

MS. BARBARA CLASS (Orcid ID : 0000-0002-2361-9821)

Article type : Research Article

Editor : Holger Schielzeth

Corresponding Author Email ID: barbara.class@utu.fi

## **A statistical methodology for estimating assortative mating for phenotypic traits that are labile or measured with error**

Barbara Class<sup>1</sup>, Niels J. Dingemans<sup>2</sup>, Yimen G. Araya-Ajoy<sup>3</sup>, Jon E. Brommer<sup>1</sup>

<sup>1</sup>Department of Biology, University of Turku, University Hill, Turku 20014, Finland

<sup>2</sup>Behavioural Ecology, Department of Biology, Ludwig Maximilian University of Munich (LMU), Planegg-Martinsried, Germany

<sup>3</sup>Centre for Biodiversity Dynamics (CBD), Department of Biology, Norwegian University of Science and Technology (NTNU), N-7491 Trondheim, Norway

### **Keywords**

Assortative mating, Non-random assortment, Repeatability, Quantitative genetics, Common environment, Mixed-effect modelling, Bivariate model

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/2041-210X.12837

This article is protected by copyright. All rights reserved.

## Summary

1. Assortative mating in wild populations is commonly reported as the correlation between males' and females' phenotypes across mated pairs. Theories of partner selection and quantitative genetics assume that phenotypic resemblance of partners captures associations in “intrinsically determined” trait values. However, when considering traits with a repeatability below one (labile traits or traits measured with error), the correlation between phenotypes of paired individuals can arise from shared environmental effects on the phenotypes of paired individuals or correlated measurement error.
2. We introduce statistical approaches to estimate assortative mating in labile traits, or traits measured with error in the presence of shared environmental effects. These approaches include i) the correlation between the mean phenotypes of males and females, ii) the correlation between randomized values of individuals and iii) the between-pair correlation derived from a bivariate mixed model.
3. We use simulations to show that the performance of these different approaches depends on the number of repeated measures within individuals or pairs, which is determined by study design, and rates of survival and divorce.
4. We conclude that short-term environmental effects on phenotypes of paired individuals likely inflate estimates of assortative mating when not statistically accounted for. Our approach allows investigation of this important issue in assortative mating studies for labile traits (e.g., behavior, physiology, or metabolism) in both socially monogamous and other mating systems, and groupings of individuals outside a mating context.

## Introduction

Phenotypes of males and females are often correlated within breeding pairs of socially monogamous species. For instance, human partners are more similar than expected by chance with regards to physical traits (Silventoinen, Kaprio, Lahelma, Viken & Rose 2003), socioeconomic status (Mascie-Taylor 1987), education (Mare 1991; Lewis & Oppenheimer 2000), intelligence (Mascie-Taylor 1989; Plomin & Deary 2015), and personality (Mascie-Taylor 1989; Keller, Thiessen & Young 1996; Glicksohn & Golan 2001). Individuals can also be paired with partners of dissimilar phenotype, such as in white-throated sparrows *Zonotrichia albicollis*, where two color morphs exist and over 90% of the population pairs disassortatively (Thornycroft 1975). Nonrandom assortment of individuals can result from individuals choosing their partner based on phenotypic similarity/dissimilarity (Jiang, Bolnick & Kirkpatrick 2013), or from other processes (Burley 1983). This nonrandom assortment generates a positive ('assortative mating') or negative ('disassortative mating') correlation between the homologous traits of mated individuals (Wright 1921).

Nonrandom assortment and its mechanisms have been extensively studied in the wild because of its evolutionary implications. For instance, assortative mating has been proposed to drive speciation (Johannesson, Rolán-Alvarez & Ekendahl 1995; Seehausen, Alphen & Witte 1997; Maan & Seehausen 2011; Bolnick & Kirkpatrick 2012; Langerhans & Makowicz 2013). Assortative mating for heritable traits, furthermore, can increase the additive genetic variance, or affect the strength of genetic correlations by causing linkage disequilibrium, and influence micro-evolution (Gimelfarb 1986; Falconer & MacKay 1996; Lynch & Walsh 1998). Importantly, individuals are assumed to mate non-randomly with respect to individual-specific properties, such as an individual's genetic make-up or expression of a trait that is highly repeatable or fixed (Kirkpatrick & Barton 1997). When applied to less repeatable

Accepted Article

(“labile”) traits, such as behavior, individuals are assumed to mate (dis)assortatively with respect to their average phenotypes. Nevertheless, other processes can generate a correlation between phenotypes of mated individuals in the wild (Jiang et al. 2013) and need to be taken into account (Pérez-Figueroa, de Uña-Alvarez, Conde-Padín & Rolán-Alvarez 2008; Snowberg & Bolnick 2012).

Assortative mating studies focus primarily on fixed traits (Jiang et al. 2013). Recent research, however, increasingly addresses nonrandom mating with respect to labile traits (e.g., behaviour: Dingemanse, Both, Drent & Tinbergen 2004; Ariyomo & Watt 2013; Kralj-Fišer, Sanguino Mostajo, Preik, Pekár & Schneider 2013; Montiglio, Wey, Chang, Fogarty & Sih 2016; body condition and ornaments: Jiang et al. 2013). Importantly, because labile traits (such as behaviour, hormone levels, or aspects of immunology) are repeatedly expressed within an individual, they can be plastically adjusted to environmental conditions between time points. Labile traits therefore vary not just *among* but also *within* individuals. As we will detail below, the assumption that an individual’s trait value (i.e., phenotype) measured at some point during its life reflects its “intrinsic” individual-specific value is a rather strong assumption for labile traits. Hence, partners may look more alike because they both respond similarly to a shared environmental factor, were measured by the same observer, etc., rather than because they were assortatively mated.

Here, we discuss the importance of acknowledging multi-level variation when interpreting correlations between phenotypes of partners for labile phenotypic traits or traits measured with error. We mathematically describe the relative contributions of assortative mating, correlated environmental effects and correlated measurement error to correlations between phenotypes of partners, highlighting the role of the trait’s repeatability in mediating each mechanism’s overall effect. We further compare the utility of various statistical approaches used to estimate nonrandom assortment for labile phenotypes and investigate the

performance of each approach using simulations. We provide R scripts for researchers interested in estimating nonrandom assortment. Finally, we provide guidelines with regard to study design and statistical analyses of assortative mating for a range of biological systems.

