Open access • Journal Article • DOI:10.1007/S00265-017-2308-2

Within- and between-individual (co)variance partitioning reveals limited pleiotropic effects of testosterone on immune function, sexual signaling, and parental investment — Source link

Arne Iserbyt, Marcel Eens, Wendy Baetens, Anke Vermeulen ...+1 more authors

Institutions: University of Antwerp

Published on: 30 Mar 2017 - Behavioral Ecology and Sociobiology (Springer Berlin Heidelberg)

Topics: Challenge hypothesis, Animal ecology, Parental investment and Testosterone (patch)

Within- and between-individual (co)variance partitioning reveals limited pleiotropic effects of testosterone on immune function, sexual signaling, and parental investment

Authors: Arne Iserbyt[a]*, Marcel Eens[a], Wendy Baetens, Anke Vermeulen[a], Wendt Müller[a]

Title: Within- and between-individual (co)variance partitioning reveals limited pleiotropic effects of testosterone on immune function, sexual signaling and parental investment

Affiliation: [a] Department of Biology, Behavioural Ecology and Ecophysiology Group, University of Antwerp, Universiteitsplein 1, B-2610 Wilrijk, Belgium.

* **Corresponding author**:

Email: arne.iserbyt@uantwerpen.be

Tel: +3232652347

Fax: +3232652271

Abstract

How and why individuals differ from each other is a central question in behavioral and evolutionary ecology, because selection particularly acts on this among-individual variation. It is therefore important to accurately partition phenotypic variances into their within- and between-individual components. Partitioning covariances into both components can also inform about underlying mechanistic pathways that potentially interlink trait expression. In the current study we applied such a (co)variance partitioning approach to test key predictions of two central hypotheses in behavioral ecology, namely the immunocompetence handicap hypothesis and the challenge hypothesis. To this end, we assessed potential pleiotropic effects of testosterone on male sexual signaling, immune function and parental care. We here repeatedly measured a set of relevant traits in 47 breeding pairs of captive canaries (*Serinus canaria*). We found that a within-individual increase in female testosterone level suppressed immune function. Furthermore, testosterone levels were positively related to male song repertoire size as an important component of sexual signaling at the between-male level. These were, however, the only relevant significant correlations. Overall, our data do therefore not convincingly support the hypotheses tested, and suggest rather limited hormonal pleiotropic effects of testosterone on immune function, parental care and male sexual signaling, at least in our study system.

Keywords

challenge hypothesis, hormonal pleiotropy, immunocompetence handicap hypothesis, parental care, sexual signaling, suites of traits, trade-off

Significance statement

Phenotypic variances and covariances can nowadays be partitioned easily into within- and between-individual components. These components inform about trait repeatability and the extent to which multiple traits form phenotypic suites of traits, as well as about their joint underlying mechanistic pathways. Testosterone for example, may be key to mediate the expression of suites of traits in many vertebrate species. We here used captive canaries and repeatedly measured male and female traits, relevant in the context of endocrinology, sexual signaling, immune function and parental investment. For both sexes, we report particularly weak correlations between testosterone and all other measured traits at both within- and between-individual levels. Our case study thus questions the pleiotropic effects of testosterone, but exemplifies the applicability and relevance of (co)variance partitioning in behavioral ecological research.

Introduction

Steroid hormones such as testosterone (T) regulate the development of secondary sexual traits like weaponry, bright coloration or complex songs in many male vertebrates (reviewed by Adkins-Regan 2005; Kempenaers et al. 2008), which are typically positively related to male reproductive success. Their expression is thought to honestly signal the quality of the bearer, as only high quality individuals are supposed to sustain the costs associated with their production and maintenance (Grafen 1990; Zahavi and Zahavi 1997). Indeed, high plasma T levels, and thus enhanced sexual signaling, has been associated with a number of costs (summarised by Kempenaers et al. 2008). The two most often mentioned costs are a reduction in pathogen and parasite resistance (cf. immunocompetence handicap hypothesis: Folstad and Karter 1992; Roberts et al. 2004) and a reduction of parental care (cf. a key prediction of the challenge hypothesis: Wingfield et al. 1990; Hirschenhauser and Oliveira 2006; Wingfield 2016). These concurrent positive and negative effects of T on life-history traits, along with the involved physiological trade-offs have been extensively studied, though nearly exclusively in males (but see Ketterson et al. 2005; Zysling et al. 2006). This makes T an important within-individual mechanistic component causing hormone-mediated suites of traits (a mechanism known as hormonal pleiotropy, e.g. Hau 2007; Ketterson et al. 2009) that may be adaptive because it enables an organism to adjust its phenotype to a novel environment via phenotypic shifts (McGlothlin and Ketterson 2008). But the evolution of a single trait may be hampered when entwined with the expression and thus selection on other traits, via a common physiological mechanism (e.g. hormonal pleiotropy) or via a genetic mechanism (e.g. gene pleiotropy or linkage disequilibrium), which also depends on the magnitude and sign of the correlation (Wolf and Weissing 2012).

Phenotypic correlations among traits occur both at a within- and between-individual level. Within-individual correlations indicate the similarity in plastic responses of two focal traits across repeated measurements. In other words, these correlations specify the extent of simultaneous changes in expression of phenotypic traits in response to a common internal or environmental factor. Traits may

thus change in concert, which could be driven by a common mechanistic pathway. Between-individual correlations then indicate the extent of permanent environmental (e.g. maternal effects) and genetic correlations (e.g. by gene pleiotropy or linkage disequilibrium) between focal traits, and may cause multiple traits to respond to selection as a unit (cf. suites of traits; Sinervo and Svensson 2002; Dochtermann and Dingemanse 2013). Covariance partitioning of phenotypic traits may therefore be a first indicative tool to reveal the potential for constrained evolutionary trajectories. Such constraining effects can be substantial, as has been shown for a variety of behavioral (Wolf and Weissing 2012) and endocrine (Ketterson et al. 2009) traits. Phenotypic covariances can nowadays easily be partitioned into within- and between-individual components, which allows for testing basic predictions from central hypotheses such as the immunocompetence handicap and the challenge hypotheses.

Decomposing phenotypic variance into within- and between-individual components (Dingemanse and Dochtermann 2013), is also relevant for our understanding of the (response to) selection of a given trait. Between-individual variation describes how much individuals differ from each other in their average phenotype, and mostly relates to both 'permanent' environmental factors (e.g. maternal effects) and heritable genetic effects (Badyaev and Uller 2009; Dingemanse and Dochtermann 2013). Selection particularly acts on this among-individual variation However, traits often vary substantially within individuals too, which is particularly true for state-dependent physiological and behavioral traits (Rands and Johnstone 2006; Williams 2008; Westneat et al. 2014). Such variation within individuals reflects the phenotypic plasticity of an individual's genotype responding to variable environmental and internal conditions (and to some extent stochastic processes and measurement error; Dingemanse and Dochtermann 2013; Baugh et al. 2014; Cleasby et al. 2015). The extent to which within- or between-individual variation contributes to explain the total phenotypic variation is commonly characterized by the repeatability of a trait (Lessells and Boag 1987; Bell et al. 2009).

