A generalization of the informational view of non-random mating:

Models with variable population frequencies.

A. Carvajal-Rodríguez

Departamento de Bioquímica, Genética e Inmunología. Universidad de Vigo, 36310
Vigo, Spain.

*: A. Carvajal-Rodríguez. Departamento de Bioquímica, Genética e Inmunología.

Universidad de Vigo, 36310 Vigo, Spain. Phone: +34 986813828

email: acraaj@uvigo.es

Keywords: Encounter-mating, monogamy, mate choice, sexual selection, sexual isolation, Kullback-Leibler divergence, population genetics.

Abstract

Mate choice may generate non-random mating patterns. It has been recently shown that the mating distribution caused by mate choice can be expressed as a gain in information with respect to random mating. In that model, the population phenotypic frequencies were assumed as constant during the breeding season. In the present work such restriction was relaxed to consider different encounter-mating processes in which the population frequencies of available individuals change over mating rounds. As with the constant case, here we describe the change in the mating phenotypes by the flow of information with respect to random mating. This information can be partitioned into sexual selection, sexual isolation and a mixed effect. Likewise, the pairwise statistics for total change, sexual selection and sexual isolation are generalized for variable population frequencies.

The new tests had more power for the detection of the effects of non-random mating when the population frequencies vary during the breeding season. The differences in power were high for sexual selection but slight for sexual isolation scenarios. However, the application of the new formulas require the estimation of frequencies at each mating round. Therefore, choosing one or another type of statistics would depend on the biological scenario as well as in the availability and easiness to split the sampling in more than one mating round.

19

20

1. Introduction

Mate choice can be defined in general, as any aspect of an organism phenotype that leads to an individual to engage in sexual activity with some partners more likely than with others (Rosenthal, 2017).

From a population genetics point of view, mate choice is defined by its effects: it is the observed mating frequency deviation with respect to random mating. So defined, the effects of mate choice can be partitioned into sexual selection and sexual or behavioural isolation (intersexual selection). Sexual selection is defined as the observed change in gene or phenotype frequencies in mated individuals with respect to population frequencies. Sexual isolation is defined as the deviation from random mating in mated individuals (Rolán-Alvarez and Caballero, 2000).

In a previous work (Carvajal-Rodríguez, 2018), the mating distribution caused by mate choice was expressed as the gain in information with respect to random mating. In that model, the population phenotypic frequencies were assumed as constant over the breeding season. In the present work, the previous results were extended to include the case where population frequencies vary over different mating rounds during the breeding season.

The extension to include variable population frequencies is appropriate for a better description of monogamous species in which pair formation occurs by different mating rounds. In this case, the frequencies of available individuals can change during the same breeding season (Gimelfarb, 1988).

We show that, by relaxing the condition of constant population frequencies, a new generalized information partition can be obtained. As before, this information can be partitioned into sexual selection, sexual isolation, and a mixed effect, while the

corresponding G-like tests can be performed over each information index (Carvajal-Rodríguez, 2018). The new equations become the previous ones as soon as the population frequencies are set to constant values.

Prairie vole (*Microtus ochrogaster*) is a good model organism that can help us to demonstrate the new tests. This mammal is a monogamous species that establishes long-term heterosexual relationships among adult mates. This rodent lives mainly in the grasslands of the central United States and Canada. Its living conditions imply limited food and water resources, which may have contributed to the evolution of a monogamous life strategy in this species (Williams et al., 1992; Young et al., 2011). It is known that the long-term pair-bonds in this species are regulated by a variety of neurotransmitters and driven by epigenetic events (Wang et al., 2013). Indeed, the prairie vole has become a model system for the study of the neurobiology of monogamy, social attachment and nurturing. However, for our purpose is enough to consider that, after encounter and mating, there is an oxytocin-modulated monogamous behaviour which, from the point of view of the frequency of available adults, may resemble a model without replacement. That is, the already formed mating pairs are excluded from the mating pool. Therefore, we simulated sampling in each of several mating rounds to compare variable versus constant frequency tests for three different scenarios (random mating, sexual isolation and sexual selection).

Following, we first obtained the general information equation for the total change, and then developed the partition into the different effects. Afterwards, the new model was demonstrated by example, and finally the new outcomes were summarized in the discussion.

69

70 **2. Non-random mating with variable population frequencies**

71 During the breeding season the formation of mating pairs is a two-step process in which
72 first, an encounter occurs and second, it ends or not in a successful mating (Gimelfarb,
73 1988).