### **A variance-partitioning description of assortative mating**

Nonzero phenotypic correlations between traits of partners are often viewed as evidence for nonrandom assortment with respect to individual-specific phenotypes (Wright 1921; Falconer & Mackay 1996). We use a variance partitioning approach to illustrate why this assertion only stands for phenotypic traits that do not harbor within-individual variance (i.e., ‘fixed’ traits, like adult skeleton size in species with determinate growth, measured without error). Individual-specific phenotypes exist when a repeatedly measured trait is repeatable because it harbors significant among-individual variance ( $V_{ind}$ ). Phenotypic variation ( $V_P$ ) is also attributable to within-individual plastic responses to environmental variation ( $V_e$ ) and measurement error ( $V_{me}$ ):

$$V_P = V_{ind} + V_e + V_{me} \quad \text{Eqn.1}$$

The latter variance component ( $V_{me}$ ) would be estimated as “residual” variance ( $V_\epsilon$ ) in statistical analyses (i.e.,  $V_\epsilon = V_{me}$ ) as would variation not attributable to individual differences (i.e.  $V_\epsilon = V_{me} + V_e$ ) in statistical models not modeling environmental effects causing within-individual variation ( $V_e$ ). While  $V_e$  is 0 for “fixed” traits where phenotypic variation represents among-individual variation (unless measured with error),  $V_e$  is known to explain a large portion of variation in labile traits. For example, meta-analyses imply that, respectively, 63%, 55%, and 85% of phenotypic variance in behavior (Bell, Hankinson & Laskowski 2009), metabolism or hormones (Holtmann, Lagisz & Nakagawa 2017) is within-individuals.

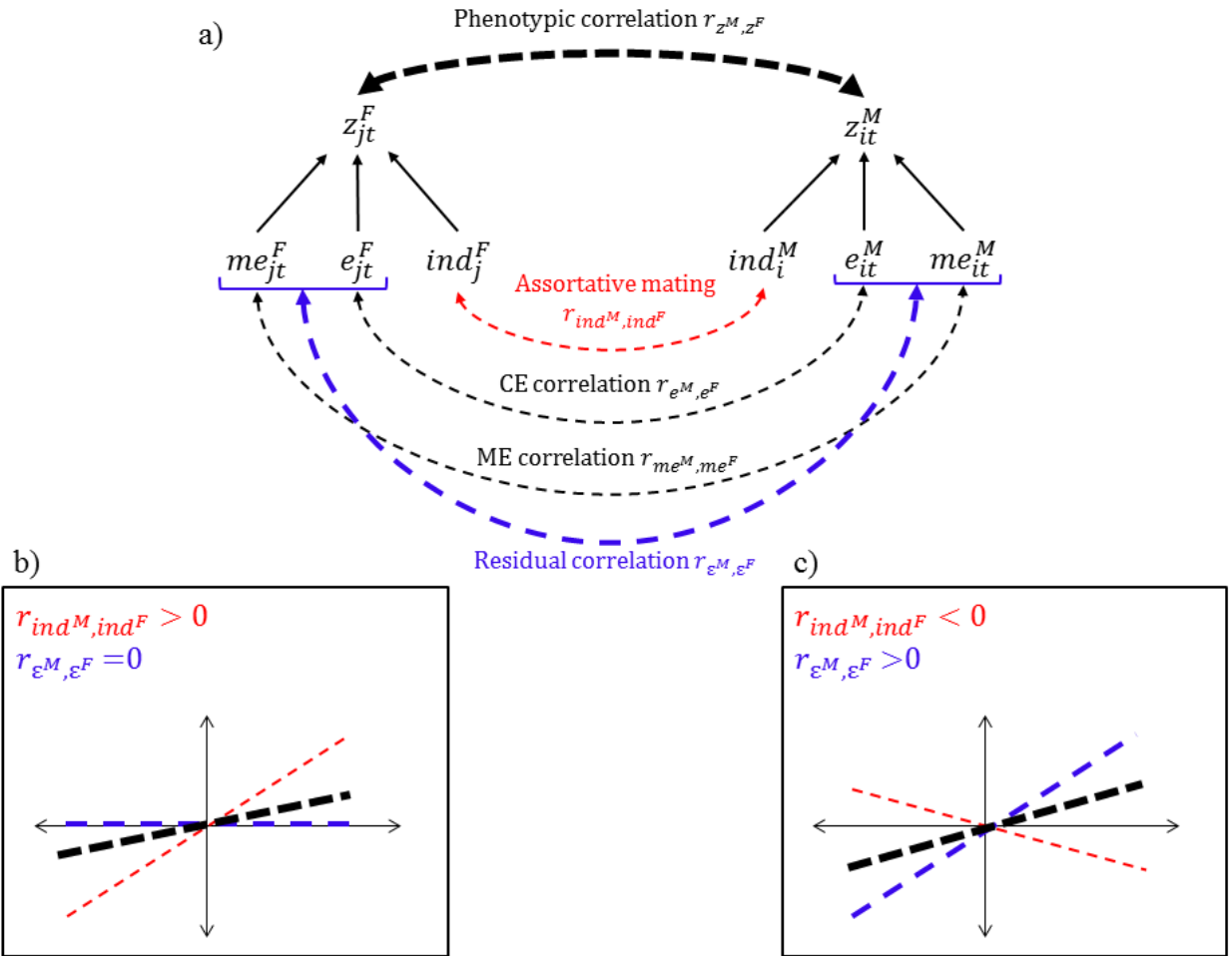
The multi-level nature of phenotypic variation is illustrated in Figure 1a, where values for a labile trait measured at the same time point  $t$  in a male  $i$  ( $i = 1 \dots n^M$ ) and a female  $j$  ( $j = 1 \dots n^F$ ) forming a pair are determined by the sum of their individual-specific values ( $ind_i^M, ind_j^F$ ), environmental effects ( $e_{it}^M, e_{jt}^F$ ), and measurement error ( $me_{it}^M, me_{jt}^F$ ). We assume here, for simplicity, that each level follows a univariate normal distribution, and that phenotypes are not affected by other factors, such as their partner's phenotype (further assumptions are described in Text S1). The existence of variance at multiple levels implies that covariance (correlation) between partners' phenotypes can also exist at multiple levels. While the correlation between individual-specific values  $r_{ind^M, ind^F}$  represents our statistical definition of "true" nonrandom mating for a repeatable trait, the environmental correlation  $r_{e^M, e^F}$  represents the correlation across the sexes in within-individual responses to the environment which arises because individuals forming a pair respond to the same environmental factor (i.e. food availability or predation risk) varying spatio-temporally across pairs. In other words, a positive correlation between male and female phenotype can arise because individuals increase their response when breeding in a type A habitat and decrease it when breeding in a type B habitat in a situation where habitat types are heterogeneously distributed across pairs. Finally,  $r_{me^M, me^F}$  represents correlated measurement error across the sexes, existing, for instance, because the same observer measures both partners at the same time. The phenotypic correlation ( $r_{z^M, z^F}$ ) between paired individuals thus results from the sum of assortative mating, common environmental correlation and correlated measurement error weighed by the proportional contribution of each level to the total phenotypic variance (Roff 1997). Hence, the effect of assortative mating on the phenotypic correlation is proportional to the trait's geometric mean repeatability ( $R_{geom}$ ), representing the weighted average of the two sex-specific repeatabilities (see Supplementary Text S1 for a step-by-step