In the current study we aimed to provide an empirical example of how (co)variance partitioning can be applied to test central ecological and evolutionary hypotheses. We focused on the steroid hormone

testosterone, which plays a pivotal role in behavioral ecology. More precisely, previous studies highlighted the dual role of T within individuals, in which the extent of sexual signaling is traded off against immune function (immunocompetence handicap hypothesis, Folstad and Karter 1992; Roberts et al. 2004) or parental care (as part of the challenge hypothesis, Wingfield et al. 1990; Hirschenhauser and Oliveira 2006). The majority of these studies focused on the within-individual correlated immunological and behavioral responses to experimentally enhanced T levels (e.g. Duffy et al. 2000; Mougeot et al. 2004). In fact, the same expectations hold for between-individual trade-offs (van Oers et al. 2011). In this case, individuals with relatively high T levels should be characterized by a suite of correlated behavioral and physiological traits. We here assessed the hormonal pleiotropic effects of T (Hau 2007; McGlothlin and Ketterson 2008) on three important traits, i.e. immune function, sexual signaling and parental investment, which were repeatedly measured using captive canaries (*Serinus canaria*) as a model species. We then applied mixed modeling tools (Dingemanse and Dochtermann 2013) to partition the (co)variances of these traits at two hierarchical levels, i.e. within-individuals and between-individuals. In summary, we aimed (1) at assessing the repeatability of each investigated trait under constant environmental conditions (2) at elucidating the mechanistic role of peak plasma T in mediating within-individual trait correlations and (3) at describing between-individual trait correlations as a preliminary indication for the presence of suites of correlated, possibly co-adapted phenotypic traits.

Methods

Study species and animal keeping

During spring 2014, we used a total of 47 adult male and 47 adult female Fife Fancy canaries (*Serinus canaria*) that originated from our own outbred pedigreed population. Males and females were housed separately in two large indoor aviaries, were exposed to a long light regime to induce reproductive activity (14:10, L:D), and had access to seeds and water *ad libitum*. Egg food was provided twice a

week. After five weeks, males were moved to individual cages (50x64x40 cm, GEHU cages, The Netherlands) for the next three weeks to stimulate territorial behaviors. Subsequently, each male was paired by allocating a female to its cage. Nest cups and nesting materials were provided. Progress on nest building, egg laying and incubation was monitored daily (Iserbyt et al. 2015a; Iserbyt et al. 2015b). A first clutch with at least one egg was produced in 45 nests and a second clutch (see further) with at least one egg was produced in 42 nests. Hatching was synchronized within broods to facilitate cross-fostering (Estramil et al. 2013). Successful hatching of at least one egg occurred in 37 nests. Enriched egg food was provided on a daily basis after hatching (day 0). Nestlings were cross-fostered at an early stage (day 0) to standardize parental work load across nests with all foster nests containing four unrelated nestlings. Nestlings were cross-fostered with respect to nestling age (max. 12h age difference between nestlings) to minimize size differences among nestlings and, if possible, with respect to egg order (nestlings hatched from eggs with different egg order) to equalize maternal effects across nests. Within each nest, nestlings were individually marked with a unique color on their back, using non-toxic markers (Artline®70N), which was reapplied when necessary. They were ringed with a unique code on day 7. A total of 36 pairs successfully raised offspring, with on average 3.54 ± 0.12 (SE) fledglings per nest.

Physiological measurements

Plasma testosterone levels

Blood (± 100 µl) was sampled from the alar wing vein for measuring plasma testosterone (T) levels near the expected seasonal peak in T concentrations (Johnsen 1998; Parisot et al. 2005; Ketterson et al. 2005; Goymann et al. 2007). For males, blood was sampled before the pairs were formed, which corresponds to the period of territorial establishment. For females this was done on the day that the first egg was laid (see Fig. 1). Blood was always sampled between 10am and 1pm to reduce potential diurnal variation in T levels (Kempenaers et al. 2008). Plasma concentrations of T were measured using an enzyme-linked immunoassay ELISA kit (Enzo Life Sciences, ADI-900-065), full details of

148 the analytical procedure can be found in Iserbyt *et al*. (2015a). Analytical sensitivity of the ELISA-
149 assay is 5.67 pg/ml and shows 100% cross-reactivity with testosterone, 14.63% with 19-
150 hydroxytestosterone, 7.2% with androstenedione and <1% with other steroid compounds. The assay
151 kit was validated by ensuring parallelism of serial dilutions of plasma samples with the standard 4-
152 parameter logistic curve. Plasma samples were diluted 1:15 in the supplied assay buffer and
153 concentrations of T were assessed with the standard curve for each assay separately. Each plasma
154 sample was analyzed in duplicate (within-assay measurement error: 0.9%). Furthermore, a set of
155 randomly chosen samples (N = 39) was measured twice in different assays, indicating an across-assay
156 measurement error of 3.2%. Similarly, within- and across-assay coefficients of variability based on
157 standard dilutions were 2.51% (N = 50) and 16.4% (N=10), respectively.

158

159 *Immune functioning*

160 Although the effect of blood sampling on an individual's health state is probably very limited (e.g.
161 most males resume singing within one hour and most females continued incubation), a three–day
162 recovery period was allowed before the assessment of immune status. An immune response was
163 induced by injecting phytohaemagglutinin (PHA, 30µg dissolved in 30µl PBS [0.01M])
164 subcutaneously in the left wing web (patagium). Injections took place between 10am and 1pm.
165 Patagium thickness was measured with a pressure-sensitive micrometer (Mitutoyo, accuracy 0.01mm),
166 first prior to injection and a second time exactly 24h post-injection. On both days, measurements were
167 taken six times independently (Smits et al. 1999). A mixed modelling approach, taking into account
168 sex differences and repeated sampling of individuals indicated that 30.0% of all variance is explained
169 by measurement error. The mean of the six independent measurements was therefore used to reduce
170 this error. The wing web swelling in response to an injection of PHA is a commonly used proxy for
171 the contemporary individual immune status as it is indicative for the accumulation of circulating T-
172 lymphocytes to an injected mitogen (Smits et al. 1999).

173

174 ***Behavioral measurements***

175 *Song behavior*

176 Bird song, a sexually selected trait, was recorded for three hours (mean ± 1SE: 187 ± 23 min) between
177 9 am and 3 pm, one or two days before blood sampling. To this end an omnidirectional AV-Jefe
178 TCM141 tie clip microphone was used, which was connected to a M-audio microtrack II recorder
179 (recording settings: WAV, 44,100 Hz, 16 Bits, Mono). The microphone was placed centrally on the
180 trellised front of the cage and as such was pointed towards the cage's interior. The entire cage front
181 was covered with clear perspex to reduce background noise. Spectrographic analyses were performed
182 using Avisoft-SASLab Pro 5.2.07 (Avisoft, Berlin, Germany; spectrogram parameters: FFT length
183 512, frame size 100% and overlap of 50 %). The first ten and last five recorded minutes of the
184 recording were neglected to avoid potential human disturbance effects. Number of song bouts and
185 number of unique syllables per bout were both counted for the entire recording. Song bouts were at
186 least 1.5 seconds and were delineated by a pause of at least 0.4s (Leitner et al. 2001). Forty-four
187 individuals were successfully recorded. Birdsong is a multifaceted behavior that provides conspecifics
188 with information about the signaler's genetic quality or condition (reviewed by Møller and
189 Pomiankowski 1993; Gil and Gahr 2002). Among the different song components we here quantified
190 three key song properties. First, repertoire size is scored as the average number of unique syllables per
191 bout, of which its magnitude or complexity is thought to rise with (genetic) quality (e.g. Hasselquist et
192 al. 1996, but see Byers and Kroodsma 2009). Second, song effort (i.e. the relative time spent singing
193 during the recording; Madison et al. 2015), is considered as a performance-related trait and hence
194 reflects a male's current condition. Finally, song consistency quantifies the ability to consistently
195 repeat specific song elements and is known to contain a variety of dynamic information on the male
196 quality, including age, health status and motor skills (Rivera-Gutierrez et al. 2010; Sakata and
197 Vehrencamp 2012; Vehrencamp et al. 2013). To assess consistency, a specific type of syllable was
198 chosen that is used by most individuals in the population and is defined as a "common syllable"
199 (Vergauwen et al. 2014). Ten of these common syllables were selected for each male, all belonging to
200 different song bouts. The syllables were uploaded in Avisoft-CORRELATOR version 3.1 and the

spectrograms were subsequently compared using an analysis of spectrographic cross-correlation. The resulting matrix of pairwise comparisons was averaged to obtain a single consistency estimate per individual.

