The information theory formalism unifies the detection of the patterns of sexual selection and assortative mating for both discrete and quantitative traits.

A. Carvajal-Rodríguez*

Centro de Investigación Mariña (CIM), Departamento de Bioquímica, Genética e Inmunología. Universidade de Vigo, 36310 Vigo, Spain.

*: A. Carvajal-Rodríguez. Departamento de Bioquímica, Genética e Inmunología. Universidade de Vigo, 36310 Vigo, Spain. Phone: +34 986813828

email: acraaj@uvigo.es

Keywords: quantitative traits, sexual selection, assortative mating, Kullback-Leibler divergence

Abstract

Sexual selection plays a crucial role in modern evolutionary theory, offering valuable insight into evolutionary patterns and species diversity. Recently, a comprehensive definition of sexual selection has been proposed, defining a trait as subject to sexual selection when variations in that trait are nonrandomly associated with variations in access to gametes, a limiting resource. Previous research on discrete traits demonstrated that non-random mating can be effectively quantified using the Kullback-Leibler symmetrized divergence, capturing information acquired through mating influenced by mutual mating propensities instead of random occurrences. This novel theoretical framework allows for detecting and assessing the strength of sexual selection and assortative mating.

In this study, we aim to achieve two primary objectives. Firstly, we demonstrate the seamless alignment of the previous theoretical development, rooted in information theory and mutual mating propensity, with the new sexual selection definition. Secondly, we extend the theory to encompass quantitative traits. Our findings reveal that sexual selection and assortative mating can be quantified effectively for quantitative traits by measuring the information gain concerning the random mating pattern. Specifically, the gain in information within the sexual selection pattern can be expressed as a function of the sexual selection intensity index and the ratio of variances. Additionally, the measure capturing the underlying information of assortative mating is a function of the square of the correlation coefficient, taking values within the non-negative real number set $[0, +\infty)$.

Introduction

Sexual selection is a fundamental concept in modern evolutionary theory, offering valuable insights into various evolutionary patterns and the diversity of species. Originally proposed by Darwin in (1871), sexual selection describes the competition among individuals of one sex to secure mates from the other sex. However, the concept of sexual selection has been a subject of controversy since its inception (Andersson 1994; Prum 2012; Parker 2014; Parker and Pizzari 2015; Roughgarden et al. 2015; Alonzo and Servedio 2019; Lehtonen 2022; Vries and Lehtonen 2023). There is ongoing debate surrounding its precise definition (Fitze and Galliard, 2011), and its role as a key component of modern evolutionary biology has even faced challenges (Parker and Pizzari, 2015; Roughgarden et al., 2006; but see Shuker, 2010).

Recently, Shuker and Kvarnemo (2021) put forth a comprehensive definition of sexual selection, characterizing it as "any selection that arises from fitness differences associated with nonrandom success in the competition for access to gametes for

fertilization". In essence, a trait falls under sexual selection when variations in that trait are nonrandomly associated with variations in access to gametes, which serve as a limiting resource. From an evolutionary perspective, this process of sexual selection generates a discernible pattern of change in gene frequency, influencing the frequencies of corresponding genotypes and phenotypes. It is important to emphasize the use of the term "pattern of change" when discussing sexual selection, as it helps to clarify whether we are referring to the dynamic process or to the recognizable pattern in genetic or phenotypic changes.

Darwin identified two primary biological mechanisms of sexual selection: mate competition and mate choice (Ng et al. 2019 and references therein). The mate choice process refers to the effects of certain traits expressed in one sex that leads to the non-random allocation of reproductive effort with members of the opposite sex (Edward 2015). Mate choice may produce a mating pattern where similar phenotypes or genotypes mate with one another more or less frequently than would be expected under a random mating pattern i.e. it produces a deviation from random mating in mated individuals (Rolán-Alvarez and Caballero 2000; Carvajal-Rodríguez 2018b). When the observed mating pattern between similar phenotypes is more frequent that the expected by chance, we refer to positive assortative mating and when it is less frequent we refer to negative assortative mating. Sometimes in the literature the positive assortative mating is just referred as assortative mating and the negative assortative mating as disassortative mating.

Mate choice when it produces positive assortative mating can have important evolutionary effects as it alters the variance of quantitative traits (Fisher 1918), can generate sexual selection (de Cara et al. 2008; Carvajal-Rodríguez 2018b) and under certain conditions, may facilitate local adaptation and speciation even in the presence of

gene flow (Kirkpatrick and Ravigné 2002; Coyne and Orr 2004; Gavrilets 2004; Rettelbach et al. 2013; Cotto and Servedio 2017; Sachdeva and Barton 2017; Fernández-Meirama et al. 2022; Yukilevich 2023).

When examining phenotypes involving discrete traits, assortative mating can be identified through mating data using the pair sexual isolation statistic (Rolán-Alvarez and Caballero 2000; Carvajal-Rodríguez and Rolan-Alvarez 2006). For quantitative traits, assessing assortative mating can be achieved by computing the Pearson correlation coefficient within mated pairs (Jiang et al. 2013). However, it is important to note that this measure comes with known problems and limitations (Clancey et al. 2022; Fernández-Meirama et al. 2022).

In a previous study (Carvajal-Rodríguez 2018b), it was demonstrated that non-random mating can be effectively quantified using the Kullback-Leibler symmetrized divergence. This measure captures the information acquired when mating is influenced by mutual mating propensities, rather than occurring randomly. By employing this novel theoretical framework, it becomes possible to detect and assess the strength of sexual selection and assortative mating (Carvajal-Rodríguez 2020; Estévez et al. 2020; Lau et al. 2021). However, it should be noted that the original framework was designed for discrete classes, and in this current work, we aim to extend its applicability to continuous cases.

The objective of this work is twofold. Firstly, we aim to demonstrate that the previous theoretical development, which is rooted in information theory and the concept of mutual mating propensity (Carvajal-Rodríguez 2018b, 2020), seamlessly aligns with the new sexual selection definition put forth by Shuker and Kvarnemo (2021). In this context, sexual selection emerges as a consequence of variations in the mutual mating propensity, without requiring specific assumptions about the genetic basis of the

involved traits. Secondly, we intend to illustrate that, for quantitative traits, the pattern of sexual selection, quantified as the difference between the phenotypic distribution of mated individuals and the population distribution, can be expressed as a function of the sexual selection intensity index (sensu Ng et al. 2019) and the ratio of variances. Additionally, we emphasize that the measure capturing the underlying information of assortative mating is not the correlation itself, but rather a function of the square of the correlation coefficient, that takes values within the non-negative real number set [0, +∞).

Fitness and sexual selection

Fitness, defined as an organism's ability or propensity to survive and reproduce relative to others in the same population (Hedrick 2005; Hansen 2018) drives natural selection by leading to changes in gene frequencies (Hartl 2020). Identifying natural selection relies on recognizing the patterns of change it produces in frequencies.

Consider a population with various types, each representing a set of individuals with specific properties (traits, genotypes, etc.). Let N_i be the number of individuals of type *i*, and suppose a selection episode changes this number to N'_i. The absolute fitness, or growth rate, of type *i* during this episode is denoted as $W_i = N'_i/N_i$. To compare different types' fitness fairly, we compute the relative fitness, w_i, as the ratio of their absolute fitness to the mean population fitness. Thus, for a given type *i* with frequency p_i, the new frequency p'_i after a selection episode will be given by (Bürger 2000; Hansen 2018)

$$p_i' = p_i w_i \quad (1).$$

Recently, a simple theoretical framework, similar to the model presented in (1), has been developed to describe the influence of mating competition and mate choice on the emergence of sexual selection and assortative mating patterns (Carvajal-Rodríguez

2018b, 2019, 2020). This theoretical framework has enabled the development of new statistical tests for detecting sexual selection and assortative mating. Moreover, by linking variation in mating ability (i.e., gamete access for fertilization) to the resulting frequency patterns, this framework facilitates the construction of predictive models and supports multi-model inference based on observed data patterns (Carvajal-Rodríguez 2020).

In the following sections, I will briefly review and elucidate this theoretical framework, demonstrating its compatibility with Shuker and Kvarnemo's (SK hereafter) definition of sexual selection ("any selection that arises from fitness differences associated with nonrandom success in the competition for access to gametes for fertilization"; Shuker and Kvarnemo 2021).