Accepted Article

explanation). The influence of the residual correlation  $r_{\varepsilon^M, \varepsilon^F}$  (i.e., the combined effects of  $r_{me^M, me^F}$  and  $r_{e^M, e^F}$ ) on the phenotypic correlation is thus proportional to  $1 - R_{geom}$ .

$$r_{z^M, z^F} = R_{geom} * r_{ind^M, ind^F} + (1 - R_{geom}) * r_{\varepsilon^M, \varepsilon^F} \quad \text{Eqn.2}$$

Because many labile traits have repeatabilities below 0.5, their phenotypic correlations will largely represent the residual within-pair correlation caused by the environment shared by pair members or measurement error. Obviously, the phenotypic correlation can represent an unbiased estimate of assortative mating when the repeatability is one or when correlations do not differ between levels (i.e.,  $r_{ind^M, ind^F} = r_{\varepsilon^M, \varepsilon^F}$ ). In all other cases, phenotypic correlations will systematically under- or over-estimate assortative mating (Figure 1b,c). In sum, depending on the sign and magnitude of the residual within-pair correlation, and the trait's repeatability, the phenotypic correlation between partners' traits will either underestimate or overestimate true assortative mating. This problem has not been explicitly recognized in assortative mating studies, most of which rely on phenotypic correlations. We detail various approaches to estimate assortative mating for labile traits or traits measured with error below.



**Figure 1. Correlations at multiple levels shape the overall phenotypic correlation between paired individuals. In a), phenotypes of paired males and females ( $z_{it}^M, z_{jt}^F$ ) are determined by individual-specific values ( $ind_i^M, ind_j^F$ ), environmental effects ( $e_{it}^M, e_{jt}^F$ ) and measurement error ( $me_{it}^M, me_{jt}^F$ ), where 1/3 of the phenotypic variation is due to individual differences and the remainder to residual variation ( $\epsilon = me + e$ ). Correlations can exist at each level (dashed arrows); the phenotypic correlation equals (Eqn. 2)  $r_{z^M, z^F} = 1/3 * r_{ind^M, ind^F} + 2/3 * r_{\epsilon^M, \epsilon^F}$ . The phenotypic correlation thus b) underestimates or c) overestimates the correlation caused by assortative mating.**

### Statistical approaches to estimate assortative mating

We introduce here three statistical alternatives to the “classical” phenotypic approach to estimate assortative mating in labile traits or traits measured with error. We study whether



these approaches can provide unbiased estimates of assortative mating, and how this depends on the amount of replication of pairs and individuals in the data.

(a) *Correlation between mean phenotypes of males and females of each unique pair (CIM)*

This approach seeks to correct for repeated measures within unique breeding pairs, by using individual means instead of unique phenotypic values (Roulin 1999). Here, a unique pair is defined by the unique combination of a male and a female forming a pair; a single individual can hence be a partner in multiple pairs. The CIM approach estimates the correlation between the mean value of males and females ( $r_{\bar{z}^M, \bar{z}^F}$ ) within a pair, which intuitively approximates our measure of assortative mating (i.e.,  $r_{ind^M, ind^F}$ ). Indeed, the individual-mean of all phenotypic values tends to correspond to its individual-specific value (Text S1) and the correlation between the means of mates tends to the correlation between their individual specific values. As detailed by Snijders & Bosker (1999) and Dingemanse et al. (2012), for correlations between two traits assayed repeatedly within a single individual, the correlation between mean values does only approximate the correlation between individual-specific values though under specific conditions; their exact relationship is:

$$r_{\bar{z}^M, \bar{z}^F} = \sqrt{\left(\frac{V_{ind}^M}{V_{ind}^M + \frac{V_{\epsilon}^M}{n}}\right) \left(\frac{V_{ind}^F}{V_{ind}^F + \frac{V_{\epsilon}^F}{n}}\right)} * r_{ind^M, ind^F} + \left(1 - \sqrt{\left(\frac{V_{ind}^M}{V_{ind}^M + \frac{V_{\epsilon}^M}{n}}\right) \left(\frac{V_{ind}^F}{V_{ind}^F + \frac{V_{\epsilon}^F}{n}}\right)}\right) * \frac{r_{\epsilon^M, \epsilon^F}}{n}$$

Eqn. 3

where  $n$  represents the number of repeated measures collected for each unique individual (here assumed to be the same across sexes). Eqn. 3 clarifies that this approximation *only* holds for specific cases where the sample size per individual ( $n$ ) is relatively large. The majority of empirical studies, by contrast, have access to very few repeated measures per

individual, implying that the approach is typically expected to return estimates biased towards the residual correlation.

*(b) Correlation between randomized values of individuals (CIR)*

In situations where the within-pair residual correlation ( $r_{\varepsilon^M, \varepsilon^F}$ ) is zero, the correlation between individual-specific values (Eqn. 2) simplifies to:

$$r_{ind^M, ind^F} = \frac{r_{z^M, z^F}}{R_{geom}} \quad \text{Eqn. 4}$$

A zero within-pair residual correlation may be achieved by randomizing all observed phenotypic values of an individual over its records, followed by the calculation of a simple phenotypic correlation to approximate  $r_{z^M, z^F}$  in Eqn.4. This approach should return unbiased estimates of assortative mating despite utilizing the phenotypic correlation because the randomization procedure produces data where the residual within-pair correlation is zero. A drawback is that it requires multiple analyses, to estimate repeatability for each sex, and calculate the phenotypic correlation ( $r_{z^M, z^F}$ ) for each run of the randomized dataset. The calculated uncertainty associated with the individual-level correlation ( $r_{ind^M, ind^F}$ ) should thus be inflated because, in addition to the normal sampling bias and estimation uncertainty, randomization bias and estimation uncertainty for repeatability cause additional noise.

*(c) Between- and within-pair correlations derived from bivariate mixed-effects models*

Mixed-effect models can also be used to estimate trait (co)variances at multiple hierarchical levels within a single dataset (Lynch & Walsh 1998; Wilson et al. 2009; Dingemanse & Dochtermann 2013), and may be applied to the level of the unique pairs in assortative mating analyses. Instead of defining phenotypes at the individual level, one considers the same data categorized at the level of unique pairs. For every male  $i$  and female  $j$  forming a unique pair  $p$  ( $p = 1 \dots n_p$ ), where  $n_p$  is the number of unique pairs in the data, the expression of a labile trait