Parental provisioning

Parental provisioning behavior was assessed from video recordings that were taken between 10am and 1pm when the nestlings were 10 and 11 days old (see e.g. Estramil et al. 2013). In order to standardize hunger levels across nests, we handfed Orlux (Versele-Laga) to all nestlings until satiation. Nestlings were then returned to their nest and food was removed for one hour prior to video recording, which lasted 2h. Duration (s) of all observed feeding events were scored for males and females separately. For males, we further distinguished feeding events towards chicks (direct feeding) and towards their partner (i.e. mate-feeding: Galván and Sanz 2011; also known as allo-feeding: Estramil et al. 2014). Provisioning was estimated as the total duration of feeding behavior corrected for the length of the recorded video and is presented as minutes per hour. Provisioning estimates like this are commonly used proxies to describe parental care in birds (Mutzel et al. 2013; Bowers et al. 2014; Iserbyt et al. 2015b).

To minimize observer bias, blinded methods were used when all behavioral data were recorded and analyzed.

Morphometrics and maternal investment

Standard morphometric measurements were taken prior to blood sampling. Tarsus length was measured with a digital caliper (Digimatic, Mitutoyo; accuracy 0.01 mm) and body mass with a digital balance (Kern EMB200-2; accuracy 0.01g). Tarsus length remains nearly constant in adult birds, the within-individual variation therefore nearly exclusively represents measurement error (females: 7.2%;

males: 6.4%). Additionally, we quantified the clutch size and egg mass as a measure of female reproductive investment.

Repeated measurements

All traits under study were measured twice to assess their repeatability, with the timing of sampling being adjusted to the specific characteristics of the trait (see Fig. 1). For males, song recordings, blood sampling, immune challenge and morphological measurements were spaced out by 10 days during the period of territorial establishment. Female traits (testosterone level, immune challenge, morphometrics) were measured and sampled on the first day of egg laying of respectively the first and second clutch (Mean ± SE interval: 35 ± 1 days; N = 42). For females that failed to produce a clutch (first clutch: N = 2; second clutch: N = 5), these measurements were taken seven days after the peak of measurements in our experiment. Parental provisioning was scored when parents were feeding offspring of the first clutch on day 10 and day 11, as hatching of the second clutch was inhibited. Female reproductive traits were measured for both clutches while paired with the same male. Except for egg mass, data collected from females that failed to produce a first and/or second clutch were considered as biologically meaningful and are hence included in the statistical analyses (see further). Sample sizes for within- and between-individual correlations (Tables 2 and 3) may show slight deviations from the sample sizes mentioned above, due to technical difficulties during the first or the repeated physiological or behavioural measurement.

Statistical analyses

All analyses were performed in SAS 9.3 (SAS Institute Inc., Cary, NC, U.S.A.), following the equations and the exact scripts as published by Dingemanse and Dochtermann (2013). Normality assumptions were checked for all variables via Shapiro Wilk tests. To meet these assumptions, testosterone levels and mate feeding were log-transformed. All variables were centered by conversion

to standardized z-scores (similar to Baugh et al. 2014; Thys et al. 2017). Specifically, univariate mixed models including random intercepts were used to partition within- and between-individual variances. Repeatability is then calculated as the proportion of total variance explained by between-individual variation (Lessells and Boag 1987). Similar to the convention on effect sizes of correlations (Møller and Jennions 2002; Garamszegi et al. 2012), we consider repeatability values <0.3 as low, between 0.3 and 0.5 as intermediate and >0.5 as high.

Next, bivariate mixed models with random intercepts for the factor 'individual' were used to partition within- and between-individual covariances among all pairwise trait combinations (equations 7a,b detailed in Dingemanse and Dochtermann 2013). Specifically, these models generate two matrixes. The first matrix provides the between-individual variances for both individual traits, the between-individual covariance and the between-trait covariance. The second matrix provides within-individual and within-trait variances, as well as the within-individual and between-trait covariances. Within- and between-individual correlations are then calculated separately as the respective covariance divided by the square root of both multiplied within-individual (or between-individual) trait variances (see equations 7c,d in Dingemanse and Dochtermann 2013). We chose to perform multiple bivariate mixed models, rather than one multivariate mixed model, because individuals with one or more missing values were entirely excluded by the statistical program. Thus the bivariate mixed models made optimal use of the given data, by including the highest possible sample size for each trait combination (see Table 2 and 3). These models were performed for males and females separately. Significance levels were determined via likelihood-ratio tests. For this, the absolute difference in 2*LogLikelihood from an original model and its respective alternative model was compared against a χ^2-distribution. For the univariate analyses, this alternative model excluded the random individual effect. For the alternative models in the bivariate analyses, either the between- or within-covariance component was fixed to zero (Dingemanse and Dochtermann 2013).

Results

276

277 *Within- and between-individual variation*

278 All but one of the investigated traits were significantly repeatable in both males and females (see
279 Table 1; Fig. 2). The immune response was least repeatable, but approached significance in males (r=
280 0.26 ± 0.15; P= 0.083) and in females (r= 0.27 ± 0.15; P=0.054; Fig. 2b). As expected, morphological
281 traits were highly repeatable (male: $r \geq$ 0.90; female: $r \geq$ 0.84). All behavioral traits varied more
282 between than within individuals (provisioning behavior: $r \geq$ 0.56, Fig. 2c; song activity: $r \geq$ 0.62).
283 Likewise, all female reproductive traits were highly repeatable (clutch size: r= 0.54 ± 0.17; average
284 egg mass: r=0.72 ± 0.19).

285

286 *Within- and between-individual correlations*

287 The only significant within-individual correlation in females was found between peak plasma T levels
288 and immune response (r_0 = -0.19; χ^2_1 = 4.2; P = 0.040; see Table 2; Fig. 3e). Specifically, this negative
289 correlation indicates that an increase of the immune response from the first to the second measurement
290 corresponds with a decrease in plasma T levels. The magnitude and sign of this correlation was
291 similar, but not significant, in males (r_0 = -0.19; χ^2_1 = 1.7; P = 0.19; see Table 3; Fig. 3e). In fact, none
292 of the tested within-individual correlations were significant in males (see Table 3).

293 Peak plasma T levels correlated negatively with body mass in a between-female comparison (r_{ind}= -
294 0.42; χ^2_1 = 5.13; P = 0.024). In general, female body mass was positively associated with tarsus length
295 (r_{ind}= 0.33; χ^2_1 = 4.9; P = 0.034), clutch size (r_{ind}= 0.51; χ^2_1 = 7.49; P = 0.006), and egg mass (r_{ind}=
296 0.36; χ^2_1 = 6.10; P = 0.013). Heavy females also tended to provide less food to their nestlings (r_{ind}= -
297 0.36; χ^2_1 = 3.51; P = 0.061). Furthermore, a negative between-individual correlation was observed
298 between immune response and tarsus length (r_{ind}= -0.49; χ^2_1 = 3.90; P = 0.048). No other significant
299 between-individual correlations were observed for females (see Table 2; Fig. 3b,c).