Fitness for mating and fertilization

We will focus on the change in frequencies due to fitness differences associated with access to gametes for fertilization. First, we must specify the concept of individual fitness for mating and fertilization. Secondly, we will express for the discrete case, the SK definition of sexual selection in terms of information and then, extend the definition to the continuous case.

Let a female character X with a total of k_1 discrete possible trait values and a male character Y with k_2 discrete values. Characters X and Y can be the same or not. We define M_{ij} as the mutual ability or propensity to access gametes for fertilization between a female with value i for X and a male with value j for Y. Let p_{1i} be the population frequency of type i females and p_{2j} the frequency of type j males so that $q_{ij} = p_{1i}p_{2j}$ is the probability of random pairing of gametes from type i females with gametes from type j males for fertilization. The mean propensity is

$$M=\sum_{i=1}^{k1}\sum_{j=1}^{k2} q_{ij} M_{ij}$$

and the relative mutual propensity $m_{ij}=\frac{M_{ij}}{M}$.

The probability q'_{ij} of fertilization between any pair having such a combination of traits is (Carvajal-Rodríguez 2018b)

$$q'_{ij}=q_{ij} m_{ij} \quad (2).$$

Equation (2) represents the expected frequencies of fertilization between types *i* and *j* as a function of the expected frequencies per random gamete encounter and the relative values of mutual propensity. The equation clearly resembles equation (1) for the change in frequencies because of selection. In fact, the absolute value M_{ij} can be expressed in a similar way to absolute fitness i.e., as the ratio of the number N'_{ij} of fertile matings between types *i* and *j* after selection over the expected number N_{ij} if there were no selection (random fertilization).

$$M_{ij}=\frac{N'_{ij}}{N_{ij}}.$$

Our focus is on identifying the individual propensity to access gametes for fertilization, rather than the mutual propensity of pairs. We refer to this individual propensity as the fitness for mating and fertilization. In this context, we define the absolute marginal propensity M_{Fi} of the female of type *i* as

$$M_{Fi}=\sum_{j=1}^{k2} p_{2j} M_{ij}.$$

The relative marginal propensity m_{Fi} is obtained by dividing by the mean propensity *M*

$$m_{Fi}=\frac{M_{Fi}}{M} \qquad (3).$$

Similarly, the relative marginal propensity of males of type *j* is

$$m_{Mj}=\sum_{i=1}^{k1} p_{1i} m_{ij}=\frac{M_{Mj}}{M}.$$

Thus, the relative marginal propensity of a type is the average relative ability of individuals of that type to access gametes for fertilization, i.e., the relative fitness for mating (and fertilization).

While different mutual propensities shape mate choice, different marginal propensities within a sex will produce a pattern of sexual selection (see Figure1 in Carvajal-Rodríguez 2020). This pattern arises due to the unequal ability to access the gametes of the opposite sex for fertilization. Consequently, this sexual selection aligns perfectly with SK's definition.

Importantly, the concepts of mutual and marginal propensities, as defined, do not presuppose any specific genetic basis underlying the phenotypes involved in mating. Furthermore, they do not imply any particular mating preference function.

Having formalized the concepts of mutual and marginal propensities it is then possible to model mate competition and choice and the patterns of sexual selection and assortative mating they produce (see Carvajal-Rodríguez 2018b, 2020). The pattern of sexual selection is caused mainly by two biological mechanisms: mate competition and mate choice (Lewontin et al. 1968; Arnold and Wade 1984; Endler 1986; Rolán-Alvarez and Caballero 2000; Rolán-Alvarez et al. 2015).

The process of mate competition refers, in a broad sense, to access to mating through courtship, intrasexual aggression and/or competition for limited reproductive resources (Andersson 1994; Kokko et al. 2012; Wacker and Amundsen 2014). Mate competition

may generate a pattern of sexual selection (a change in frequencies of the trait under study) in the sex that competes.

The process of mate choice is the behavioral propensity to mate with individuals having certain traits and occurs whenever the effects of traits expressed in one sex leads to non-random allocation of reproductive investment with members of the opposite sex (Edward 2015 and references therein). Choice may be mediated by phenotypic (sensorial or behavioral) properties that affect the propensity of individuals to mate with certain phenotypes (Jennions and Petrie 1997).The observed pattern driven by mate choice can be a change in trait frequency in the other sex (sexual selection) and/or assortative mating.

The model in (2) facilitates the development of new statistical tests, based on information theory, to detect patterns of sexual selection and assortative mating (Carvajal-Rodríguez 2018b). Moreover, it enables to formalize the necessary and sufficient conditions for random mating in terms of mating propensities. Once these conditions are defined, we can generate different mating and fertilization fitness scenarios that produce the intended patterns of sexual selection and assortative mating, allowing the application of powerful model inference techniques (Carvajal-Rodríguez 2020).

The formalization of the mate choice and competition processes from the model presented in (2) have been carried out only for discrete traits and here we extend the mathematical framework to the continuous case. As a result, we will see that the pattern of sexual selection can be expressed as a function of the sexual selection intensity index and the ratio of variances, and the pattern of assortative mating is captured by a function of the square of the Pearson correlation instead of the correlation per se.

Sexual selection and assortative mating in quantitative traits

Consider two continuous traits in a population, X in females and Y in males, and let be the mating probability distribution Q'(x,y) with probability density $q'(x,y)$ representing the mating (and fertilization) probability for pairs with values in the infinitesimal interval ([x,x+dx],[y,y+dy]). Alternatively consider the probability distribution $Q(x,y)$ of random mating from the product of densities $q(x,y)=f(x)g(y)$ of X and Y, respectively.

Note that (2) still holds so that $q'(x,y)$ depends on the density $q(x,y)$ and the infinitesimal mutual mating propensity m_{xy} which is the relative mutual propensity to access gametes for fertilization between a female with value in the infinitesimal interval [x,x+dx] and a male with value in the infinitesimal interval [y,y+dy].

In earlier work, Carvajal-Rodríguez (2018b) showed for discrete traits, that non-random mating can be characterized as the divergence between the mating distribution, Q', derived from mutual mating propensities, and the distribution of random mating, Q. For the continuous case we have that the divergence between the two distributions Q' and Q is given by

$$J=\int_{-a}^{+a}\int_{-a}^{+a}(q'(x,y)-q(x,y))\log\frac{q'(x,y)}{q(x,y)}dxdy \qquad (4),$$

where $q(x,y)=f(x)g(y)$.

For simplicity and without lost of generality we assume that the range is the same for X and Y so that, $X,Y\in[-a,a]$.

The distribution function of X in the population

$$F(x)=\int_{-a}^{x}f(u)du$$

and Y

$$G(y)=\int_{-a}^{y} g(v)\,dv,$$

where $f(u)$ and $g(v)$ are the respective probability densities.

From the probability density $q'(x,y)$, the marginal densities for X and Y in the matings are

$$f_1(x)=\int_{-a}^{+a} q'(x,y)\,dy$$

$$f_2(y)=\int_{-a}^{+a} q'(x,y)\,dx$$

Sexual selection

The difference $q'(x,y)$-$q(x,y)$ in (4) can be partitioned as follows

$$q'(x,y)-q(x,y)=f_1(x)f_2(y)-q(x,y)+q'(x,y)-f_1(x)f_2(y),$$

similarly, the logarithm

$$\log\frac{q'(x,y)}{q(x,y)}=\log\left(\frac{f_1(x)f_2(y)}{q(x,y)}\frac{q'(x,y)}{f_1(x)f_2(y)}\right)=\log\frac{f_1(x)f_2(y)}{q(x,y)}+\log\frac{q'(x,y)}{f_1(x)f_2(y)}$$

Recall that the pattern of sexual selection is quantified by comparing the expected mate types resulting from random mating among mated individuals with those calculated using the population frequencies (Hartl and Clark 1997; Rolán-Alvarez and Caballero 2000). In the continuous scenario, this is formalized by the Jeffreys divergence

$$J_{PSS}=\int_{-a}^{+a}\int_{-a}^{+a}(f_1(x)f_2(y)-q(x,y))\log\frac{f_1(x)f_2(y)}{q(x,y)}\,dxdy \qquad (5)$$

and because

$$\int_{-a}^{+a} f_1(x)\,dx = \int_{-a}^{+a} f_2(y)\,dy = 1 \quad \text{and} \quad \int_{-a}^{+a} f(x)\,dx = \int_{-a}^{+a} g(y)\,dy = 1$$

we can separate by sex

$$J_{PSS} = J_{PS1} + J_{PS2} \quad \text{where}$$

$$J_{PS1} = \int_{-a}^{+a} (f_1(x) - f(x)) \log \frac{f_1(x)}{f(x)}\,dx$$

$$J_{PS2} = \int_{-a}^{+a} (f_2(y) - g(y)) \log \frac{f_2(y)}{g(y)}\,dy.$$

Therefore, similar to the approach applied in the discrete case (Carvajal-Rodríguez 2018b), J_{PSS} quantifies sexual selection for quantitative traits as the change in phenotype distribution resulting from differential mating and fertilization. The information captured in J_{PSS} is the sum of the sexual selection within each sex.

