2157

Lande R, Arnold SJ (1983) The Measurement of Selection on Correlated Characters. Evolution 37:1–18

Lande R, Shannon S (1996) The role of genetic variation in adaptation and population persistence in a changing environment. Evolution 50:434–437

Lessells CM, Ruuskanen S, Schwabl H (2016) Yolk steroids in great tit Parus major eggs: variation and covariation between hormones and with environmental and parental factors. Behav Ecol Sociobiol 70:843–856

Lutyk D, Tagirov M, Drobniak S, Rutkowska J (2017) Higher growth rate and gene expression in male zebra finch embryos are independent of manipulation of maternal steroids in the eggs. Gen Comp Endocrinol 254:1–7

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer, Sunderland, MA

Mankiewicz JL, Godwin J, Holler BL, Turner PM, Murashige R, Shamey R et al. (2013) Masculinizing effect of background color and cortisol in a flatfish with environmental sex-determination. Integr Comp Biol 53:755–765

Marri V, Richner H (2014) Yolk carotenoids increase fledging success in great tit nestlings. Oecologia 176:371–377

Masel J (2006) Cryptic Genetic Variation Is Enriched for Potential Adaptations. Genetics 172:1985–1991

Merilä J (1997) Expression of Genetic Variation in Body Size of the Collared Flycatcher Under Different Environmental Conditions. Evolution 51:526–536

Merilä J, Fry JD (1998) Genetic variation and causes of genotype-environment interaction in the body size of blue tit (Parus caeruleus). Genetics 148:1233–1244

Navara KJ, Hill GE, Mendonça MT (2006) Yolk Testosterone Stimulates Growth and Immunity in House Finch Chicks. Physiol Biochem Zool 79:550–555

Nilsson JÅ, Åkesson M, Nilsson JF (2009) Heritability of resting metabolic rate in a wild population of blue tits. J Evol Biol 22:1867–1874

Oltman DWC, Onoghue POD, Ogg JTH, Ianchet MF (2005) Selection and Genetic (Co) Variance in Bighorn Sheep. Evolution 59:1372–1382

Perrier C, Delahaie B, Charmantier A (2018) Heritability estimates from genome-wide relatedness matrices in wild populations: application to a passerine, using a small sample size. Mol Ecol Resour 18:838–853

Peters A (2007) Testosterone and carotenoids: an integrated view of trade-offs between immunity and sexual signalling. BioEssays 29:427–430

Pick JL, Ebneter C, Hutter P, Tschirren B (2016) Disentangling Genetic and Prenatal Maternal Effects on Offspring Size and Survival. Am Nat 188:628–639

Pitala N, Siitari H, Gustafsson L, Brommer JE (2009) Ectoparasites help to maintain variation in cell-mediated immunity in the blue tit-hen flea system. Evol Ecol Res 11:79–94

Podmokła E, Drobniak SM, Rutkowska J (2018) Chicken or egg? Outcomes of experimental manipulations of maternally transmitted hormones depend on administration method – a meta-analysis. Biol Rev 93:1499–1517

Poissant J, Wilson AJ, Coltman DW (2010) Sex-specific genetic variance and the evolution of sexual dimorphism: a systematic review of cross-sex genetic correlations. Evolution 64:97–107

Postma E (2014) Four decades of estimating heritabilities in wild vertebrate populations: improved methods, more data, better estimates? In: Charmantier Garant D, Kruuk LEBA (ed) Quantitative Genetics in the Wild. Oxford University Press, Oxford, United Kingdom, p 16–33

Price GP (1970) Selection and Covariance. Nature 227:520–521

Queitsch C, Sangster TA, Lindquist S (2002) Hsp90 as a capacitor of phenotypic variation. Nature 417:618–624

R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Remeš V (2011) Yolk androgens in great tit eggs are related to male attractiveness, breeding density and territory quality. Behav Ecol Sociobiol 65:1257–1266

Roberts ML, Buchanan KL, Hasselquist D, Evans MR (2007) Effects of testosterone and corticosterone on immunocompetence in the zebra finch. Horm Behav 51:126–134

Roberts RL, Zullo AS, Carter CS (1997) Sexual differentiation in prairie voles: the effects of corticosterone and testosterone. Physiol Behav 62:1379–1383

Roulin A, Almasi B, Rossi-Pedruzzi A, Ducrest AL, Wakamatsu K, Miksik I et al. (2008) Corticosterone mediates the condition-dependent component of melanin-based coloration. Anim Behav 75:1351–1358

Rowiński PK, Rogell B (2017) Environmental stress correlates with increases in both genetic and residual variances: a meta-analysis of animal studies. Evolution 71:1339–1351

Rubolini D, Romano M, Boncoraglio G, Ferrari RP, Martinelli R, Galeotti P et al. (2005) Effects of elevated egg corticosterone levels on behavior, growth, and immunity of yellow-legged gull (Larus michahellis) chicks. Horm Behav 47:592–605

Ruuskanen S, Doligez B, Pitala N, Gustafsson L, Laaksonen T (2012) Long-term fitness consequences of high yolk androgen levels: Sons pay the costs. Funct Ecol 26:884–894

Ruuskanen S, Gienapp P, Groothuis TGG, Schaper SV, Darras VM, Pereira C et al. (2016) Heritable variation in maternally derived yolk androgens, thyroid hormones and immune factors. Heredity 117:184–190