Male plasma testosterone levels did not correlate significantly with any of the other traits (between-individual correlations), except for a positive relationship with repertoire size (r_{ind}= 0.59; χ^2_1 = 4.0; P = 0.046; Fig. 3a). Song activity (r_{ind}= 0.37; χ^2_1 = 5.8; P = 0.015) and body mass (r_{ind}= 0.32; χ^2_1 = 4.6; P = 0.032) were positively correlated with tarsus length. Male body mass was also positively associated with the immune response (r_{ind}= 0.41; χ^2_1 = 3.9; P = 0.048). The only significant negative between-individual correlation was found between song consistency and mate feeding (r_{ind}= -0.52; χ^2_1 = 4.6; P = 0.032). No other significant between-individual correlations were observed for males (see Table 3; Fig. 3b,c).

Discussion

Covariance partitioning is a particularly useful statistical tool to decompose phenotypic variances and covariances into their within- and between-individual components. These components contain different information about the extent to which multiple traits are linked via a common mechanism or form phenotypic suites of traits. This approach has only recently been applied within an ecological framework, and it may also be highly suited to test central hypotheses in behavioral ecology, as we exemplify in this study.

Within-individual correlations

Within-individual correlations inform about potential mechanistic pathways, because they indicate how quantitative changes in time for one trait correspond with changes in one or several other traits (Dingemanse and Dochtermann 2013). In the current study, we found weak evidence for correlated temporal changes across traits within individual males and females. In fact, none of these correlations were significant after correcting for a potential type-I error (false positives, see also last part of this discussion). However, these weak or absent within-individual correlations may also be caused by a type-II error (false negatives), or due to the particularly stable laboratory conditions between repeated measurements. That said, female peak testosterone levels correlated negatively with their immune

response to the PHA challenge. Despite the relatively high measurement error in this latter trait, this observation is in line with predictions from the immunocompetence handicap hypothesis (Folstad and Karter 1992). A similar, though insignificant within-individual trade-off was observed in males, given the power issue as described above. It is worth emphasizing that the mechanistic role of T has been extensively studied, though nearly exclusively in males (but see Ketterson et al. 2005; Zysling et al. 2006). However, females are also known to circulate significant levels of androgens, to possess androgen receptors, as well as androgen metabolizing enzymes (Staub and De Beer 1997). Despite increased research interest, the adaptive value of female T and its causal effects on the expression of phenotypic and behavioral traits, as well as on immune function remain equivocal (Owen-Ashley et al. 2004; Ketterson et al. 2005; Jawor et al. 2006; Goymann et al. 2007; O'Neal et al. 2008; Iserbyt et al. 2015a). A more elaborate discussion on the immunocompetence handicap hypothesis and challenge hypothesis is given below.

Between-individual correlations

Elevated T levels are thought to be beneficial for males as they enhance sexual signaling (inter-sexual selection), competitiveness (intra-sexual selection) and thus reproductive success via increased mating opportunities (Oliveira 2004; Kempenaers et al. 2008). We indeed found a strong positive (between-individual) correlation between male T levels and repertoire size. Repertoire size has traditionally been considered as a sexually selected song trait in many bird species (Gil and Gahr 2002), including canaries (Gil et al. 2004). Indeed, male repertoire size has been positively related to female preferences (Searcy 1992; but see Byers and Kroodsma 2009), extra-pair paternity (Hasselquist et al. 1996) and shown to influence egg yolk composition (Gil et al. 2004). However, it has been suggested by by Byers and Kroodsma (2009) that repertoire size may not have originated through sexual selection per se, but may rather represent a by-product of social selection. It has been hypothesized that multiple song characteristics reflect different messages (Gil and Gahr 2002) and hence, selection for increasingly complex social communication may have favored large repertoire sizes. This multiple

352 message hypothesis may at least in part explain why we did not observe any significant correlations
353 among our three different song parameters.

354

355 Current theory predicts that sexually selected traits should be costly in order to function as honest
356 indicators of quality in mate choice (Grafen 1990; Zahavi and Zahavi 1997; Gil and Gahr 2002).
357 These costs may lead to trade-offs, the two most prominent in this context are captured by the
358 immunocompetence handicap hypothesis (Folstad and Karter 1992) and the challenge hypothesis
359 (Wingfield et al. 1990). These mechanistic trade-offs are typically studied at a within-individual level
360 by measuring the immunological and behavioral responses to experimentally enhanced T levels (e.g.
361 Duffy et al. 2000; Mougeot et al. 2004). At a between-individual level however, males with relatively
362 high T values could be characterized by a suite of correlated behavioral and physiological traits, a
363 mechanism known as hormonal pleiotropy (Reviewed by Hau 2007; McGlothlin and Ketterson 2008;
364 Ketterson et al. 2009). A recent empirical example comes from great tits (*Parus major*) where artificial
365 selection lines for divergent personalities ('fast- bold' and 'slow-shy' explorers) showed a correlated
366 response in plasma levels of testosterone, immune response (van Oers et al. 2011) and corticosterone
367 (Baugh et al. 2012), which in turn may alter several metabolic functions. Our data also indicate that
368 the expression of repertoire size is T-dependent (see also Van Hout et al. 2012), but we could not
369 confirm this for any other song parameters (here: song activity and consistency) and neither did we
370 find evidence for the immunocompetence handicap hypothesis or a trade-off with parental care in
371 males as predicted by the challenge hypothesis. After accounting for type-I errors and considering the
372 probability of type-II errors, we may have to conclude that the weak and limited number of between-
373 individual correlations with T questions the role of hormonal pleiotropy in shaping an individual's
374 phenotype, at least in our study system.

375 Tables 2 and 3 also depict several other interesting between-individual correlations. Although
376 discussing these correlations in detail goes beyond the aims of this manuscript, and bearing in mind
377 potential type-I and type-II errors (see above and details below), it is worth noting that our

morphological traits (body mass and tarsus length) were significantly related to multiple physiological and behavioral traits. This suggests that specific suites of morphological, physiological and behavioral traits exists which are partly linked by a common mechanism, like gene pleiotropy or linkage disequilibrium. It is thus possible that past artificial selection in this domesticated canary type may have caused correlated responses for several other linked phenotypic traits.

Within- versus between-individual variation

All other studied traits except the immune response were significantly repeatable, including traits that are considered to be labile and state-dependent (Westneat et al. 2014). These are traits related to female reproduction (clutch size $r = 0.54$; egg mass $r = 0.71$), behavior (range $r = 0.35$-0.87) and endocrinology (peak T level male $r = 0.36$; female $r = 0.31$). Such overall high and significant between-individual variation relative to within-individual variation is important from an evolutionary perspective because repeatability estimates are often considered as an upper bound estimate of heritability (Lessells and Boag 1987; Dohm 2002; Greives et al. 2016), which forms a prerequisite for evolutionary responses to selection. We observed small sex differences in repeatability estimates, which contrasts with the general pattern across taxa that males are more repeatable than females for example in behavioral traits (Bell et al. 2009). It has been hypothesized that T-related male sexually selected traits should to be more predictable, providing females with reliable cues during mate choice (summarized by Bell et al. 2009). Females are also considered as the more responsive sex, e.g. towards the needs of the brood (Johnstone and Hinde 2006; Nakagawa and Schielzeth 2010; Iserbyt et al. 2015b), inherently implying lower repeatability of female parental care compared with males. However, we did not find evidence for this. The absence of sex-differences in repeatability in our study may in part be explained by the stable laboratory environment and the relatively short interval between repeated measurements (i.e. days to weeks).