Normally distributed traits

Let assume that $q'(x,y)$ is a bivariate normal density and traits X and Y are normally distributed. Then, the marginal density function for the trait in females within matings is $f_1(x) \sim N(\mu_1,\sigma_1^2)$, and in the population is $f(x) \sim N(\mu_x,\sigma_x^2)$. The marginal density function for the trait in males within matings is $f_2(y) \sim N(\mu_2,\sigma_2^2)$ and in the population is $g(y) \sim N(\mu_y,\sigma_y^2)$ which are normally distributed and J_{PS1} and J_{PS2} can be expressed in closed form as (Kullback 1997; Belov and Armstrong 2011)

$$J_{PS1} = \frac{1}{2}\left(\frac{\Phi_1^2+1}{\Phi_1} + \frac{\Phi_1+1}{\Phi_1}\frac{(\mu_1-\mu_x)^2}{\sigma_x^2} - 2\right) \quad \text{where} \quad \Phi_1 = \sigma_1^2/\sigma_x^2.$$

$$J_{PS2} = \frac{1}{2}\left(\frac{\Phi_2^2+1}{\Phi_2} + \frac{\Phi_2+1}{\Phi_2}\frac{(\mu_2-\mu_y)^2}{\sigma_y^2} - 2\right) \quad \text{where} \quad \Phi_2 = \sigma_2^2/\sigma_y^2.$$

with nJ_{PS1} (nJ_{PS2}) asymptotically distributed as χ^2 with 2 degrees of freedom under the null hypothesis that the females (males) within matings are distributed with mean μ_x (μ_y) and variance σ_x^2 (σ_y^2) where n is the sample size of matings.

Note that in the above expressions, the term for the difference of the means is the square of the standardized sexual selection intensity (SSI) index (Erlandsson and Rolán-Alvarez 1998; Ng et al. 2019; Caballero 2020) so that

$$J_{PS1}=\frac{1}{2}\left(\frac{\Phi_1{}^2+1}{\Phi_1}+\frac{\Phi_1+1}{\Phi_1}SSI_1{}^2-2\right)$$

$$J_{PS2}=\frac{1}{2}\left(\frac{\Phi_2{}^2+1}{\Phi_2}+\frac{\Phi_2+1}{\Phi_2}SSI_2{}^2-2\right) \qquad (6)$$

where SSI_1 is the SSI index for females and SSI_2 for males.

If we want to test only the mean value of the female trait X in the mating sample, we set Φ_1=1 so that

$$J_{PS1\Phi_1=1}=\frac{(\mu_1-\mu_x)^2}{\sigma_1{}^2}$$

and with a sample size of n matings we have $(nJ_{PS1\Phi=1})^{0.5} \sim t_{n-1}$. Similarly, for the males we have $(nJ_{PS2\Phi=1})^{0.5} \sim t_{n-1}$ with

$$J_{PS2\Phi_2=1}=\frac{(\mu_2-\mu_y)^2}{\sigma_2{}^2}.$$

For large sample sizes both types of tests (χ^2 or t) work similar, for a smaller sample size the χ^2 test might be slightly liberal and the t-test should be preferred.

Assortative mating

The assortative mating pattern corresponds to the deviation from random mating distribution measured in mated individuals i.e. the divergence

$$J_{PSI}=\int_{-a}^{+a}\int_{-a}^{+a}(q'(x,y)-f_1(x)f_2(y))\log\frac{q'(x,y)}{f_1(x)f_2(y)}dxdy \qquad (7).$$

Normally distributed traits

As before, suppose that $q'(x,y)$ is a bivariate normal density and that traits X and Y are normally distributed. The divergence for assortative mating in (7) is the Jeffreys (or Kullback-Leibler symmetrized) divergence between the joint distribution and the product of the marginals and serves as a measure of the relationship between X and Y that can be expressed as a function of the correlation coefficient (see Appendix A.1 and Kullback 1997) so,

$$J_{PSI}=\int_{-a}^{+a}\int_{-a}^{+a}(q'(x,y)-f_1(x)f_2(y))\log\frac{q'(x,y)}{f_1(x)f_2(y)}dxdy=\frac{\rho^2}{1-\rho^2} \qquad (8).$$

Thus, as long as the quantitative traits we are measuring are normally distributed, the natural measure of assortative mating is a function of the square of correlation coefficient that takes values within the set of nonnegative real numbers $[0, +\infty)$.

The statistical significance of $J_{PSI}>0$ can be tested since nJ_{PSI} is asymptotically distributed as χ^2 with 1 degree of freedom (Kullback 1997). To distinguish between positive and negative assortative mating, we only need to consider the sign of the Pearson correlation coefficient ρ.

Finally, the total divergence in (4) can be expressed as the sum of sexual selection (5) and assortative mating (7) divergences plus a factor E

$$J=J_{PS1}+J_{PS2}+J_{PSI}+E$$

where (see Appendix A.2)

$$E=\int_{-a}^{+a}\int_{-a}^{+a}(f_1(x)f_2(y)-q(x,y))\log\frac{q'(x,y)}{f_1(x)f_2(y)}dxdy\neq 0\Leftrightarrow J_{PSS}\neq 0\wedge J_{PSI}\neq 0.$$

So the E component only appears when both sexual selection ($J_{PSS} \neq 0$) and assortative mating ($J_{PSI} \neq 0$) are present.

Simulations

Detection of sexual selection

We will utilize the program MateSim (Carvajal-Rodríguez 2018a) to simulate the mating distribution for quantitative traits using a logistic model (Xie et al. 2015). This model represents an open-ended preference function, capable of generating a pattern of sexual selection in both females and males.

The simulation procedure consists of two main steps. First a population of size N is generated, comprising $N/2$ females and $N/2$ males. The individuals are randomly sampled from a standard normal distribution with a mean of 0 and a standard deviation of 1, $Z(0,1)$. Second, a sample of n ($n<N$) matings is obtained as the individuals interact to form pairs for mating. The mating probability of a male j with phenotype Z_j of mating with any female F is evaluated as follows:

$$P(F,j)=\exp(-1+\alpha Z_j)/[1+\exp(-1+\alpha Z_j)]$$

and the same applies for a female i with phenotype Z_i to mate with any male M.

$$P(i,M)=\exp(-1+\alpha Z_i)/[1+\exp(-1+\alpha Z_i)]$$

Therefore, the mating probability after encounter for a pair $i \times j$ will be

$$p_{ij}=P(i,M)\times P(F,j).$$

The parameter α (≥0) accounts for the strength of sexual selection, if α=0 there is random mating, the higher the value of α, the greater the strength of sexual selection.

The tests in (6) are able of distinguishing between random mating and sexual selection for different sample sizes and sexual selection intensities. Not surprisingly, the greater the sexual selection and the size of the sample, the greater the power of detection and the value of the information recovered by the J_{PS1} and J_{PS2} indexes (Table 1). For a large sample size, n=500, the correlation between the strength of sexual selection α and the J indexes was 0.99. For smaller sample size, n=50, the correlation was 0.97, indicating that J_{PS1} and J_{PS2} indices are good proxies of the strength of sexual selection.