$z$  measured for the male (superscript  $M$ ) and the female (superscript  $F$ ) is defined such that male and female phenotypes vary across pairs ( $pair$ ):

$$z_{pt}^M = \mu^M + pair_p^M + \varepsilon_{pt}^M$$

$$z_{pt}^F = \mu^F + pair_p^F + \varepsilon_{pt}^F$$

Eqn.5

where the pair-specific effect ( $pair$ ) and the residual within-pair (co)variances affecting male and female phenotypes are both assumed to follow a bivariate normal distribution with zero mean and (co)variance matrix to be estimated. The pair-identity correlation ( $r_{pair^M, pair^F}$ ) and within-pair residual correlation ( $r_{\varepsilon^M, \varepsilon^F}$ ) is calculable from these (co)variance matrices using standard approaches. This model may seem somewhat removed from our original definition of assortative mating (Eqn.2). By contrast, the pair-level bivariate mixed model simply represents a sophisticated multi-level version of the “traditional” phenotypic approach (defining assortative mating as the association of traits across sexes within pairs). The advantage of the mixed model is that it uses all phenotypic measurements: it includes all repeated measures made on unique pairs (instead of just 1 in the phenotypic approach). It also statistically controls for unequal replication within pairs (which approach (a) does not), and it directly estimates the residual within-pair correlation (instead of controlling for in approach (b)). Partitioning of variance across and within pairs requires replication of measurements on the pair level, which differs from replications on the individual level if individuals change partners across observations. Because the identity of the individual is not included in the pair-level model, we assume that all pairs are independent, and do not consider while additional individual identity effects (i.e., an individual mated with different partners is indistinguishable from other same-sex individuals). The pair-level mixed model approach

uses data efficiently when two members of the same pair re-mate in subsequent breeding attempts because it results in replication at the pair level.

Our summary detailed above implies that the performance of each approach should depend on the number of observations per individual and/or unique pair (Table 1). As individuals can be a member of >1 unique pair, the number of unique pairs is equal or greater than the number of same-sex individuals. Individual-level approaches thus have more replication than pair-level approaches. In the next section, we test the relative performance of each approach.

We do so using simulations, focusing on populations of two hypothetical species where the amount of replication on the individual and pair levels are similar (the “immortal albatross” scenario) or different (the “bluish tit” scenario) due to stable or unstable pair composition.

**Table 1: Level of replication associated with each approach and characteristics affecting it**

<b>Approach</b>	<b>Replicate level</b>	<b>Replication affected by</b>
Phenotypic	Unique pair	number of unique pairs (once per pair)
CIM	Individual	number of individuals and number of trials per individual
CIR	Individual	number of individuals and number of trials per individual
Bivariate mixed model	Unique pair	number of unique pairs and number of trials per unique pair

### **Simulation studies**

We simulated phenotypic data (e.g. aggressiveness) in two populations of birds representing stable (albatrosses) and unstable (tits) pair composition scenarios, with both populations consisting of pairs breeding every year at the same site. We varied amount of replication at

the individual or pair level, and investigated the distribution of the estimates obtained with each approach (Table 1). We considered scenarios where the phenotypic correlation between the sexes was due to i) assortative mating (i.e., correlations between *ind* values across sexes) only or, ii) common environmental effects only, while measurement error was assumed uncorrelated between partners.

#### *Methods for simulating assortative mating*

We implemented an approach that simulated phenotypic data for any population size with i) varying survival and divorce rates, ii) varying repeatabilities and iii) varying correlations between mated individuals on different hierarchical levels. To model re-mating in the population due to divorce, we first generated the identities of individuals forming pairs and then the phenotypic values (using a “network approach”; Text S2). Specifically, we generated identities of individuals forming pairs in the population, reflecting the mating structure of the population. Every year,  $n_{pt}$  pairs of individuals are formed; the number of males and females is constant over the years. From one year to the other, a certain number of males and females survive (determined by  $p_{surv}$ , a proportion of the number of males and females). Males and females that do not survive are replaced by the same number of new males and females. Every year, single individuals are mated randomly but if both partners of a pair survive, they can re-mate with probability  $p_{remate}$ . Phenotypic data, finally, is generated for every individual using different steps. First, individual-specific values, with correlation of  $\rho_{ind}$  between sexes, were randomly drawn. Second, we added environmental effects common to the pair (correlation of  $\rho_{env}$ ) on top. Third, we added error to each individual’s measurement of its phenotype, which was uncorrelated between sexes. All effects followed a Gaussian distribution. The complete procedure is detailed in Supplementary Text S2.

We first generated data for the “immortal albatross” population, characterized by an exceptionally high longevity and fidelity ( $p_{surv} = 1$ ,  $p_{remate} = 1$ ). In this population, composed of 200 breeding pairs (200 males and 200 females), individuals do not die, are always mating with the same partner and are tested for their aggressiveness once every year. Hence, individuals and pairs all have the same amount of replication, equal to the number of years of data collection. Aggressiveness scores of paired individuals can be affected (in a correlated or independent way) by an environmental factor that both share during a breeding season but which can change across breeding seasons (e.g. environmental characteristics of their territory, see Araya-Ajoy & Dingemanse 2017). In this simulated example, individuals breed in nests that were randomly assigned to pairs, thus potentially changing every year (contrary to real albatross species, see Bried, Pontier & Jouventin 2003). Phenotypic data was generated for two scenarios (see Text S3 for further scenarios) where the assortative mating ( $\rho_{ind}$ ) and environmental ( $\rho_{env}$ ) correlations are: i)  $\rho_{ind} = 0.5$  and  $\rho_{env} = 0$  and ii)  $\rho_{ind} = 0$  and  $\rho_{env} = 0.5$ , for study durations 3 and 10 years. Mean aggressiveness in males and females was zero-centered and individual, environmental and residual variances, were respectively 0.3, 0.35 and 0.35, and the same between sexes (i.e., repeatability = 0.3). Phenotypic data was generated for each combination of scenario and sampling design 100 times. Using the same methodology, we generated phenotypic data for the “bluish tit” population, characterized by moderate rates of survival and remating ( $p_{surv} = 0.6$  and  $p_{remate} = 0.5$ ). For this scenario, we varied the number of years (3, 5, 10), the annual number of breeding pairs (50, 100, 200, 300, 500), the number of repeats per individual per year (1, 2), and the trait’s repeatability (0.1 to 0.7), applied to both the ‘assortative mating only’ and ‘common environment only’ scenarios.

### *Data analysis*

For each dataset generated in each scenario, we estimated assortative mating for each approach. Pearson's correlations were calculated between males and females' mean values and randomized values, and the phenotypic correlation (defined as the Pearson's correlation of the first phenotypic value collected for each male and female within each unique pair; Class, Kluever & Brommer 2014). For the bivariate mixed model, where male and female aggressiveness are the response variables, the pair identity and a pair-specific environmental factor (e.g. nest identity), were fitted as random effects. This model differs from the model introduced above (Eqn.5) in that it additionally decomposes the within-pair covariance into covariance due to the common environment versus correlated measurement error. This extension thus allowed for testing the performance of the bivariate mixed model when an extra random effect is modelled (note that when this factor is unknown or not modelled, the common environmental covariance is estimated as residual (i.e., within-pair) covariance). We calculated the distribution (mode and 95% credible intervals) for the different types of correlations (indicative of assortative mating, common environment, and residual correlations) over all iterations and scenarios. For the "bluish tit" scenario, we also quantitatively compared bias, imprecision, coverage and power for the CIM, CIR and bivariate mixed model approaches in each dataset (Text S5). All the simulations and analyses were performed in R (R Core Team 2016, R version 3.3.0.), using the package "asreml" (VSN International, Hemel Hempstead, UK; Butler, Cullis, Gilmour & Gogel 2009). We provide the code for generating all simulations and analyses (Text S5).