74 When males and females can have multiple mates, or when only a fraction of the
75 individuals of both sexes do actually mate in the current breeding season, it is
76 reasonable to assume that the availability of the different phenotypes are not affected by
77 the matings already occurred. Therefore, the population frequencies can be considered
78 constant during the reproductive season. In this case, the process of mating resembles a
79 sampling with replacement from the point of view of the available phenotypes
80 (Carvajal-Rodríguez, 2018).

81 On the other hand, when we consider a monogamous species in which most of the
82 sexually mature individuals are involved in a series of mating rounds, then the
83 availability of individuals is affected by the previous matings. In this case, the
84 population phenotypic frequencies are not constant during the breeding season and the
85 process resembles a sampling without replacement.

86

87 *2.1 General model*

88 Let consider a population where the mating occurs in various sequential rounds so that
89 in the r^{th} round the number of available (unmated) females is n_{1r} and the number of
90 available males is n_{2r} . Thus, the relative frequency of the female phenotype i is $p_{1ri} = n_{1ri}$
91 $/ n_{1r}$, i.e. the number of available i -type females divided by the total number of available

females. If the model is without replacement, the frequencies may be different in the different mating rounds. Similarly, for males with phenotype j we have $p_{2rj} = n_{2rj} / n_{2r}$. During the breeding season, the absolute frequencies of adults will be updated depending on the type of the encounter-mating scheme. Under the individual encounter-mating model (Gimelfarb, 1988) there is only one mate by round. In this case, if the new mating includes a female of i - type then in the next round we have $n_{1(r+1)i} = n_{1ri} - 1$ and $n_{1r+1} = n_{1r} - 1$, and similarly for males. In general, if each mating round involves a number U_r of matings then in the next round, we will have $n_{1r+1} = n_{1r} - U_r$ available females and $n_{2r+1} = n_{2r} - U_r$ available males, and the corresponding phenotypic classes updated depending on the specific matings that occurred.

Now, consider the encounter of a female of type i with a male of type j . The mutual mating propensity m_{ij} is the number of matings, after encounter of i with j , in a given environment (Carvajal-Rodríguez, 2018).

In mating round r there is a probability $q_{rij} = p_{1ri} \times p_{2rj}$ of encounter between i and j . Therefore, the number of matings $i \times j$ at this round is $q_{rij}m_{ij}$. The sum over the mating rounds gives the total number of $i \times j$ matings and so, the observed relative frequencies of these matings after R rounds can be expressed as

$$q'_{Rij} = \frac{m_{ij}}{M_R} \sum_{r=1}^R q_{rij} \quad (1)$$

where $M_R = \sum_{i,j} (m_{ij} \sum_{r=1}^R q_{rij})$.

Note that M_R defines the total number of matings at the end of the season and can be expressed as the sum of matings at each round so that

$$M_R = \sum_{r=1}^R M_r \text{ with } M_r = \sum_{i,j} q_{rij} m_{ij}$$

Also note that the maximum possible number of matings occurring in round r is

$$\max(M_r) = \max\{m_{ij}, \dots\} \text{ for every } i \text{ and } j.$$

When mating is at random, the mating propensity is constant i.e. $m_{ij} = m$ for every pair i ,

j and so

$$M_R = m \times R.$$

Then, by noting the expected frequencies under random mating as q_{Rij} and by

substituting the constant propensity in (1) we obtain

$$q_{Rij} = \frac{1}{R} \sum_{r=1}^R q_{rij}$$

Note that if the frequencies are constant over the different mating rounds then we

recover the formulas as given in (Carvajal-Rodríguez, 2018). So that with constant

frequencies

$$q_{Rij} = q_{ij}; M_R = R \sum_{i,j} m_{ij} q_{ij} = RM; q'_{Rij} = q'_{ij} = q_{ij} \frac{m_{ij}}{M}.$$

When the population frequencies are variable over the breeding season, the correct

pairwise statistic for measuring the deviation from random mating is given by $RPTI_{ij} =$

q'_{Rij}/q_{Rij} , which is the ratio of the frequency of the observed pair types divided by the

expected pair types under random mating, calculated from the population frequencies

when these are variable over the same breeding season. If the population frequencies

can be considered constant, the $RPTI_{ij}$ statistic becomes the pair total index (PTI_{ij}) as

originally defined in (Rolán-Alvarez and Caballero, 2000).