Table 1. Mating under a logistic model. The parameter α accounts for the intensity of sexual selection. A value of $\alpha = 0$ corresponds to random mating. Mating is with replacement. N is the population size and n the mating sample size. The quantitative trait distribution is $Z(0,1)$. Number of runs: 10,000. % sig: The % of tests with p-value ≤0.05. The values in parentheses correspond to the average value of the statistic for the total number of significant runs.

N	***n***	***α***	**% sig. (J_{PS1})**	**%sig. (J_{PS2})**
10^4	500	0	5.3 (0.016)	4.3 (0.016)
10^4	500	0.1	28.5.0 (0.019)	30.0 (0.019)
10^4	500	0.25	96.6 (0.039)	96.1 (0.038)
10^4	500	0.5	100 (0.126)	100 (0.128)
10^4	500	1	100 (0.40)	100 (0.41)
10^4	500	1.5	100 (0.710)	100 (0.701)
10^4	500	2	100 (0.981)	100 (0.985)
500	50	0	5.0 (0.146)	5.1 (0.141)
500	50	0.1	9.5 (0.155)	8.6 (0.15)
500	50	0.25	25.9 (0.168)	28.2 (0.179)
500	50	0.5	64.3 (0.215)	66.3 (0.219)
500	50	1	99.9 (0.586)	99.5 (0.476)
500	50	1.5	100 (0.845)	100 (0.918)
500	50	2	100 (1.107)	100 (1.044)

Detection of assortative mating: J_{PSI} versus correlation

We will utilize the program MateSim (Carvajal-Rodríguez 2018a) to simulate the mating distribution for quantitative traits using a FND Gaussian function (Carvajal-Rodríguez and Rolán-Alvarez 2014). The FND function with parameter C>0 generate positive assortative mating patterns that can be measured by the Pearson correlation coefficient ρ and also by J_{PSI} as given in (8).

The simulation procedure is as before but the preference is evaluated following the FND function

$$FND = e^{-\left(\frac{C^2 D^2}{s^2 D^2_{max}}\right)}.$$

If C=0 the mating is random, if C>0 there is mate choice that generates positive assortative mating. The parameter D represents the phenotypic difference between male and female in the pair. If the D is zero the preference is the maximum. The parameter s is the tolerance to mate with a partner that does not correspond to the preference optimum. i.e. s= 0.08 implies eight times more tolerance for non-optimal mating than s=0.01. D_{max} is the maximum phenotypic difference in the population between any two individuals, regardless of sex. For a phenotypic difference D/D_{max} between a pair, the strength of the choice depends on the ratio between the choice and tolerance parameters, C/s.

For a sample of n matings, the significance of ρ can be computed as $\frac{\rho\sqrt{(n-2)}}{\sqrt{(1-\rho^2)}}$ that follows a t-distribution with n-2 degrees of freedom. We have already seen that nJ_{PSI} is asymptotically distributed as χ^2 with 1 degree of freedom.

In Table 2 both types of tests are performed for mating distributions generated under different choice and tolerance values. When the population and sample size are large, 10,000 and 500, respectively, from choice values of 0.2, both the correlation and the J_{PSI} have greater than 97% power to detect assortative mating. The difference is that the correlation value already reaches its maximum from choice values of 0.4.

When the population and sample sizes are smaller, specifically 500 and 50, respectively, we observe a few noteworthy points. Firstly, it is not surprising that J_{PSI} tends to be somewhat liberal in these cases, as the test nJ_{PSI} is only asymptotically distributed as χ^2, and its behavior improves with higher sample sizes. Secondly, both methods demonstrate approximately 40% power when the choice is 0.2 and the tolerance is 0.08. However, they exhibit almost 100% power in other cases, such as when the tolerance is 0.01 or the choice is from 0.4 and beyond. The difference between ρ and J_{PSI} is again the saturation of the value of ρ as the strength of choice increases.

Table 2. One sex is choosy. Choice implies preference for similarity. Mating is with replacement. *N* is the population size and *n* the sample size of matings. The quantitative trait distribution is *Z*(0,1). Number of runs: 10,000. % sig: The % of tests with *p*-value ≤0.05. The values in parentheses correspond to the average value of the statistic for the total number of significant runs.

N	***n***	***Choice***	***s***	**% sig. (ρ)**	**%sig. ($\rho^2/(1-\rho^2)$)**
10^4	500	0	--	5.1 (0.01)	5.2 (0.01)
10^4	500	0.2	0.08	97.9 (0.18)	97.9 (0.03)
10^4	500	0.4	0.08	100 (0.55)	100 (0.43)
10^4	500	0.8	0.08	100 (0.85)	100 (2.60)
10^4	500	1	0.08	100 (0.89)	100 (3.72)
10^4	500	0.2	0.01	100 (0.95)	100 (9.3)
10^4	500	0.4	0.01	100 (0.99)	100 (59.0)
10^4	500	0.8	0.01	100 (1.0)	100 (194.0)

10^4	500	1	0.01	100 (1.0)	100 (413.4)
500	50	0	--	5.1 (0.03)	6.2 (0.12)
500	50	0.2	0.08	38.6 (0.37)	41.7 (0.16)
500	50	0.4	0.08	99.95 (0.61)	99.97 (0.64)
500	50	0.8	0.08	100 (0.90)	100 (4.59)
500	50	1	0.08	100 (0.93)	100 (6.68)
500	50	0.2	0.01	100 (0.96)	100 (13.54)
500	50	0.4	0.01	100 (0.99)	100 (56.68)
500	50	0.8	0.01	100 (1.0)	100 (183.08)
500	50	1	0.01	100 (1.0)	100 (403.27)

The greater the sample size and the choice strength, the greater the power to detect assortative mating. In Figure 1 the correlation between the choice strength (*C*/*s*) and both statistics can be appreciated. For the ρ statistic the correlation with choice strength is 0.6 for both large and small sample size, for J_{PSI} the correlation with the choice strength is 0.97.

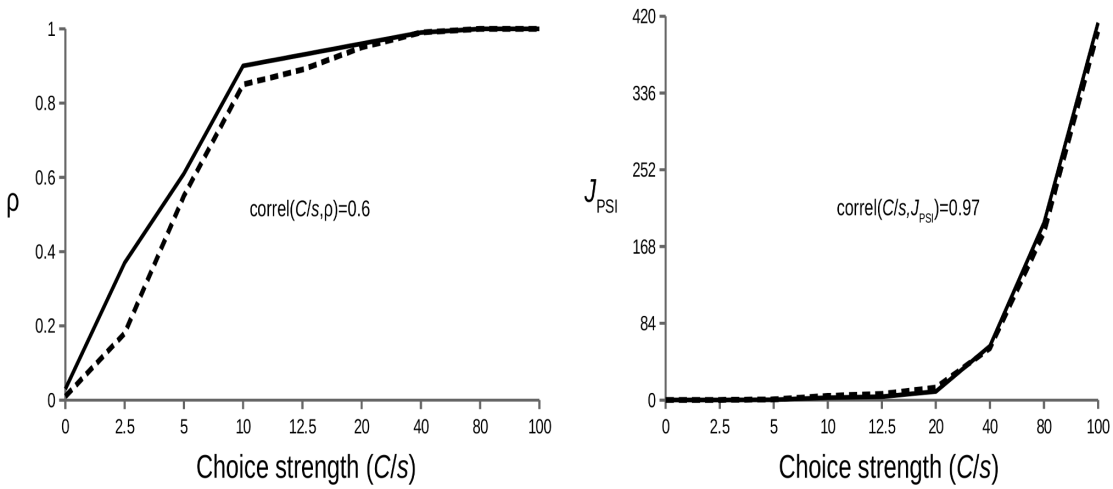

Figure 1. Continuous line *N*=10,000, *n*=500. Dashed line *N*=500, *n*=50.

The results were almost identical (Appendix A.3 Table S1) when we consider a bias for the mating preference, i.e. the choosy males preferring females 1.5 times bigger than themselves (Lau et al. 2021).

We were also interested in evaluating the performance of the J_{PSI} index across different functions and mating scenarios. To achieve this, we employed the logistic function with $\alpha = 2$ (which implies sexual selection without mate choice) with mating involving replacement, as well as individual and mass mating without replacement (see Appendix A.3 Table S2). In the case of large population and sample sizes, both ρ and J_{PSI} exhibited the expected false positive rate of 5%. However, with lower sample sizes, J_{PSI} showed a slightly liberal false positive rate of 6%.