### *Comparison of the different approaches for both populations*

The performance of the approaches for albatrosses and tits is visualized in figure 2, which shows the distribution of estimates for the "assortative mating only" (Scenario 1) and

“common environmental correlation only” (Scenario 2) simulations for populations composed of 200 pairs measured once per year for either 3 or 10 years, and a trait repeatability of 0.3. Note that while the number of repeats per individual and pair is equal to the number of years in the albatross population, this number varies, respectively from 1.66 to 2.17 for individuals and from 1.14 to 1.19 for pairs for the 3 and 10 year studies in the tit population.

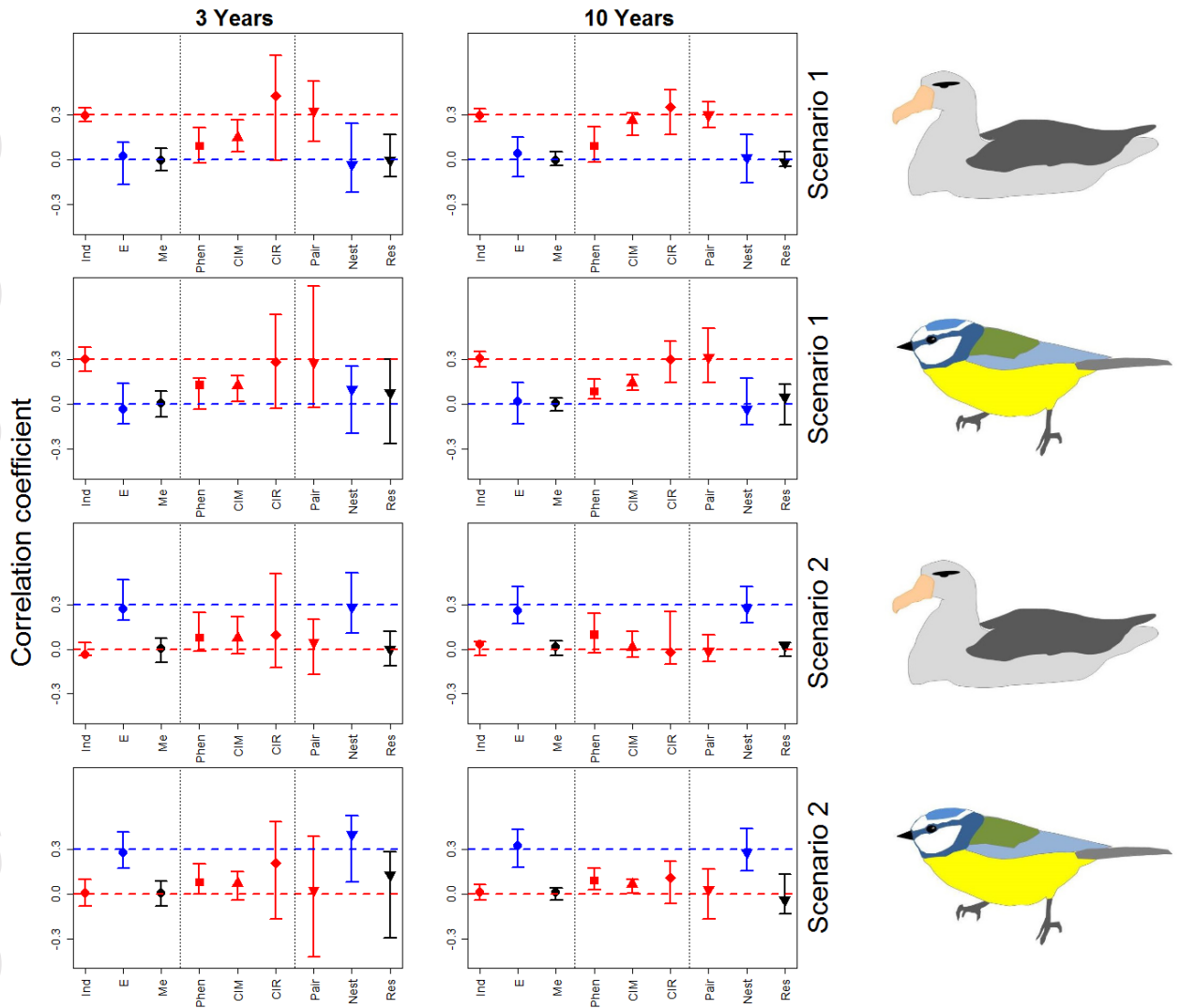
As predicted, the phenotypic correlation is biased when the assortative mating vs. common environment correlation differ (Figure 2). In scenario 1, the distribution of the phenotypic correlation does not cover the true assortative mating correlation and often includes zero. In scenario 2, the phenotypic correlation tends to overestimate true assortative mating and is biased towards the common environment correlation.

The performance of the CIM approach depends on the population and number of years. Though performing reasonably well in albatrosses with many years of data, it remains biased in tits, where increasing the number of years only reduces imprecision (width of the distribution of estimates). In line with Eqn. 3, the CIM estimates are generally biased towards the phenotypic correlation when the number of replicates at the individual level is low (see Text S4 for further detail). By contrast, the performance of the CIR approach is affected by the number of years and the specific scenario. Though relatively unbiased in scenario 1 (where the common environmental correlation is by definition 0), this approach is biased towards the phenotypic correlation in scenario 2. It overestimates assortative mating for both species for 3-year studies, because the low number of replicates does not enable the randomization procedure to properly suppress environmental correlations between mated individuals. Although the CIR approach has a relatively low precision, it becomes more precise within increasing number of years (see also Text S4). Hence, the performance of both individual-level approaches critically depends on the number of replicates observations per



Accepted Article  
individuals. Importantly, both approaches are biased (towards the phenotypic correlation) when the number of replicates is low (for Scenario 2), which increases the risk of spuriously detecting assortative mating when it is absent (Type I error, see Text S4).

By contrast, the pair-level bivariate mixed model accurately estimates the true assortative mating correlation; regardless of number of replicates, population, or scenario, it returns relatively unbiased estimates for the common environment and residual correlations. Though least precise (compared to CIM and CIR approaches), its imprecision decreases as the number of replicates on the pair level increases (i.e., with increasing numbers of years, or when pairs are stable). Although the bivariate mixed model is the most conservative approach as shown by our power analysis, this approach becomes the most powerful when pairs are measured twice per year (Text S4). This is because repeated measures within years increase the pair-level replication, which consequentially increases power. By contrast, the individual-level approaches (CIM and CIR) are less affected by repeated measures within years (Text S4). Finally, our simulations also demonstrated a positive influence of repeatability and the number of pairs on the accuracy and precision of the estimates of assortative mating and the power to detect it (Text S4).