As with the constant frequency case we can measure the change with respect to random mating

$$\Delta q_{Rij} = q'_{Rij} - q_{Rij} = \left(\frac{m_{ij}}{M_R} - \frac{1}{R}\right) \sum_{r=1}^R q_{rij}$$

and from (1) we can also express the propensity as a ratio of frequencies

$$m_{ij} = M_R \frac{q'_{Rij}}{\sum_{r=1}^R q_{rij}} = \frac{M_R}{R} \frac{q'_{Rij}}{q_{Rij}}$$

to finally obtain the expression for the mean population change in terms of the logarithm of the mating propensity as

$$\Delta L_R = \sum_{i,j} \Delta q_{Rij} \log \left(\frac{q'_{Rij}}{q_{Rij}} \right) = J(q_R', q_R) \equiv J_{RPTI} \quad (2)$$

which is the Kullback-Leibler symmetrized divergence that measures the gain in information when the differential mating propensity moves the population from mating frequencies q_R to q'_R or vice versa. Note that J_{RPTI} becomes the information equation for the constant case when the population frequencies are set to constant values (J_{PTI} in eq. (2) in Carvajal-Rodríguez, 2018).

As with J_{PTI} , if we take the natural logarithm in (2) and multiply by the total number of matings, the obtained quantity is well approximated by a chi-square under the null hypothesis of random mating. The possible difficulty for performing this test, is being able to estimate the expected frequencies under random mating over the distinct mating rounds.

2.2. The information partition

To perform the information partition, we have to take into account that the expected pair types under random mating calculated from mated individuals, must be computed within each mating round.

In round r , the number of matings involving females of phenotype i is

$$\sum_j^{males} q_{rij} m_{ij}$$

Then, the relative frequency of females i calculated from the matings within this round is

$$p'_{1ri} = \frac{\sum_j^{males} q_{rij} m_{ij}}{\sum_i^{females} \sum_j^{males} q_{rij} m_{ij}} = \frac{p_{1ri} \sum_j^{males} p_{2rj} m_{ij}}{M_r} = \frac{p_{1ri} m'_{fri}}{M_r}$$

where $M_r = \sum_i^{females} \sum_j^{males} q_{rij} m_{ij}$ and m'_{fri} is the marginal propensity of females with phenotype i in the round r .

Similarly, the frequency of males calculated from the matings,

$$p'_{2rj} = \frac{p_{2rj} \sum_i^{females} p_{1ri} m_{ij}}{M_r} = \frac{p_{2rj} m'_{mrj}}{M_r}$$

where m'_{mrj} is the marginal propensity of males with phenotype j in the round r . Also, note that $\sum_i p'_{1ri} = M_r / M_r = 1$ and $\sum_j p'_{2rj} = M_r / M_r = 1$.

Then, after R rounds we get the probability of random mating $i \times j$ computed from matings as

$$P'_{Rij} = \frac{\sum_r p'_{1ri} p'_{2rj}}{\sum_{ij} p'_{1ri} p'_{2rj}} = \frac{\sum_r p'_{1ri} p'_{2rj}}{R} \quad (3)$$

Now, we may express the components of J_{RPTI} in (2) as

$$\Delta(q_{Rij}) = (q'_{Rij} - P'_{Rij}) + (P'_{Rij} - q_{Rij}) = \Delta(P'_{Rij}) + \Delta(P_{Rij}) \quad (4)$$

and

$$RPTI_{ij} = RPSI_{ij} \times RPSS_{ij} \quad (5)$$

where $RPSI_{ij} = q'_{Rij} / P'_{Rij}$ and $RPSS_{ij} = P'_{Rij} / q_{Rij}$ correspond to the generalized version of the pair sexual isolation and pair sexual selection statistics respectively, for the case of variable population frequencies over different mating rounds. These statistics become the original ones when the population frequencies are constant.

By substitution of (4) and (5) in (2) we obtain the information partition

$$J_{RPTI} = J_{RPSI} + J_{RPSS} + E_R \quad (6)$$

where

$$J_{RPSI} \equiv \sum_{i,j} \Delta P'_{Rij} \log \left(\frac{q'_{Rij}}{P'_{Rij}} \right)$$

$$J_{RPSS} \equiv \sum_{i,j} \Delta(P_{Rij}) \log \left(\frac{P'_{Rij}}{q_{Rij}} \right)$$

$$E_R = J_{RPTI} - (J_{RPSI} + J_{RPSS})$$

We have generalized the information partition of total change due to non-random mating, for models without replacement in which the population frequencies are updated after each mating round. The original partition developed in (Carvajal-Rodríguez, 2018), emerges as a particular case from (6) when the population frequencies are constant or the number of mating rounds are reduced to one ($R = 1$).