We also simulate a mate choice scenario for a bimodal population where two more or less extreme phenotypes coexist (Appendix A.3 Table S3). The power for the detection of assortative mating was similar to the unimodal case and the correlation of the statistics with the strength of choice was greater the larger the sample size, being about 0.58-0.65 for ρ and 0.95-0.97 for J_{PSI} (Figure 2).

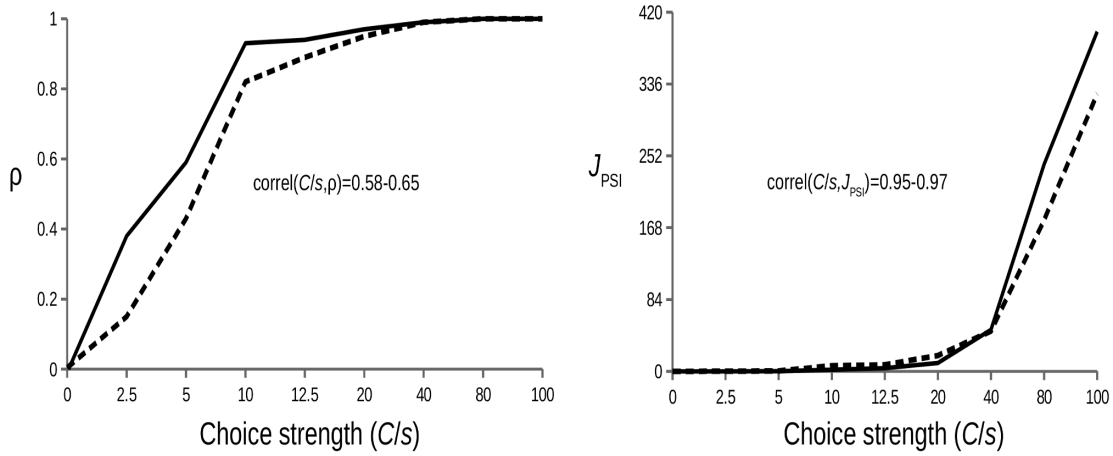

Figure 2. Bimodal population with mixture proportion π=0.5, μ_2-μ_1=3 and σ_1+σ_2 = 2.

Lastly, we examined how being choosy in one or both sexes affects the ρ and J_{PSI} indexes and found no differences (compare Table S4 in Appendix A.3 with Table 1).

5. Example of application

Echinolittorina malaccana is a marine intertidal snail commonly found on rocky shores throughout the Indo-West Pacific region (Reid 2007). The primary mechanism of mate choice in this species appears to be male preference for females slightly larger than themselves. Assortative mating for size, measured by pairwise correlation, has been estimated to be around 0.5 (Ng et al. 2019; López-Cortegano et al. 2020). In a previous study, size (shell length, a continuous trait) has been transformed into a qualitative trait (discrete size classes) enabling the use of multimodel inference techniques. These analyses provided deeper insights into mate choice and confirmed that the most suitable model involves mate choice with bias, resulting in both assortative mating and sexual selection for size (Lau et al. 2021).

We will use the same onshore data from mating pairs of *E. malaccana* used in (Lau et al. 2021) and available from (Ng et al. 2019; Rolan-Alvarez et al. 2021) to apply the new indices J_{PSS} and J_{PSI} directly to the shell length trait (Table 3).

Table 3. Analysis of *Echinolittorina malaccana* data.

Test	***N***	***n***	J_{PS1} **(pval)**	J_{PS2} **(pval)**	**ρ (pval)**	J_{PSI} **(pval)**
Chi^2	876	228	0.11 ($<10^{-5}$)	0.03 (0.031)	0.57 ($<10^{-5}$)	0.49 ($<10^{-5}$)

Data from Ng et al 2019.

Sexual selection was observed in both sexes, although it appeared to be weaker in males. This finding aligns with previous research, indicating that sexual selection in *E. malaccana* females is likely driven by male preference for slightly larger females, while in males, it may result from male-male competition (Ng et al. 2019; Lau et al. 2021). As for assortative mating, both the Pearson correlation coefficient and the new J_{PSI} index supported and validated previous findings.

6. Discussion

There is a vast amount of literature exploring mate choice, competition, assortative mating, and sexual selection as influential forces in the evolution of traits that may not always be directly favoured by natural selection (Darwin 1871; Andersson 1994; reviewed in Lindsay et al. 2019). These concepts are traditionally grouped under the term 'theory of sexual selection'. Sexual selection theory stands as one of the most active areas of evolutionary research, engaging diverse fields and inference methodologies, given its relevance in population genetics, evolutionary ecology, animal behavior, sociology, and psychology. Two recent important reviews were published on the concept and definition of sexual selection (Alonzo and Servedio 2019; Shuker and Kvarnemo 2021). Alonzo and Servedio discuss the challenges in achieving a consensus definition, while Shuker and Kvarnemo propose a new definition for sexual selection.

In this study, we demonstrate that the SK definition of sexual selection can be effectively applied within a modeling framework that utilizes information theory to describe mate choice and competition processes, leading to patterns of sexual selection and assortative mating. Under the information theory formalism, these patterns are described in terms of Jeffreys divergence.

Importantly, this novel framework, initially developed for discrete traits, is general and doesn't assume a priori whether the choice is exercised by males, females, or both, nor does it make assumptions about the sex that competes. Furthermore, it is agnostic concerning the genetic basis or evolutionary mechanisms underlying complex mate choice and competition processes. The informational framework starts with a concept of mutual mating fitness or mutual mating propensity (sensu Carvajal-Rodríguez 2018b, 2020), where the marginal values correspond to individual mating fitness. These mating fitnesses allow us to measure the pattern of sexual selection according to the SK definition.

In addition, we have extended the informational framework to encompass quantitative traits. For both discrete and quantitative traits, sexual selection is detected by comparing the distribution of mating phenotypes with that of the overall population. Similarly, assortative mating is detected by comparing the observed distribution in mated individuals with the expected distribution under random mating of those same individuals.

When the traits are normally distributed, the sexual selection indexes J_{PS1} (females) and J_{PS2} (males) can be expressed as a function of the standardized sexual selection intensity index and the ratio of sample and population variances. Simulations have shown that J_{PS} indexes have a correlation of 0.99 with the strength of sexual selection.

Furthermore, the assortative mating index J_{PSI} possesses as much statistical power as ρ, with the added advantage of being highly correlated with the strength of mate preference, which is not generally true for the correlation coefficient (Clancey et al. 2022; Fernández-Meirama et al. 2022).

Finally, we applied the quantitative trait information indices developed in this work to a previously published dataset on size assortative mating and sexual selection (Ng et al. 2019; Lau et al. 2021). We confirmed previous findings and validate the quantitative approach for the new indices of sexual selection and assortative mating.

In conclusion, the application of information theory to describe patterns of sexual selection and assortative mating (Carvajal-Rodríguez 2018b, 2019, 2020), has paved the way for extending these developments to quantitative traits. This study demonstrates that sexual selection and assortative mating can be effectively quantified as information gain concerning the random mating pattern. Just like the previous developments for discrete traits, the results obtained with quantitative traits indicate that employing information theory to describe mating processes and patterns holds great promise as a valuable approach to understanding both the causes and effects of sexual selection.

Acknowledgements

Dedicated to my sister Paci. This work was supported by Xunta de Galicia (Grupo de Referencia Competitiva, ED431C 2020/05), Ministerio de Economía y Competitividad (PID2022-137935NB-I00) and by Fondos FEDER (“Unha maneira de facer Europa”).

Appendix

A.1. Assortative mating with normally distributed traits

The general formulation is

$$J_{PSI}=\int_{-a}^{+a}\int_{-a}^{+a}(q'(x,y)-r'(x,y))\log\frac{q'(x,y)}{r'(x,y)}dxdy$$

where $r'(x,y)= f_1(x)f_2(y)$

If q' is bivariate normal and f_1 and f_2 are normally distributed we have

$$q'(x,y)=\frac{1}{|2\Pi\Sigma_1|^{0.5}}\exp[-\frac{1}{2}(z-\mu_1)'\Sigma_1^{-1}(z-\mu_1)]$$

where Σ_1 is the variance-covariance matrix (see below), $|\Sigma_1|$ is the determinant, z is the column vector of the two variables x,y and z' is the row vector of x,y and μ_1 is the column vector (μ_{1x},μ_{1y}).