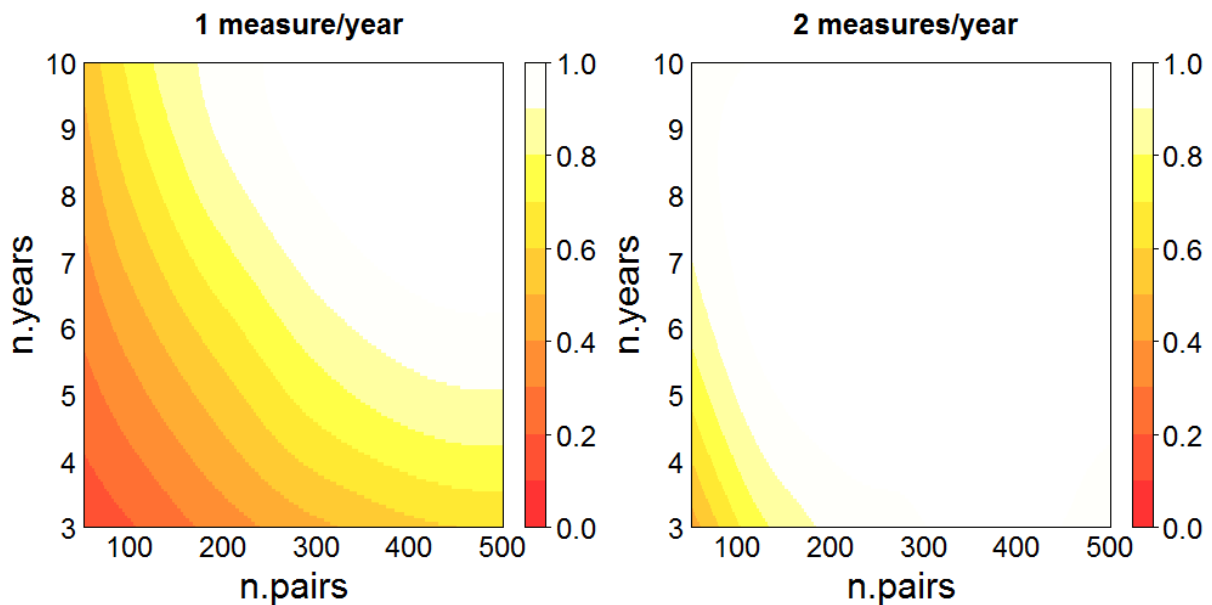


**Figure 2: Distributions (point estimate and 95% credible intervals) of types of correlations simulated and estimated for scenarios 1 (assortative mating only) and 2 (common environment only) in ‘immortal albatross’ and ‘bluish tit’ populations studied over 3 or 10 breeding seasons. The left-side of each panel depicts the simulated correlations of individual-specific values (Ind, red), common environment correlations (E, blue), and the measurement error correlations (Me, black). The variation around these estimates represents sampling error while the horizontal dashed lines represent the “true” values (red line for  $\rho_{\text{ind}}$ , blue line for  $\rho_{\text{env}}$ , and black line for  $\rho_{\text{res}}$ ). The middle part of each panel depicts the estimated phenotypic correlation (Phen, red squares), the correlation of individual means (CIM, red triangles), and the correlation of randomly drawn values (CIR, red diamonds). The right-side of each panel, represented by tip-down triangles, depicts the correlations estimated by the bivariate model: the between-pairs correlation (Pair, red), the correlation due to common environment effects (Nest, blue), and the residual correlation (Res, black). In each population, 200 unique pairs are measured every year and repeatability is 0.3.**

*Choosing the best statistical approach and sampling designs for ‘bluish tits’*

We compared the performance of all approaches to estimate assortative mating in a hypothetical tit population, representing the most realistic natural scenario in that individuals may die or change partner between time intervals. Our simulations clearly show that the individual-level approaches are inappropriate to estimate assortative mating in such populations whereas the bivariate mixed model still provides consistently unbiased estimates of assortative mating. However, the precision and power of the bivariate mixed model is generally lower compared to the other approaches, implies that it requires more data. Encouragingly, adopting a study design where repeated measures are taken on the level of the pair drastically increases statistical power. For instance, given a trait repeatability of 0.3, one

could monitor a population of 150 pairs during 3 years instead of 10 by measuring each pair twice per year (Figure 3). Note that for this sampling design, fitting a random effect designating the unique combination of pair and year in the bivariate mixed model becomes necessary (in our simulations this was done by fitting nest identity). Overall, we stress that our comparison between the approaches applies to this worked example, but that our code provided in Supplementary Text S5 will enable simulating a variety of other scenarios.



**Figure 3: Power to detect assortative mating ( $\rho_{\text{ind}}=0.3$ ) in scenario 1 using the bivariate mixed-model approach when the repeatability is 0.3, as a function of the number of years, the number of pairs monitored each year and the number of measures per year. Power is represented by different color levels ranging from red (0 to 10% power) to white (90 to 100% power) as power increases.**

## General discussion

We highlighted in this paper that there are various types of processes (e.g., correlated responses to an environmental factor common to both individuals forming a pair, and measurement error correlated across pairs) that can greatly bias the estimation of assortative mating for labile phenotypic traits. As we demonstrated mathematically, assortative mating

can be described using a variance partitioning approach analogous to the one used to by quantitative geneticists to study the covariance between multiple traits within individuals (Falconer & Mackay 1996; Lynch & Walsh 1998). We introduced different statistical approaches that can be used to estimate non-random mating for repeatable traits in wild populations, all hinging on having sufficient amounts of replication on the individual or pair levels. Because the performance of each approach depends on the characteristics of the population, we advise using the code provided in the Supplementary Text S5 to: i) choose the best statistical approach given the data already collected or ii) choose the best sampling design and the best statistical approach based on the characteristics of the population and the trait repeatability.

Importantly, the approaches described in this paper are not restricted to assortative mating for repeatedly expressed traits such as behavior, physiology or metabolism but can be applied to i) any trait showing intra-individual variability (i.e., repeatability < 1). This includes truly “fixed” traits that are measured with error and where measurement error can be correlated between two partners forming a pair, ii) other non-random mating situations where the traits of mated partners are not homologous, and iii) other mating systems (e.g. social polyandry). Indeed, the pair-level approach can be used whenever paired individuals can be repeatedly measured during their pair bond; individual-level approaches can instead be applied to estimate assortative mating as long as individuals can be identified and measured repeatedly and the pair bonds between them are known. In addition, the pair-level approach can be extended to a group-level approach to study interactions between more than two individuals, where individuals that have been measured repeatedly for a specific trait are recorded (e.g. neighbors).