Our study may also have practical implications for future sampling designs. Given the relatively high repeatability in most of our traits, a single snapshot in time might be sufficient to quantify (between-individual) physiological and behavioral trait variation, at least within a defined period of the

405 reproductive cycle (e.g. during territory defense or during egg laying) and under constant
406 environmental conditions (Trösch et al. 2017). However, artificial selection during domestication may
407 also have reduced sensitivity to environmental conditions (Price 1999), which could also explain the
408 relatively high repeatability.

409

410 *Covariance partitioning in behavioral ecological studies – a preliminary conclusion*

411 Covariance partitioning is a particularly useful statistical tool, but it requires large sample sizes to
412 draw firm conclusions. The statistical power depends on the sample size (here: maximum 47
413 individuals per sex), the number of measurements per individual (here: 2), the repeatability of a trait
414 (here: on average in females: $r = 0.49$; males: $r = 0.50$), and the observed magnitude of the between-
415 individual correlation (here: on average for females: $r_{ind} = 0.28$; males: $r_{ind} = 0.20$). With these
416 parameters, the predicted power is about 0.3 and it would require 200 individuals per sex to reach a
417 power threshold of 0.8 (according to Dingemanse and Dochtermann 2013). Thus our study may face a
418 high probability of type-II error, and several true correlations may have remained undetected.
419 However, the statistical power is proportional to the magnitude of r_{ind} (Dingemanse and Dochtermann
420 2013), which makes it very difficult to achieve the required sample size when between-individual
421 correlations are low, as is often the case (see Tables 2 and 3). Thus whether typical behavioral
422 ecological studies reach the power threshold of 0.8 in cases of low r_{ind} remains to be shown. We
423 would also like to add a final cautionary note concerning the potential inflation of false positive
424 correlations (i.e. type-I errors) as a result of multiple testing with the bivariate mixed models.
425 Applying Bonferroni corrections to these models would lower the significance level to $\alpha = 0.002$ and α
426 $= 0.005$ for between- and within-female correlations and to $\alpha = 0.001$ and $\alpha = 0.003$, respectively for
427 males. None of the correlations in table 2 and 3 were found to be significant following these
428 corrections.

429 Covariance partitioning, once the above mentioned aspects are taken into account, is nevertheless very
430 promising as it allows to elucidate mechanistic pathways within individuals and informs about the

between-individual covariance structure among traits of interest (for a recent example, see Thys et al. 2017). Future studies elucidating the mechanistic pathways within individuals could benefit from combining this statistical approach with experimental manipulations, for example by comparing behavioral and physiological traits before and after elevating T levels (e.g. Ketterson et al. 2005; Lynn 2008; but see Goymann and Wingfield 2014). It may also be a logical next step to apply a behavioral reaction norm concept (Dingemanse et al. 2010). In this approach, traits are compared across different environmental conditions, for example, before and after exposure to a territorial intruder (Wingfield 1985). Finally, relating T profiles with various physiological and behavioral traits across successive stages of reproduction, thus capturing temporal dynamics of within-individual correlations, may form another promising line of research (also discussed in Iserbyt et al. 2015a).

Acknowledgements

We thank Geert Eens and Peter Scheys for their assistance in taking care of the birds, as well as the entire Family Ecology Unit for their general support and discussion during the breeding season. Bert Thys and four anonymous reviewers provided constructive comments, which helped to improve the clarity of this manuscript. This research was financially supported by FWO Flanders (project ID: 1517815N and 12I1916N to A.I. and project ID: G.0102.12N to W.M. and M.E.) and the University of Antwerp (project ID: 22959 to W.M.).

Compliance with ethical standards

Ethical standards

All experiments were carried out in accordance with the guidelines of the Ethical Committee of the University of Antwerp, Belgium (ID: 2014-72).

Conflict of interest

The authors declare no conflict of interest.

456

457

458 **References**

459

460 Adkins-Regan E (2005) Hormones and Animal Social Behaviour. Princeton University Press,
461 Princeton, New Jersey

462 Badyaev AV, Uller T (2009) Parental effects in ecology and evolution: mechanisms, processes and
463 implications. Philos Trans R Soc Lond B Biol Sci 364:1169–1177. doi: 10.1098/rstb.2008.0302

464 Baugh AT, van Oers K, Dingemanse NJ, Hau M (2014) Baseline and stress-induced glucocorticoid
465 concentrations are not repeatable but covary within individual great tits (*Parus major*). Gen
466 Comp Endocrinol. doi: 10.1016/j.ygcen.2014.08.014

467 Bell AM, Hankison SJ, Laskowski KL (2009) The repeatability of behaviour: a meta-analysis. Anim
468 Behav 77:771–783. doi: 10.1016/j.anbehav.2008.12.022

469 Bowers EK, Nietz D, Thompson CF, Sakaluk SK (2014) Parental provisioning in house wrens: effects
470 of varying brood size and consequences for offspring. Behav Ecol 25:1485–1493. doi:
471 10.1093/beheco/aru153

472 Byers BE, Kroodsma DE (2009) Female mate choice and songbird song repertoires. Anim Behav
473 77:13–22. doi: 10.1016/j.anbehav.2008.10.003

474 Cleasby IR, Nakagawa S, Schielzeth H (2015) Quantifying the predictability of behaviour: statistical
475 approaches for the study of between-individual variation in the within-individual variance.
476 Methods Ecol Evol 6:27–37. doi: 10.1111/2041-210X.12281

477 Dingemanse NJ, Dochtermann NA (2013) Quantifying individual variation in behaviour: mixed-effect
478 modelling approaches. J Anim Ecol 82:39–54. doi: 10.1111/1365-2656.12013

479 Dingemanse NJ, Kazem AJN, Réale D, Wright J (2010) Behavioural reaction norms: animal

480 personality meets individual plasticity. Trends Ecol Evol 25:81–89. doi:
481 10.1016/j.tree.2009.07.013

482 Dochtermann NA, Dingemanse NJ (2013) Behavioral syndromes as evolutionary constraints. Behav
483 Ecol 24:806–811. doi: 10.1093/beheco/art002

484 Dohm MR (2002) Repeatability estimates do not always set an upper limit to heritability. Funct Ecol
485 16:273–280. doi: 10.1046/j.1365-2435.2002.00621.x

486 Duffy DL, Bentley GE, Drazen DL, Ball GF (2000) Effects of testosterone on cell-mediated and
487 humoral immunity in non-breeding adult European starlings. Behav Ecol 11:654–662. doi:
488 10.1093/beheco/11.6.654

489 Estramil N, Eens M, Müller W (2013) Coadaptation of offspring begging and parental provisioning -
490 an evolutionary ecological perspective on avian family life. PLoS One 8:e70463. doi:
491 10.1371/journal.pone.0070463

492 Estramil N, Eens M, Müller W (2014) On the coadaptation of offspring begging and parental supply-a
493 within-individual approach across life stages. Behav Ecol Sociobiol 68:1481–1491. doi:
494 10.1007/s00265-014-1756-1