Similarly, the product of the two marginals is

$$r'(x,y)=\frac{1}{|2\Pi\Sigma_2|^{0.5}}\exp[-\frac{1}{2}(z-\mu_2)'\Sigma_2^{-1}(z-\mu_2)]$$

where μ_2 is the column vector (μ_{2x},μ_{2y}).

Note that in our case, using males and females from matings (joint distribution versus random mating distribution obtained from mating data) implies that mean and variances are equal between distributions ($\mu_1=\mu_2$, $\sigma_1=\sigma_2$), i.e. $\mu_{1x}=\mu_{2x}$, $\mu_{1y}=\mu_{2y}$ and $\sigma_{1x}=\sigma_{2x}$, $\sigma_{1y}=\sigma_{2y}$.

Let

$$\Sigma_1=\begin{pmatrix}\sigma_{xx} & \sigma_{xy}\\ \sigma_{xy} & \sigma_{yy}\end{pmatrix},\ \Sigma_1^{-1}=\frac{1}{S}\begin{pmatrix}\sigma_{yy} & -\sigma_{xy}\\ -\sigma_{xy} & \sigma_{xx}\end{pmatrix}\text{ with } S=\sigma_{xx}\sigma_{yy}-\sigma_{xy}^2$$

and

$$\Sigma_2 = \begin{pmatrix} \sigma_{xx} & 0 \\ 0 & \sigma_{yy} \end{pmatrix}, \ \Sigma_2^{-1} = \frac{1}{\sigma_{xx}\sigma_{yy}} \begin{pmatrix} \sigma_{yy} & 0 \\ 0 & \sigma_{xx} \end{pmatrix}$$

In this case the Kullback-Leibler divergences are (c.f. Kullback 1997 (1.6) p. 190)

$$KL(q' \| r') = \frac{1}{2} \log \frac{|\Sigma_2|}{|\Sigma_1|} - 1 + \frac{1}{2} tr(\Sigma_1 \Sigma_2^{-1}) = \frac{1}{2} \log \frac{\sigma_{xx}\sigma_{yy}}{S}$$

and

$$KL(r' \| q') = \frac{1}{2} \log \frac{|\Sigma_1|}{|\Sigma_2|} - 1 + \frac{1}{2} tr(\Sigma_2 \Sigma_1^{-1}) = \frac{1}{2} \log \frac{S}{\sigma_{xx}\sigma_{yy}} - 1 + \frac{\sigma_{xx}\sigma_{yy}}{S}$$

where *tr* is the trace (sum of diagonal elements) of a square matrix.

Therefore,

$$J_{PSI} = KL(q' \| r') + KL(r' \| q') = \frac{\sigma_{xx}\sigma_{yy}}{S} - 1 = \frac{\sigma_{xy}{}^2}{\sigma_{xx}\sigma_{yy} - \sigma_{xy}{}^2} = \frac{\rho^2}{1-\rho^2}.$$

A.2. General formula for non-random mating information

Consider two continuous traits, X in females and Y in males from a population and let the probability distribution Q'(x,y) with probability density $q'(x,y)$ that represents the mating probability for pairs with values in the infinitesimal interval ([x,x+dx],[y,y+dy]). Consider alternatively the probability distribution $Q(x,y)$ of random mating with the product of densities $q(x,y)=f(x)g(y)$ from X and Y, respectively. The divergence between the two distributions Q' and Q is given by

$$J=\int_{-a}^{+a}\int_{-a}^{+a}(q'(x,y)-q(x,y))\log\frac{q'(x,y)}{q(x,y)}dxdy, \quad \text{where } q(x,y)=f(x)g(y) \qquad \text{(A1).}$$

The difference $q'(x,y)$-$q(x,y)$ can be partitioned as follows

$$q'(x,y)-q(x,y)=q'(x,y)-f_1(x)f_2(y)+f_1(x)f_2(y)-q(x,y),$$

where $f_1(x)$ and $f_2(y)$ are the marginal densities for X and Y

$$f_1(x)=\int_{-a}^{+a}q'(x,y)dy$$

$$f_2(y)=\int_{-a}^{+a}q'(x,y)dx \qquad \text{(A2).}$$

Similarly, the logarithm

$$\log\frac{q'(x,y)}{q(x,y)}=\log\left(\frac{q'(x,y)}{f_1(x)f_2(y)}\frac{f_1(x)f_2(y)}{q(x,y)}\right)=\log\frac{q'(x,y)}{f_1(x)f_2(y)}+\log\frac{f_1(x)f_2(y)}{q(x,y)}.$$

Now note that the densities $f(x)$ and $g(y)$ depend on the population distribution, while the marginal densities $f_1(x)$ and $f_2(y)$ depend on the mating distribution so they are not necessarily equal. Then we have

$$J_{PSS}=\int_{-a}^{+a}\int_{-a}^{+a}(f_1(x)f_2(y)-q(x,y))\log\frac{f_1(x)f_2(y)}{q(x,y)}dxdy$$

and

$$J_{PSI}=\int_{-a}^{+a}\int_{-a}^{+a}(q'(x,y)-f_1(x)f_2(y))\log\frac{q'(x,y)}{f_1(x)f_2(y)}dxdy$$

therefore, as in the discrete case, the measure of non-random mating can be divided into

$$J=J_{PSS}+J_{PSI}+E \qquad \text{(A3).}$$

The component E is

$$E=\int_{-a}^{+a}\int_{-a}^{+a}(A+B)dxdy=\int_{-a}^{+a}\int_{-a}^{+a}Adxdy+\int_{-a}^{+a}\int_{-a}^{+a}Bdxdy=\int_{-a}^{+a}\int_{-a}^{+a}Adxdy$$

$$A=(f_1(x)f_2(y)-q(x,y))\log\frac{q'(x,y)}{f_1(x)f_2(y)}$$

$$B=(q'(x,y)-f_1(x)f_2(y))\log\frac{f_1(x)f_2(y)}{q(x,y)}.$$

We will show that

$$\int_{-a}^{+a}\int_{-a}^{+a}Bdxdy=\int_{-a}^{+a}\int_{-a}^{+a}[(q'(x,y)-f_1(x)f_2(y))\log\frac{f_1(x)f_2(y)}{q(x,y)}]dxdy=0 \qquad \text{(A4)}$$

so that,

$$E=\int_{-a}^{+a}\int_{-a}^{+a}(f_1(x)f_2(y)-q(x,y))\log\frac{q'(x,y)}{f_1(x)f_2(y)}\neq 0\Leftrightarrow J_{PSS}\neq 0\wedge J_{PSI}\neq 0 \qquad \text{(A5).}$$

Now, first recall that $q(x,y)=f(x)g(y)$ and then

$$\log\frac{f_1(x)f_2(y)}{q(x,y)}=\log\frac{f_1(x)}{f(x)}+\log\frac{f_2(y)}{g(y)}.$$

Also,

$$\int_{-a}^{+a}f_1(x)dx=\int_{-a}^{+a}\int_{-a}^{+a}q'(x,y)dxdy=\int_{-a}^{+a}f_2(y)dy=1$$

so that,

$$\int_{-a}^{+a}\int_{-a}^{+a}[f_1(x)f_2(y)\log\frac{f_1(x)}{f(x)}]dxdy=\int_{-a}^{+a}f_1(x)\log\frac{f_1(x)}{f(x)}dx\int_{-a}^{+a}f_2(y)dy=\int_{-a}^{+a}f_1(x)\log\frac{f_1(x)}{f(x)}dx$$

(A6),

and similarly

$$\int_{-a}^{+a}\int_{-a}^{+a}[f_1(x)f_2(y)\log\frac{f_2(y)}{g(y)}]dxdy=\int_{-a}^{+a}f_2(y)\log\frac{f_2(y)}{g(y)}dy \quad \text{(A7).}$$