We showed that the partitioning of male and female traits into a (co)variance on the between- and within-pair level can be used to approximate the correlation in individual-specific values

between partners, and hence capture “true” assortative mating for individual-level differences in (labile) phenotypic traits. An interesting corollary of this method is that it allows quantifying (if known) the extent to which environmental effects common to the pair (e.g., territory quality) or residual correlations (if unknown) drive the phenotypic resemblance of pair members. Indeed, these phenomena are likely to be common for labile traits, which respond plastically to environments often varying in time and space between and within pairs. It is nevertheless important to note that environmental factors that are conserved across repeated measures of a pair can generate a covariance which can be confounded with assortative mating using a pair-level approach. Solving this issue would require partitioning the between-pair covariance further (i.e. genetic and permanent environmental levels). Estimating the environmental source of resemblance between pair members is of relevance e.g. in the study of sexual selection where theories assume that mate selection is causing correlations in “intrinsically” determined differences between individuals (Kirkpatrick & Barton 1997). Thus, for phenotypic resemblance of partners to have evolutionary consequences, it cannot be a purely residual (within-pair) correlation. Quantification of the residual correlation across pairs may also be relevant in evolutionary quantitative genetics. This is because assortative mating estimated as the phenotypic correlation  $r$  between the parents’ traits is typically assumed to reflect the correlation between parents’ breeding values  $m$  such that  $m = h^2r$ , where  $h^2$  is the heritability of the trait (Falconer & Mackay 1996). Absence of residual (environmental) correlation between pair members is a necessary (but insufficient) condition for the equality  $m = h^2r$  to hold, and the bivariate pair-level mixed model approach introduced here would hence allow for a “first test” of this condition.

Finally, we assumed that individuals do not respond to the phenotype of their partner (i.e., no indirect effects) despite the fact that individuals forming monogamous pairs and providing biparental care likely adjust their phenotype to either resemble their partner or

specialize on different tasks (“social niche hypothesis”, Dingemanse & Araya-Ajoy 2015). For instance, the “personalities” of spouses converge in humans (Rammstedt, Spinath, Richter & Schupp 2013). Similarly, paired individuals become more behaviorally and physiologically similar with time in other species (Hile, Plummer & Striedter 2000; Ouyang, Muturi, Quetting & Hau 2013; Laubu, Dechaume-Moncharmont, Motreuil & Schweitzer 2016). Here again, a quantitative genetic approach can be used to estimate indirect effects between paired individuals (Moore, Brodie & Wolf 1997; Dingemanse & Araya-Ajoy 2015). However, such approaches do not readily allow for accounting for assortative mating, and estimates of indirect effects might thus be biased when assortative mating occurs. Disentangling indirect effects (partners become more alike after mating) from assortative mating (like mates with like) would either require phenotyping individuals before and after they form pairs, which unfortunately can be difficult to implement, or further development of statistical approaches using data of individuals mating repeatedly with different partners.

Environmental heterogeneity across unique pairs is common in wild populations where individuals forming a pair might look more alike because of effects of shared environments. We here use examples inspired by bird species with biparental care and pair stability over a breeding season, because such mating systems provide an intuitive grasp of mating dynamics and pair-specific environmental effects. However, environmental heterogeneity may well bias the phenotypic resemblance between pair members also in sexually reproducing species without biparental care. In such species, assortative mating on the phenotypic level would arise as a correlation between the phenotypes (e.g. body mass) of male and female measured at mating. Correlated plasticity of the focal trait in response to spatio-temporal heterogeneity in an environmental factor may also produce trait resemblance in pair members in such systems, thereby biasing the phenotypic correlation relative to the “true assortative mating” correlation, as detailed above. Although such biases have yet to be

demonstrated in wild populations, it likely inflates estimates of assortative mating, when not corrected for. Using the methodological tools discussed in this paper, an empirical question to evaluate now is how common shared environmental effects on the phenotypes of paired individuals actually are in natural populations.

### **Acknowledgements**

We thank Otso Ovaskainen for his valuable input regarding the multivariate network approach and we gratefully acknowledge Wolfgang Forstmeier, Mihai Valcu, and Daiping Wang for proposing the randomization of individual values as an approach to estimate assortative mating. We also thank two anonymous reviewers for their comments on this manuscript. BC was funded by University of Turku Graduate School (UTUGS), JEB by the Finnish Academy (project 289456).

### **Authors' contributions**

All authors designed the study, discussed the results and implications and commented on the manuscript at all stages. BC performed the simulations and data analysis with technical support and advice from JEB and YGA. BC, NJD and JEB drafted the manuscript.

### **Data accessibility**

This manuscript does not include any data. The code of the simulations can be found in supplementary text S5.

### **References**

- Araya-Ajoy, Y.G. & Dingemanse, N.J. (2017) Repeatability, heritability, and age-dependence in the aggressiveness reaction norms of a wild passerine bird. *Journal of Animal Ecology*, **86**, 227–238.
- Ariyomo, T.O. & Watt, P.J. (2013) Disassortative mating for boldness decreases reproductive success in the guppy. *Behavioral Ecology*, **24**, 1320–1326.
- Bell, A.M., Hankison, S.J. & Laskowski, K.L. (2009) The repeatability of behaviour: a meta-analysis. *Animal Behaviour*, **77**, 771–783.



- Bolnick, D.I. & Kirkpatrick, M. (2012) The relationship between intraspecific assortative mating and reproductive isolation between divergent populations. *Current Zoology*, **58**, 484–492.
- Bried, L., Pontier, D. & Jouventin, P. (2003) Mate fidelity in monogamous birds: a re-examination of the Procellariiformes. *Animal Behaviour*, **65**, 235–246.
- Burley, N. (1983) The meaning of assortative mating. *Ethology and Sociobiology*, **4**, 191–203.
- Butler, D.G., Cullis, B.R., Gilmour, A.R. & Gogel, B.J. (2009) ASReml-R reference manual Version 3.
- Class, B., Kluehn, E. & Brommer, J.E. (2014) Evolutionary quantitative genetics of behavioral responses to handling in a wild passerine. *Ecology and Evolution*, **4**, 427–440.
- Dingemanse, N.J., Both, C., Drent, P.J. & Tinbergen, J.M. (2004) Fitness consequences of avian personalities in a fluctuating environment. *Proceedings of the Royal Society B*, **271**, 847–852.
- Dingemanse, N.J., Dochtermann, N.A. & Nakagawa, S. (2012) Defining behavioural syndromes and the role of ‘syndrome deviation’ in understanding their evolution. *Behavioral Ecology and Sociobiology*, **66**, 1543–1548.
- Dingemanse, N.J. & Dochtermann, N.A. (2013) Quantifying individual variation in behaviour: Mixed-effect modelling approaches. *Journal of Animal Ecology*, **82**, 39–54.
- Dingemanse, N.J. & Araya-ajoy, Y.G. (2015) Interacting personalities : behavioural ecology meets quantitative genetics. *Trends in Ecology & Evolution*, **30**, 88–97.
- Falconer, D.S. & Mackay, T.F.C. (1996) *Introduction to Quantitative Genetics (4th Edition)*. Longman, Harlow, U.K.
- Gimelfarb, A. (1986) Is offspring—midparent regression affected by assortative mating of parents? *Genetical research*, **47**, 71–75.
- Glicksohn, J. & Golan, H. (2001) Personality, cognitive style and assortative mating. *Personality and Individual Differences*, **30**, 1199–1209.
- Hile, A.A.G., Plummer, T.K.T. & Striedter, G.F.G. (2000) Male vocal imitation produces call convergence during pair bonding in budgerigars, *Melopsittacus undulatus*. *Animal behaviour*, **59**, 1209–1218.
- Holtmann, B., Lagisz, M. & Nakagawa, S. (2017) Metabolic rates, and not hormone levels, are a likely mediator of between-individual differences in behaviour: A meta-analysis. *Functional Ecology*, **31**, 685–696.
- Jiang, Y., Bolnick, D.I. & Kirkpatrick, M. (2013) Assortative mating in animals. *The American naturalist*, **181**, E125–E138.