495 Folstad I, Karter AJ (1992) Parasites, bright males, and the immunocompetence handicap. Am Nat
496 139:603–622.

497 Galván I, Sanz JJ (2011) Mate-feeding has evolved as a compensatory energetic strategy that affects
498 breeding success in birds. Behav Ecol 22:1088–1095. doi: 10.1093/beheco/arr094

499 Garamszegi LZ, Markó G, Herczeg G (2012) A meta-analysis of correlated behaviours with
500 implications for behavioural syndromes: Mean effect size, publication bias, phylogenetic effects
501 and the role of mediator variables. Evol Ecol 26:1213–1235. doi: 10.1007/s10682-012-9589-8

502 Gil D, Gahr M (2002) The honesty of bird song: multiple constraints for multiple traits. Trends Ecol
503 Evol 17:133–141. doi: 10.1016/S0169-5347(02)02410-2

504 Gil D, Leboucher G, Lacroix A, et al (2004) Female canaries produce eggs with greater amounts of

testosterone when exposed to preferred male song. Horm Behav 45:64–70. doi: 10.1016/j.yhbeh.2003.08.005

Goymann W, Landys MM, Wingfield JC (2007) Distinguishing seasonal androgen responses from male-male androgen responsiveness–Revisiting the Challenge Hypothesis. Horm Behav 51:463–476. doi: 10.1016/j.yhbeh.2007.01.007

Goymann W, Wingfield JC (2014) Male-to-female testosterone ratios, dimorphism, and life history–what does it really tell us? Behav Ecol 25:685–699. doi: 10.1093/beheco/aru019

Grafen A (1990) Biological signals as handicaps. J Theor Biol 144:517–546. doi: 10.1016/S0022-5193(05)80088-8

Greives TJ, Dochtermann NA, Stewart EC (2016) Estimating heritable genetic contributions to innate immune and endocrine phenotypic correlations: A need to explore repeatability. Horm Behav. doi: 10.1016/j.yhbeh.2016.11.015

Hasselquist D, Bensch S, von Schantz T (1996) Correlation between male song repertoire, extra-pair paternity and offspring survival in the great reed warbler. Nature 381:229–232. doi: 10.1038/381229a0

Hau M (2007) Regulation of male traits by testosterone: implications for the evolution of vertebrate life histories. BioEssays 29:133–144. doi: 10.1002/bies.20524

Hirschenhauser K, Oliveira RF (2006) Social modulation of androgens in male vertebrates: Meta-analyses of the challenge hypothesis. Anim Behav 71:265–277. doi: 10.1016/j.anbehav.2005.04.014

Iserbyt A, Eens M, Müller W (2015a) Sexually antagonistic selection during parental care is not generated by a testosterone-related intralocus sexual conflict–insights from full-sib comparisons. Sci Rep 5:17715. doi: 10.1038/srep17715

Iserbyt A, Farrell S, Eens M, Müller W (2015b) Sex-specific negotiation rules in a costly conflict over parental care. Anim Behav 100:52–58. doi: 10.1016/j.anbehav.2014.11.014

530 Jawor JM, Young R, Ketterson ED (2006) Females competing to reproduce: Dominance matters but
531 testosterone may not. Horm Behav 49:362–368. doi: 10.1016/j.yhbeh.2005.08.009

532 Johnsen TS (1998) Behavioural correlates of testosterone and seasonal changes of steroids in red-
533 winged blackbirds. Anim Behav 55:957–965. doi: 10.1006/anbe.1997.0642

534 Johnstone RA, Hinde CA (2006) Negotiation over offspring care - how should parents respond to each
535 other's efforts? Behav Ecol 17:818–827. doi: 10.1093/beheco/arl009

536 Kempenaers B, Peters A, Foerster K (2008) Sources of individual variation in plasma testosterone
537 levels. Philos Trans R Soc B - Biol Sci 363:1711–1723. doi: 10.1098/rstb.2007.0001

538 Ketterson ED, Atwell JW, McGlothlin JW (2009) Phenotypic integration and independence:
539 Hormones, performance, and response to environmental change. Integr Comp Biol 49:365–379.
540 doi: 10.1093/icb/icp057

541 Ketterson ED, Nolan V, Sandell M (2005) Testosterone in females: mediator of adaptive traits,
542 constraint on sexual dimorphism, or both? Am Nat 166:S85–S98. doi: 10.1086/444602

543 Leitner S, Voigt C, Gahr M (2001) Seasonal changes in the song pattern of the non-domesticated
544 island canary (*Serinus Canaria*) a field study. Behaviour 138:885–904. doi:
545 10.1163/156853901753172700

546 Lessells CM, Boag PT (1987) Unrepeatable repeatabilities: a common mistake. Auk 104:116–121.

547 Lynn SE (2008) Behavioral insensitivity to testosterone: why and how does testosterone alter paternal
548 and aggressive behavior in some avian species but not others? Gen Comp Endocrinol 157:233–
549 240. doi: 10.1016/j.ygcen.2008.05.009

550 Madison FN, Rouse ML, Balthazart J, Ball GF (2015) Reversing song behavior phenotype:
551 Testosterone driven induction of singing and measures of song quality in adult male and female
552 canaries (*Serinus canaria*). Gen Comp Endocrinol 61–75. doi: 10.1016/j.ygcen.2014.09.008

553 McGlothlin JW, Ketterson ED (2008) Hormone-mediated suites as adaptations and evolutionary
554 constraints. Philos Trans R Soc Lond B Biol Sci 363:1611–1620. doi: 10.1098/rstb.2007.0002

555 Møller AP, Jennions MD (2002) How much variance can be explained by ecologists and evolutionary
556 biologists? Oecologia 132:492–500. doi: 10.1007/s00442-002-0952-2

557 Møller AP, Pomiankowski A (1993) Why have birds got multiple sexual ornaments? Behav Ecol
558 Sociobiol 32:167–176. doi: 10.1007/BF00173774

559 Mougeot F, Irvine JR, Seivwright L, et al (2004) Testosterone, immunocompetence, and honest sexual
560 signaling in male red grouse. Behav Ecol 15:930–937. doi: 10.1093/beheco/arh087

561 Mutzel A, Dingemanse NJ, Araya-Ajoy YG, Kempenaers B (2013) Parental provisioning behaviour
562 plays a key role in linking personality with reproductive success. Proc R Soc B - Biol Sci
563 280:20131019.

564 Nakagawa S, Schielzeth H (2010) Repeatability for Gaussian and non-Gaussian data: A practical
565 guide for biologists. Biol Rev 85:935–956. doi: 10.1111/j.1469-185X.2010.00141.x

566 O'Neal DM, Reichard DG, Pavilis K, Ketterson ED (2008) Experimentally-elevated testosterone,
567 female parental care, and reproductive success in a songbird, the Dark-eyed Junco (*Junco*
568 *hyemalis*). Horm Behav 54:571–578. doi: 10.1016/j.yhbeh.2008.05.017

569 Oliveira RF (2004) Social modulation of androgens in vertebrates: mechanisms and function. Adv
570 Study Behav 34:165–239.