Also using the marginal density definitions from (A2)

$$\int_{-a}^{+a}\int_{-a}^{+a}[q'(x,y)\log\frac{f_1(x)}{f(x)}]dxdy=\int_{-a}^{+a}\log\frac{f_1(x)}{f(x)}dx\int_{-a}^{+a}q'(x,y)dy=\int_{-a}^{+a}f_1(x)\log\frac{f_1(x)}{f(x)}dx$$

(A8)

and

$$\int_{-a}^{+a}\int_{-a}^{+a}[q'(x,y)\log\frac{f_2(y)}{g(y)}]dxdy=\int_{-a}^{+a}f_2(y)\log\frac{f_2(y)}{g(y)}dy \quad \text{(A9).}$$

Finally, noting that (A6) equals (A8) and (A7) equals (A9) and rearranging and substituting in (A4) we obtain

$$\int_{-a}^{+a}\int_{-a}^{+a}Bdxdy=\int_{-a}^{+a}\int_{-a}^{+a}[(q'(x,y)-f_1(x)f_2(y))\log\frac{f_1(x)f_2(y)}{q(x,y)}]dxdy=(A6)+(A7)-(A8)-(A9)=0$$

A.3. Simulation results tables

Table S1. One sex is choosy. Choice implies preference for similarity with a bias of 1.5. Mating is with replacement. *N* is the population size and *n* the sample size of matings. The quantitative trait distribution is *Z*(0,1). Number of runs: 10,000. % sig: The % of tests with *p*-value ≤0.05. The values in parentheses correspond to the average value of the statistic for the total number of significant runs.

N	***n***	***Choice***	***s***	**% sig. (ρ)**	**%sig. ($\rho^2/(1-\rho^2)$)**
10^4	500	0	--	4.8 (0.004)	4.9 (0.01)

10^4	500	0.2	0.08	99.9 (0.22)	99.9 (0.05)
10^4	500	0.4	0.08	100 (0.54)	100 (0.41)
10^4	500	0.8	0.08	100 (0.87)	100 (3.19)
10^4	500	1	0.08	100 (0.89)	100 (3.99)
500	50	0	--	5 (-0.025)	5.9 (0.12)
500	50	0.2	0.08	44.7 (0.37)	48.2 (0.17)
500	50	0.4	0.08	100 (0.67)	100 (0.90)
500	50	0.8	0.08	100 (0.84)	100 (2.54)
500	50	1	0.08	100 (0.90)	100 (4.43)
500	50	0.2	0.01	100 (0.97)	100 (64.72)
500	50	0.4	0.01	100 (0.99)	100 (41.97)
500	50	0.8	0.01	100 (1)	100 (115.87)
500	50	1	0.01	100 (1)	100 (3935.1)

Table S2. Mating under a logistic model with α=2. There is no preference for similarity. *N* is the population size and *n* the sample size of matings. The quantitative trait distribution is *Z*(0,1). Number of runs: 10,000. % sig: The % of tests with *p*-value ≤0.05. The values in parentheses correspond to the average value of the statistic for the total number of significant runs.

N	***n***	**Mating**	**% sig. (ρ)**	**%sig. ($\rho^2/(1-\rho^2)$)**
10^4	500	replacement	5.3 (0.003)	5.4 (0.01)
10^4	500	individual	4.9 (0.003)	5.0 (0.01)
10^4	500	mass-encounter	4.7 (0.004)	4.8 (0.01)
500	50	replacement	5.0 (-0.01)	6.0 (0.12)
500	50	individual	5.2 (0.02)	6.3 (0.12)
500	50	mass-encounter	5.4 (0.03)	6.4 (0.12)

Table S3. Bimodal populations with mixture proportion π=0.5, μ_2- μ_1-=3 and σ_1+σ_2 = 2. One sex is choosy. Choice implies preference for similarity. Mating is with replacement. *N* is the population size and *n* the sample size of matings. The quantitative trait distribution is *Z*(0,1). Number of runs: 10,000. % sig: The % of tests with *p*-value ≤0.05. The values in parentheses correspond to the average value of the statistic for the total number of significant runs.

N	***n***	***Choice***	***s***	**% sig. (ρ)**	**%sig. ($\rho^2/(1-\rho^2)$)**
10^4	500	0	--	5.1 (0.004)	5.2 (0.01)
10^4	500	0.2	0.08	90 (0.15)	90.1 (0.03)
10^4	500	0.4	0.08	100 (0.43)	100 (0.22)
10^4	500	0.8	0.08	100 (0.82)	100 (1.98)
10^4	500	1	0.08	100 (0.89)	100 (3.79)
10^4	500	0.2	0.01	100 (0.95)	100 (9.99)
10^4	500	0.4	0.01	100 (0.99)	100 (48.67)
10^4	500	0.8	0.01	100 (1.0)	100 (242.35)
10^4	500	1	0.01	100 (1.0)	100 (397.16)
500	50	0	--	5.1 (-0.01)	6.1 (0.12)
500	50	0.2	0.08	43.2 (0.38)	46.2 (0.18)
500	50	0.4	0.08	99.7 (0.59)	99.7 (0.58)
500	50	0.8	0.08	100 (0.93)	100 (6.58)
500	50	1	0.08	100 (0.94)	100 (7.74)
500	50	0.2	0.01	100 (0.97)	100 (18.57)
500	50	0.4	0.01	100 (0.99)	100 (47.17)
500	50	0.8	0.01	100 (1.0)	100 (177.0)
500	50	1	0.01	100 (1.0)	100 (324.56)

Table S4. Both sexes are choosy. Mating with replacement. *N* is the population size and *n* the sample size of matings. Choice implies preference for similarity. The quantitative trait distribution is *Z*(0,1). Number of runs: 1,000. % sig: The % of tests with *p*-value

≤0.05. The values in parentheses correspond to the average value of the statistic for the total number of significant runs.

N	***n***	***Choice***	***s***	**% sig. (ρ)**	**%sig. ($\rho^2/(1-\rho^2)$)**
10^4	500	0	--	4.4 (0.03)	5.4 (0.06)
10^4	500	0.2	0.08	24.6 (0.25)	26.1 (0.07)
10^4	500	0.4	0.08	100 (0.47)	100 (0.3)
10^4	500	0.8	0.08	100 (0.75)	100 (1.39)
10^4	500	0.2	0.01	100 (0.93)	100 (7)
10^4	500	0.4	0.01	100 (0.98)	100 (26)
10^4	500	0.8	0.01	100 (1)	100 (106)
500	50	0	--	4.5 (0.017)	5.7 (0.12)
500	50	0.2	0.08	36.3 (0.36)	38.5 (0.16)
500	50	0.4	0.08	99.8 (0.58)	99.9 (0.57)
500	50	0.8	0.08	100 (0.81)	100 (2.1)
500	50	0.2	0.01	100 (0.96)	100 (12.7)
500	50	0.4	0.01	100 (0.99)	100 (56.7)
500	50	0.8	0.01	100 (1.0)	100 (206)

References

Alonzo, S. H., and M. R. Servedio. 2019. Grey zones of sexual selection: why is finding a modern definition so hard? Proc. R. Soc. B Biol. Sci. 286:20191325.

Andersson, M. 1994. Sexual selection. Princeton University Press, Princeton, N.J.

Arnold, S. J., and M. J. Wade. 1984. On the measurement of natural and sexual selection: applications. Evolution 720–734.

Belov, D. I., and R. D. Armstrong. 2011. Distributions of the Kullback–Leibler divergence with applications. Br. J. Math. Stat. Psychol. 64:291–309. Wiley Online Library.

Bürger, R. 2000. The Mathematical theory of selection, recombination, and mutation. John Wiley, Chichester.

Caballero, A. 2020. Quantitative Genetics. Cambridge University Press.

Carvajal-Rodríguez, A. 2019. A generalization of the informational view of non-random mating: Models with variable population frequencies. Theor. Popul. Biol. 125:67–74.

Carvajal-Rodríguez, A. 2018a. MateSim: Monte Carlo simulation for the generation of mating tables. Biosystems 171:26–30.

Carvajal-Rodríguez, A. 2020. Multi-model inference of non-random mating from an information theoretic approach. Theor. Popul. Biol. 131:38–53.

Carvajal-Rodríguez, A. 2018b. Non-random mating and information theory. Theor. Popul. Biol. 120:103–113.