- Johannesson, K., Rolán-Alvarez, E. & Ekendahl, A. (1995) Incipient Reproductive Isolation between Two Sympatric Morphs of the Intertidal Snail *Littorina saxatilis*. *Evolution*, **49**, 1180–1190.
- Keller, M.C., Thiessen, D. & Young, R.K. (1996) Mate assortment in dating and married couples. *Personality and Individual Differences*, **21**, 217–221.
- Kirkpatrick, M. & Barton, N.H. (1997) The strength of indirect selection on female mating preferences. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 1282–1286.
- Kralj-Fišer, S., Sanguino Mostajo, G.A., Preik, O., Pekár, S. & Schneider, J.M. (2013) Assortative mating by aggressiveness type in orb weaving spiders. *Behavioral Ecology*, **24**, 824–831.
- Langerhans, R.B. & Makowicz, A.M. (2013) Sexual selection paves the road to sexual isolation during ecological speciation. *Evolutionary Ecology Research*, **15**, 633–651.
- Laubu, C., Dechaume-Moncharmont, F.-X., Motreuil, S. & Schweitzer, C. (2016) Mismatched partners that achieve postpairing behavioral similarity improve their reproductive success. *Science advances*, **2**, e1501013.
- Lewis, S.K. & Oppenheimer, V.K. (2000) Educational Assortative Mating across Marriage Markets : Non-Hispanic Whites in the United States. *Demography*, **37**, 29–40.
- Lynch, M. & Walsh, B. (1998) *Genetics and analysis of quantitative traits*. Sinauer Associates, Sunderland, MA.
- Maan, M.E. & Seehausen, O. (2011) Ecology, sexual selection and speciation. *Ecology Letters*, **14**, 591–602.
- Mare, R.D. (1991) Five Decades of Educational Assortative Mating. *American Sociological Review*, **56**, 15–32.
- Mascie-Taylor, C.G. (1987) Assortative mating in a contemporary British population. *Ann.Hum.Biol.*, **14**, 59–68.
- Mascie-Taylor, C.G. (1989) Spouse similarity for IQ and personality and convergence. *Behavior genetics*, **19**, 223–227.
- Montiglio, P.-O., Wey, T.W., Chang, A.T., Fogarty, S. & Sih, A. (2016) Multiple mating reveals complex patterns of assortative mating by personality and body size. *Journal of Animal Ecology*, **85**, 125–135.
- Moore, A.J., Brodie, E.D. & Wolf, J.B. (1997) Interacting Phenotypes and the Evolutionary Process : I . Direct and Indirect Genetic Effects of Social Interactions. *Evolution*, **51**, 1352–1362.

Ouyang, J.Q., Muturi, M., Quetting, M. & Hau, M. (2013) Small increases in corticosterone before the breeding season increase parental investment but not fitness in a wild passerine bird. *Hormones and Behavior*, **63**, 776–781.

Pérez-Figueroa, A., De Uña-Alvarez, J., Conde-Padín, P. & Rolán-Alvarez, E. (2008) Comparison of two methods for analysing the biological factors contributing to assortative mating or sexual isolation. *Evolutionary Ecology Research*, **10**, 1201–1216.

Plomin, R. & Deary, I. (2015) Genetics and intelligence differences: five special findings. *Molecular Psychiatry*, **20**, 98–108.

R Core Team (2016) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rammstedt, B., Spinath, F.M., Richter, D. & Schupp, J. (2013) Partnership longevity and personality congruence in couples. *Personality and Individual Differences*, **54**, 832–835.

Roff, D.A. (1997) *Evolutionary quantitative genetics* (D.A. Roff, Ed.). Chapman & Hall, New York, NY.

Roulin, A. (1999) Nonrandom pairing by male barn owls (*Tyto alba*) with respect to a female plumage trait. *Behavioral Ecology*, **10**, 688–695.

Seehausen, O., Alphen, J. & Witte, F. (1997) Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science*, **277**, 1808–1810.

Silventoinen, K., Kaprio, J., Lahelma, E., Viken, R.J. & Rose, R.J. (2003) Assortative mating by body height and BMI: Finnish twins and their spouses. *American Journal of Human Biology*, **15**, 620–627.

Snijders, T. & Bosker, R. (1999) *Multilevel analysis: An introduction to basic and advanced multilevel modelling*. Sage, London.

Snowberg, L.K. & Bolnick, D.I. (2012) Partitioning The Effects Of Spatial Isolation, Nest Habitat, And Individual Diet In Causing Assortative Mating Within A Population Of Threespine Stickleback. *Evolution*, **66**, 3582–3594.

Throneycroft, H. B. (1975) A Cytogenetic Study of the White-Throated Sparrow , *Zonotrichia albicollis* (Gmelin). *Evolution*, **29**, 611–621.

Wilson, A.J., Réale, D., Clements, M.N., Morrissey, M.M., Postma, E., Walling, C.A., Kruuk, L.E.B. & Nussey, D.H. (2009) An ecologist's guide to the animal model. *The Journal of animal ecology*, **79**, 13–26.

Wright, S. (1921) Systems of Mating. III. Assortative Mating Based on Somatic Resemblance. *Genetics*, **6**, 144–161.

## Supporting information

**Text S1.** Step-by-step explanation for the variance-partitioning description of assortative mating

**Text S2.** Description of the network approach used to generate assortative mating data

**Text S3.** Comparison of the different approaches in the albatross example for four scenarios

**Text S4.** Quantitative comparison of the different approaches in the “bluish tit” example: bias, imprecision, coverage and power

**Text S5.** R codes used to generate and analyze data using asreml, nlme and MCMCglmm and sommer packages and R code used for the “bluish tit” simulation