Saltz JB, Bell AM, Flint J, Gomulkiewicz R, Hughes KA, Keagy J (2018) Why does the magnitude of genotype-by-environment interaction vary? Ecol Evol 8:6342–6353

Sarv T, Horak P (2009) Phytohaemagglutinin injection has a long-lasting effect on immune cells. J Avian Biol 40:569–571

Schoech SJ, Rensel MA, Heiss RS (2011) Short- and long-term effects of developmental corticosterone exposure on avian physiology, behavioral phenotype, cognition, and fitness: a review. Curr Zool 57:514–530

Schroeder J, Burke T, Mannarelli ME, Dawson DA, Nakagawa S (2012) Maternal effects and heritability of annual productivity. J Evol Biol 25:149–156

Schweitzer C, Goldstein MH, Place NJ, Adkins-Regan E (2013) Long-lasting and sex-specific consequences of elevated egg yolk testosterone for social behavior in Japanese quail. Horm Behav 63:80–87

Self SG, Liang K-Y (1987) Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions. J Am Stat Assoc 82:605

Stram DODO, Lee JWJW, Kee J, Lee JWJW, Kee J (1994) Variance components testing in the longitudinal mixed effects model. Biometrics 50:1171–1177

Teplitsky C, Mills JA, Yarrall JW, Merilä J (2009) Heritability of fitness components in a wild bird population. Evolution 63:716–726

Teplitsky C, Robinson MR, Merila J (2014) Evolutionary potential and constraints in wild populations. In: Charmantier A, Garant D, Kruuk LEB (eds) Quantitative Genetics in the Wild. Oxford University Press, New York, NY, p 190–208

Thomson CE, Bayer F, Crouch N, Farrell S, Heap E, Mittell E et al. (2017) Selection on parental performance opposes selection for larger body mass in a wild population of blue tits. Evolution 71:716–732

Tobler M, Sandell MI (2007) Yolk testosterone modulates persistence of neophobic responses in adult zebra finches, Taeniopygia guttata. Horm Behav 52:640–645

Tolvanen J, Kivelä SM, Doligez B, Morinay J, Gustafsson L, Bijma P et al. (2020) Quantitative genetics of the use of conspecific and heterospecific social cues for breeding site choice. Evolution 74:2332–2347

True HL, Lindquist SL (2000) A yeast prion provides a mechanism for genetic variation and phenotypic diversity. Nature 407:477–483

Tschirren B, Richner H, Schwabl H (2004) Ectoparasite-modulated deposition of maternal androgens in great tit eggs. Proc R Soc B Biol Sci 271:1371–1375

Tschirren B, Saladin V, Fitze PS, Schwabl H, Richner H (2005) Maternal yolk testosterone does not modulate parasite susceptibility or immune function in great tit nestlings. J Anim Ecol 74:675–682

Tschirren B, Sendecka J, Groothuis TGG, Gustafsson L, Doligez B (2009) Heritable variation in maternal yolk hormone transfer in a wild bird population. Am Nat 174:557–564

Valcu C-M, Scheltema RA, Schweiggert RM, Valcu M, Teltscher K, Walther DM et al. (2019) Life history shapes variation in egg composition in the blue tit Cyanistes caeruleus. Commun Biol 2:1–14

Vedder O, Kingma S-A, von Engelhardt N, Korsten P, Groothuis GGT, Komdeur J (2007) Conspecific brood parasitism and egg quality in blue tits Cyanistes caeruleus. J Avian Biol 38:625–629

Walsh B, Lynch M (2018) Evolution and Selection of Quantitative Traits. Oxford University Press UK, London

Wang H, Zhang F, Zeng J, Wu Y, Kemper KE, Xue A, et al. (2019). Genotype-by-environment interactions inferred from genetic effects on phenotypic variability in the UK Biobank. Sci Adv 5:eaaw3538

Wilson AJ, Nussey DH, Pemberton JM, Pilkington JG, Morris A, Pelletier F et al. (2007) Evidence for a genetic basis of aging in two wild vertebrate populations. Curr Biol 17:2136–2142

Winter V, Elliott JE, Letcher RJ, Williams TD (2013) Validation of an egg-injection method for embryotoxicity studies in a small, model songbird, the zebra finch (Taeniopygia guttata). Chemosphere 90:125–131

Wolf JB, Wade MJ (2009) What are maternal effects (and what are they not?) Philos Trans R Soc B Biol Sci 364:1107–1115

Wyman MJ, Rowe L (2014) Male Bias in Distributions of Additive Genetic, Residual, and Phenotypic Variances of Shared Traits. Am Nat 184:326–337

AUTHOR CONTRIBUTIONS

SMD—conceptualisation, funding acquisition, extensive fieldwork, fieldwork logistics, experimental procedures, laboratory analyses, data management, data analysis, first paper drafts, paper revisions; KJ, DL—extensive fieldwork, fieldwork logistics, experimental procedures, laboratory analyses, data management, paper revisions; MC, JS, AA—fieldwork, commenting on paper drafts; LG—logistic support during fieldwork, paper comments.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41437-021-00490-1.

Correspondence and requests for materials should be addressed to Szymon M. Drobniak.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.