571 Owen-Ashley NT, Hasselquist D, Wingfield JC (2004) Androgens and the immunocompetence
572 handicap hypothesis: unraveling direct and indirect pathways of immunosuppression in song
573 sparrows. Am Nat 164:490–505. doi: 10.1086/423714

574 Parisot M, Tanvez A, Lacroix A, et al (2005) Social competition and plasma testosterone profile in
575 domesticated canaries: An experimental test of the challenge hypothesis. Horm Behav 48:225–
576 232. doi: 10.1016/j.yhbeh.2005.02.011

577 Price EO (1999) Behavioral development in animals undergoing domestication. Appl Anim Behav Sci
578 65:245–271. doi: 10.1016/S0168-1591(99)00087-8

579 Rands SA, Johnstone RA (2006) Statistical measures for defining an individual's degree of

580 independence within state-dependent dynamic games. BMC Evol Biol 6:81. doi: 10.1186/1471-
581 2148-6-81

582 Rivera-Gutierrez HF, Pinxten R, Eens M (2010) Multiple signals for multiple messages: great tit,
583 *Parus major*, song signals age and survival. Anim Behav 80:451–459. doi:
584 10.1016/j.anbehav.2010.06.002

585 Roberts ML, Buchanan KL, Evans MR (2004) Testing the immunocompetence handicap hypothesis: a
586 review of the evidence. Anim Behav 68:227–239. doi: 10.1016/j.anbehav.2004.05.001

587 Sakata JT, Vehrencamp SL (2012) Integrating perspectives on vocal performance and consistency. J
588 Exp Biol 215:201–209. doi: 10.1242/jeb.056911

589 Searcy WA (1992) Song repertoire and mate choice in birds. Am Zool 32:71–80. doi:
590 10.1093/icb/32.1.71

591 Sinervo B, Svensson E (2002) Correlational selection and the evolution of genomic architecture.
592 Heredity (Edinb) 89:329–338. doi: 10.1038/sj.hdy.6800148

593 Smits JE, Bortolotti GR, Tella JL (1999) Simplifying the phytohaemagglutinin skin-testing technique
594 in studies of avian immunocompetence. Funct Ecol 13:567–572. doi: 10.1046/j.1365-
595 2435.1999.00338.x

596 Staub NL, De Beer M (1997) The role of androgens in female vertebrates. Gen Comp Endocrinol
597 108:1–24. doi: 10.1006/gcen.1997.6962

598 Thys B, Eens M, Aerts S, et al (2017) Exploration and sociability in a highly gregarious bird are
599 repeatable across seasons and in the long term but are unrelated. Anim Behav 123:339–348. doi:
600 10.1016/j.anbehav.2016.11.014

601 Trösch M, Müller W, Eens M, Iserbyt A (2017) Genes, environments and their interaction: Song and
602 mate choice in canaries. Anim Behav in press. doi: 10.1016/j.anbehav.2017.02.006

603 Van Hout AJM, Pinxten R, Darras VM, Eens M (2012) Testosterone increases repertoire size in an
604 open-ended learner: An experimental study using adult male European starlings (*Sturnus*

605 *vulgaris*). Horm Behav 62:563–568. doi: 10.1016/j.yhbeh.2012.09.008

606 van Oers K, Buchanan KL, Thomas TE, Drent PJ (2011) Correlated response to selection of
607 testosterone levels and immunocompetence in lines selected for avian personality. Anim Behav
608 81:1055–1061. doi: 10.1016/j.anbehav.2011.02.014

609 Vehrencamp SL, Yantachka J, Hall ML, de Kort SR (2013) Trill performance components vary with
610 age, season, and motivation in the banded wren. Behav Ecol Sociobiol 67:409–419. doi:
611 10.1007/s00265-012-1461-x

612 Vergauwen J, Groothuis TGG, Eens M, Müller W (2014) Testosterone influences song behaviour and
613 social dominance - but independent of prenatal yolk testosterone exposure. Gen Comp
614 Endocrinol 195:80–87. doi: 10.1016/j.ygcen.2013.10.014

615 Westneat DF, Wright J, Dingemanse NJ (2014) The biology hidden inside residual within-individual
616 phenotypic variation. Biol Rev 90:729–743. doi: 10.1111/brv.12131

617 Williams TD (2008) Individual variation in endocrine systems: moving beyond the "tyranny of the
618 Golden Mean". Philos Trans R Soc Lond B Biol Sci 363:1687–1698. doi:
619 10.1098/rstb.2007.0003

620 Wingfield JC (2016) The challenge hypothesis: Where it began and relevance to humans. Horm
621 Behav. doi: 10.1016/j.yhbeh.2016.11.008

622 Wingfield JC (1985) Short-term changes in plasma levels of hormones during establishment and
623 defense of a breeding territory in male song sparrows, *Melospiza melodia*. Horm Behav 19:174–
624 187. doi: 10.1016/0018-506X(85)90017-0

625 Wingfield JC, Hegner RE, Dufty AM, Ball GF (1990) The "challenge hypothesis": theoretical
626 implications for patterns of testosterone secretion, mating systems, and breeding strategies. Am
627 Nat 136:829–846.

628 Wolf M, Weissing FJ (2012) Animal personalities: Consequences for ecology and evolution. Trends
629 Ecol Evol 27:452–461. doi: 10.1016/j.tree.2012.05.001

630 Zahavi A, Zahavi A (1997) The handicap principle. Oxford University Press, Oxford, UK

631 Zysling DA, Greives TJ, Breuner CW, et al (2006) Behavioral and physiological responses to
632 experimentally elevated testosterone in female dark-eyed juncos (*Junco hyemalis carolinensis*).
633 Horm Behav 50:200–7. doi: 10.1016/j.yhbeh.2006.03.004

634

635

Tables

Table 1: Overview of the measured traits. Mean (± SD) values are given for untransformed and unstandardized estimates. N represents the number of individuals for which both independent measurements could be taken. The variable sample size used for each trait are caused by technical problems during one of both measurements. The presented within (V_0) - and between (V_{ind}) individual variation is based on univariate mixed models, for which the variables were standardized to z-scores. Hence, between-individual variation here also represent repeatability for each trait. P-values of significantly repeatable traits are highlighted in bold.

Parameter	Mean ± SD	N	V_0 ± SE	V_{ind} ± SE	Chi^2	P
(a) Female						
Testosterone (ng/ml)	1.01 ± 0.57	47	0.69 ± 0.14	0.31 ± 0.15	4.9	**0.027**
Wing web swell (mm)	0.55 ± 0.18	47	0.73 ± 0.15	0.27 ± 0.15	3.7	*0.054*
Tarsus length (mm)	17.83 ± 0.58	47	0.07 ± 0.02	0.93 ± 0.20	90.9	**<0.0001**
Body mass (g)	22.90 ± 2.43	47	0.18 ± 0.04	0.82 ± 0.20	48.3	**<0.0001**
Clutch size (#)	3.61 ± 1.20	47	0.46 ± 0.09	0.54 ± 0.17	16.3	**<0.0001**
Egg mass (g)	1.80 ± 0.14	42	0.28 ± 0.06	0.72 ± 0.19	30.5	**<0.0001**
Provisioning (min/h)	2.45 ± 0.98	36	0.38 ± 0.09	0.62 ± 0.20	16.6	**<0.0001**
(b) Male						
Testosterone (ng/ml)	4.80 ± 2.53	42	0.63 ± 0.13	0.37 ± 0.16	6.3	**0.012**
Wing web swell (mm)	0.41 ± 0.18	44	0.74 ± 0.15	0.26 ± 0.15	3.0	*0.083*
Tarsus length (mm)	18.16 ± 0.62	44	0.06 ± 0.01	0.94 ± 0.19	93.0	**<0.0001**
Body mass (g)	18.38 ± 1.45	45	0.07 ± 0.01	0.93 ± 0.19	84.8	**<0.0001**
Provisioning (min/h)	1.59 ± 0.67	36	0.43 ± 0.10	0.57 ± 0.19	14.1	**<0.0001**
Mate provisioning (min/h)	0.64 ± 0.72	36	0.13 ± 0.03	0.87 ± 0.22	47.8	**<0.0001**
Repertoire size (syll./bout)	6.82 ± 1.79	39	0.55 ± 0.12	0.45 ± 0.17	9.0	**0.003**
Song activity (%)	3.62 ± 3.11	44	0.37 ± 0.07	0.63 ± 0.17	22.3	**<0.0001**
Consistency	0.46 ± 0.04	30	0.64 ± 0.16	0.36 ± 0.19	4.1	**0.043**