Carvajal-Rodríguez, A., and E. Rolán-Alvarez. 2014. A comparative study of Gaussian mating preference functions: a key element of sympatric speciation models. Biol. J. Linn. Soc. 113:642–657.

Carvajal-Rodríguez, A., and E. Rolan-Alvarez. 2006. JMATING: a software for the analysis of sexual selection and sexual isolation effects from mating frequency data. BMC Evol Biol 6:40.

Clancey, E., T. R. Johnson, L. J. Harmon, and P. A. Hohenlohe. 2022. Estimation of the strength of mate preference from mated pairs observed in the wild. Evolution 76:29–41.

Cotto, O., and M. R. Servedio. 2017. The Roles of Sexual and Viability Selection in the Evolution of Incomplete Reproductive Isolation: From Allopatry to Sympatry. Am. Nat. 190:680–693.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Oxford University Press, Oxford, New York.

Darwin, C. 1871. The descent of man, and selection in relation to sex. Murray.

de Cara, M. A., N. H. Barton, and M. Kirkpatrick. 2008. A model for the evolution of assortative mating. Am. Nat. 171:580–596.

Edward, D. A. 2015. The many facets of mate choice: a response to comments on Edward. Behav. Ecol., doi: 10.1093/beheco/arv021.

Endler, J. A. 1986. Natural selection in the wild. Princeton University Press, Princeton, N.J.

Erlandsson, J., and E. Rolán-Alvarez. 1998. Sexual selection and assortative mating by size and their roles in the maintenance of a polymorphism in Swedish Littorina saxatilis populations. Hydrobiologia 378:59–69.

Estévez, D., E. Kozminsky, A. Carvajal-Rodríguez, A. Caballero, R. Faria, J. Galindo, and E. Rolán-Alvarez. 2020. Mate Choice Contributes to the Maintenance of Shell Color Polymorphism in a Marine Snail via Frequency-Dependent Sexual Selection. Front. Mar. Sci. 7.

Fernández-Meirama, M., E. Rolán-Alvarez, and A. Carvajal-Rodríguez. 2022. A Simulation Study of the Ecological Speciation Conditions in the Galician Marine Snail Littorina saxatilis. Front. Genet. 13.

Fisher, R. A. 1918. The correlations between relatives on the supposition of Mendelian inheritance. Trans. R. Soc. Edinb. 52:399–433.

Gavrilets, S. 2004. Fitness landscapes and the origin of species. Princeton University Press, Princeton, N.J.

Hansen, T. F. 2018. Fitness in Evolutionary Biology. Preprints.

Hartl, D. L. 2020. A Primer of Population Genetics and Genomics. Oxford University Press.

Hartl, D. L., and A. G. Clark. 1997. Principles of Population Genetics. Third. Sinauer Associates, Inc., Sunderland, MA.

Jennions, M. D., and M. Petrie. 1997. Variation in mate choice and mating preferences: a review of causes and consequences. Biol. Rev. 72:283.

Jiang, Y., D. I. Bolnick, and M. Kirkpatrick. 2013. Assortative Mating in Animals. Am. Nat. 181:E125–E138.

Kirkpatrick, M., and V. Ravigné. 2002. Speciation by natural and sexual selection: models and experiments. Am. Nat. 159 Suppl 3:S22-35.

Kokko, H., H. Klug, and M. D. Jennions. 2012. Unifying cornerstones of sexual selection: operational sex ratio, Bateman gradient and the scope for competitive investment. Ecol. Lett. 15:1340–1351.

Kullback, S. 1997. Information theory and statistics. Courier Corporation.

Lau, S. L. Y., G. A. Williams, A. Carvajal-Rodríguez, and E. Rolán-Alvarez. 2021. An integrated approach to infer the mechanisms of mate choice for size. Anim. Behav. 175:33–43.

Lehtonen, J. 2022. Bateman gradients from first principles. Nat. Commun. 13:3591. Nature Publishing Group.

Lewontin, R., D. Kirk, and J. Crow. 1968. Selective mating, assortative mating, and inbreeding: definitions and implications. Eugen Q 15:141–3.

Lindsay, W. R., S. Andersson, B. Bererhi, J. Höglund, A. Johnsen, C. Kvarnemo, E. H. Leder, J. T. Lifjeld, C. E. Ninnes, M. Olsson, G. A. Parker, T. Pizzari, A. Qvarnström, R. J. Safran, O. Svensson, and S. V. Edwards. 2019. Endless forms of sexual selection. PeerJ 7:e7988. PeerJ Inc.

López-Cortegano, E., C. Carpena-Catoira, A. Carvajal-Rodríguez, and E. Rolán-Alvarez. 2020. Mate choice based on body size similarity in sexually dimorphic populations causes strong sexual selection. Anim. Behav. 160:69–78.

Ng, T. P. T., E. Rolán-Alvarez, S. S. Dahlén, M. S. Davies, D. Estévez, R. Stafford, and G. A. Williams. 2019. The causal relationship between sexual selection and sexual size dimorphism in marine gastropods. Anim. Behav. 148:53–62.

Parker, G. A. 2014. The sexual cascade and the rise of pre-ejaculatory (Darwinian) sexual selection, sex roles, and sexual conflict. Cold Spring Harb. Perspect. Biol. 6:a017509–a017509.

Parker, G. A., and T. Pizzari. 2015. Sexual Selection: The Logical Imperative. Pp. 119–163 *in* T. Hoquet, ed. Current Perspectives on Sexual Selection: What's left after Darwin? Springer Netherlands, Dordrecht.

Prum, R. O. 2012. Aesthetic evolution by mate choice: Darwin's really dangerous idea. Philos. Trans. R. Soc. Lond. B Biol. Sci. 367:2253–2265.

Reid, D. G. 2007. The genus Echinolittorina Habe, 1956 (Gastropoda: Littorinidae) in the Indo-West Pacific Ocean. Zootaxa 1420:1–161.

Rettelbach, A., M. Kopp, U. Dieckmann, and J. Hermisson. 2013. Three Modes of Adaptive Speciation in Spatially Structured Populations. Am. Nat. 182:E215–E234.

Rolán-Alvarez, E., and A. Caballero. 2000. Estimating sexual selection and sexual isolation effects from mating frequencies. Evolution 54:30–6.

Rolán-Alvarez, E., A. Carvajal-Rodriguez, A. de Coo, B. Cortes, D. Estevez, M. Ferreira, R. Gonzalez, and A. D. Briscoe. 2015. The scale-of-choice effect and how estimates of assortative mating in the wild can be biased due to heterogeneous samples. Evolution 69:1845–57.

Rolan-Alvarez, E., G. Williams, T. Ng, D. Estévez-Barcia, M. Davies, S. Saltin, and R. Stafford. 2021. Data for: The causal relationship between sexual selection and sexual size dimorphism in marine gastropods. 1. Mendeley Data.

Roughgarden, J., E. Adkins-Regan, E. Akçay, JC. Crawford, R. Gadagkar, SC. Griffith, CA. Hinde, T. Hoquet, C. O'Connor, ZM. Prokop, RO. Prum, S. Shafir, SS. Snow, r D. Taylo, J. Van Cleve, and M. Weisberg. 2015. Sexual selection studies: a NESCent catalyst meeting. PeerJ Prepr. 3e680v3.

Sachdeva, H., and N. H. Barton. 2017. Divergence and evolution of assortative mating in a polygenic trait model of speciation with gene flow. Evolution 71:1478–1493.

Shuker, D. M., and C. Kvarnemo. 2021. The definition of sexual selection. Behav. Ecol. 32:781–794.

Vries, C. de, and J. Lehtonen. 2023. Sex-specific assumptions and their importance in models of sexual selection. Trends Ecol. Evol. 0. Elsevier.

Wacker, S., and T. Amundsen. 2014. Mate competition and resource competition are inter related in sexual selection. J. Evol. Biol. 27:466–477.

Xie, Y., S. Cheng, and X. Zhou. 2015. Assortative mating without assortative preference. Proc. Natl. Acad. Sci. 112:5974–5978.

Yukilevich, R. 2023. Female preferences between incipient species of Drosophila athabasca reveal divergent closed functions, epistasis, and segregation of veiled preference alleles. Evolution 77:496–508.