Table 2: Output of the bivariate mixed models for females showing between-individual correlations (± SE) above, and within- individual correlations (± SE) below the diagonal. Given that our data are converted to z-scores, the regression coefficient (β) equals the correlation coefficient (r) in each model. Hence, each value indicates the unit of change, sign, and magnitude for each correlation. The value in brackets indicate the sample size used within each bivariate model. Slight variations in sample size used for each model are caused by technical problems during one of both measurements. Within-individual correlations between tarsus length and all other traits are not presented, assuming that this trait is nearly fixed in adult birds. Similarly, within-individual correlations between female provisioning (Pr) and all other traits are not indicated, given the asynchrony in our sampling scheme (Figure 1), hence these correlations are considered biologically irrelevant. Significant correlations are indicated in bold and asterisks indicate the following: (*) 0.1 < p < 0.05; * 0.05 < p < 0.01; ** 0.01 < p < 0.001. Applying Bonferroni corrections (α level: P = 0.002 - 0.005) to reduce type-I errors resulted in insignificant correlations for all pairwise comparisons.

	T	**WS**	**Ta**	**BM**	**CS**	**EM**	**Pr**
Testosterone (T)	-	0.02±0.15 (47)	-0.14±0.15 (47)	**-0.42*±0.15 (47)**	0.26±0.15 (47)	-0.13±0.15 (45)	-0.06±0.16 (36)
Web swell (WS)	**-0.19*±0.15 (46)**	-	**-0.49*±0.14 (47)**	0.30±0.15 (47)	0.58(*)±0.14 (47)	0.32±0.14 (45)	0.32±0.17 (36)
Tarsus (Ta)	-	-	-	**0.33*±0.15 (47)**	0.18±0.14 (47)	0.21±0.14 (45)	-0.14±0.17 (36)
Body mass (BM)	-0.09±0.29 (46)	0.09±0.29 (47)	-	-	**0.51**±0.13 (47)**	**0.36*±0.13 (45)**	-0.36(*)±0.16 (36)
Clutch size (CS)	0.03±0.18 (46)	-0.18±0.23 (47)	-	-0.02±0.23 (47)	-	0.27±0.12 (45)	0.15±0.15 (36)
Egg mass (EM)	-0.11±0.21 (41)	0.04±0.19 (42)	-	0.15±0.12 (42)	-0.05±0.19 (42)	-	0.08±0.17 (35)

660 **Table 3:** Output of the bivariate mixed models for males. Within-individual correlations between male
661 provisioning (Pr) or mate-provisioning (MP) and all other traits are not indicated, given the
662 asynchrony in our sampling scheme (Figure 1). For further details, see caption table 2. Bonferroni
663 corrections (α level: P = 0.001 – 0.003) resulted in insignificant correlations for all pairwise
664 comparisons.

665

	T	WS	Ta	BM	R	SA	C	Pr	MP
Testosterone (T)	-	0.25±0.16 (43)	-0.25±0.16 (43)	-0.04±0.15 (43)	**0.59*±0.16 (41)**	0.34±0.15 (43)	0.25±0.16 (38)	-0.14±0.15 (36)	0.04±0.15 (36)
Web swell (WS)	-0.19±0.17 (43)	-	0.29±0.15 (44)	**0.41*±0.15 (44)**	0.05±0.15 (42)	-0.08±0.15 (44)	-0.31±0.16 (39)	0.38±0.17 (36)	-0.03±0.17 (36)
Tarsus (Ta)	-	-	-	**0.32*±0.14 (47)**	-0.18±0.16 (42)	**0.37*±0.15 (44)**	-0.23±0.17 (39)	-0.05±0.17 (36)	0.18±0.17 (36)
Body mass (BM)	0.00±0.05 (43)	0.23±0.50 (44)	-	-	0.10±0.16 (42)	-0.04±0.16 (44)	-0.28±0.16 (39)	-0.33±0.15 (36)	0.06±0.16 (36)
Repertoire (R)	0.02±0.15 (38)	0.01±0.20 (39)	-	0.11±0.06 (39)	-	0.2±0.16 (42)	0.15±0.17 (39)	-0.1±0.21 (35)	0.08±0.22 (35)
Song activity (SA)	0.00±0.12 (43)	0.00±0.21 (44)	-	-0.06±0.36 (44)	0.08±0.20 (39)	-	0.12±0.15 (39)	-0.08±0.18 (36)	0.2±0.17 (36)
Consistency (C)	-0.05±0.18 (29)	0.07±0.20 (30)	-	0.07±0.53 (30)	-0.04±0.19 (30)	0.17(*)±0.23 (30)	-	0.12±0.19 (33)	**-0.52*±0.18 (33)**
Provisioning (Pr)	-	-	-	-	-	-	-	-	0.08±0.17 (36)
Mate provisioning (MP)	-	-	-	-	-	-	-	-0.05±0.31 (35)	-

666

667

668 **Figure legends**

669

670 **Fig. 1:** Indicative timeline of the experimental procedure. Phenotypic measurements are indicated by
671 letters. The apostrophe (‘) indicates the second (repeated) measurement. A: song recordings (males).
672 B: blood sampling (testosterone levels), body mass and tarsus length measurements (males). C:
673 immune challenge (males). D: blood sampling (testosterone levels), body mass and tarsus length
674 measurements (females). E: immune challenge (females). F: parental provisioning (males and
675 females).

676

677 **Fig. 2:** Graphical representation of the repeatability analyses for four key traits in our study. Values
678 are centered around 0 for both measurements (m1 and m2) separately. Positive values indicate
679 relatively high values within our study population, and *vice versa*. Regression fits are drawn when
680 significant in the univariate mixed model. Filled dots and solid regression lines represent males, open
681 circles and dashed lines represent females.

682

683 **Fig. 3:** Graphical representation of the tested hypotheses at the between- (a-c) and within- (d, e)
684 individual level. Testosterone levels are related to repertoire size (a, d) which tests the sexual signaling
685 hypothesis, to wing web swelling in response to an injection of PHA (b, e) which tests the
686 immunocompetence hypothesis, and to offspring provisioning (c) which tests the trade-off hypothesis
687 with parental care. The within-individual correlation between testosterone level and offspring
688 provisioning is not presented given the asynchrony in our sampling scheme (Figure 1). Filled dots
689 represent males and open dots represent females. Trait values are centered around 0 for each sex
690 separately. Positive and negative values in graphs a-c indicate respectively relatively high and low
691 average values (mean of measurement 1 and 2) within our study population. Positive and negative
692 values in graphs d and e indicate respectively increased or decreased values for each trait from

693 measurement 1 to 2. Regression fits and 95% confidence bands are only drawn if found to be
694 significant in the bivariate mixed model.