Concerning the sexual selection information, in the previous work we showed that under the constant frequency assumption, sexual selection can be divided into its female and male components. However, the generalized index (J_{RPSS}), is no longer the exact sum of the information in both sexes. There is a new error term that increases with the number of mating rounds. To appreciate this, let consider the female and male population frequencies averaged over R mating rounds

$$p_{1Ri} = \frac{1}{R} \sum_{r=1}^R p_{1ri}$$

$$p_{2Rj} = \frac{1}{R} \sum_{r=1}^R p_{2rj}$$

And the averages for the frequencies calculated from mated individuals, i.e.

$$p'_{1Ri} = \frac{1}{R} \sum_r p'_{1ri}$$

$$p'_{2Rj} = \frac{1}{R} \sum_r p'_{2rj}$$

Now we define the sexual selection effect in females

$$J_{RS1} \equiv \sum_i \Delta p_{1Ri} \log \left(\frac{p'_{1Ri}}{p_{1Ri}} \right)$$

and in males

$$J_{RS2} \equiv \sum_j \Delta p_{2Rj} \log \left(\frac{p'_{2Rj}}{p_{2Rj}} \right)$$

where $\Delta p_{1Ri} = p'_{1Ri} - p_{1Ri}$ and $\Delta p_{2Rj} = p'_{2Rj} - p_{2Rj}$.

206 Then, the sexual selection information is

$$207 \quad J_{RPSS} = \sum_{i,j} \Delta(P_{Rij}) \log \left(\frac{P'_{Rij}}{q_{Rij}} \right) = J_{RS1} + J_{RS2} + \varepsilon_R \quad (7)$$

208 the error term arises because the logarithm for P'_{Rij} and q_{Rij} correspond to a sum of
 209 products over R , which makes sense because encounters only happens within the same
 210 round i.e., for $r \neq s$, the quantities $p_{1ri} \times p_{2sj}$ or $p_{1si} \times p_{2rj}$ have no biological sense here.
 211 For the partition in (7) to be exact ($\varepsilon_R = 0$) we need to equate the sum of products to the
 212 product of the sums, i.e. $\sum_r (p_{1ri} \times p_{2rj}) = (\sum_r p_{1ri}) \times (\sum_r p_{2rj})$ which is not true in general.
 213 The same problem occurs with p'_{1ri} and p'_{2rj} . The discrepancy increases with the number
 214 of rounds while it disappears if we have only one round or the frequencies are constant.

215 There is still another issue with the information partition estimates under the individual
 216 encounter models. The equation (3) corresponds to the formal model for the expected
 217 random mating calculated from matings instead of from population frequencies. It is
 218 usually estimated simply by the relative frequency of phenotypes within the matings at
 219 each round. However, under the individual encounter model there is only one mating
 220 per round, and estimating (3) by the count of one individual in the mating per round,
 221 produces a value of P'_{Rij} that is exactly the same as q'_{Rij} from equation (1) so that, the
 222 sexual isolation effect is not detected but included as a sexual selection effect. In fact,
 223 encounter models where the number of mating per round is equal or less than the
 224 product of the number of phenotypic classes ($k_1 \times k_2$), will be biased against the
 225 detection of sexual isolation effects if the estimation of J_{RPSI} is performed by counting
 226 matings per round. A solution for this problem will be suggested in the Example section
 227 below. On the contrary, when the number of matings per round is higher than the
 228 product of phenotypic classes (e.g. $k_1 \times k_2 = 4$) the estimation works better than the
 229 constant frequency estimators.

2.3. Generalized joint isolation index

Having defined $RPSI_{ij}$ as a generalized version of the sexual isolation pairwise statistic for variable population frequencies, we can obtain the corresponding joint isolation index as

$$I_{RPSI} = \frac{(k-1) \sum_i RPSI_{ii} - \sum_{i,j} RPSI_{ij}}{(k-1) \sum_i RPSI_{ii} + \sum_{i,j} RPSI_{ij}}$$

where k is the number of phenotypic classes ($k = k_1 = k_2$ for the joint index to be applicable). The index matches the original one as defined in (Carvajal-Rodriguez and Rolan-Alvarez, 2006) when the population frequencies are constant.

When the population frequencies are constant (models with replacement), we already know that intersexual selection (sexual isolation) emerges as a deviation of the multiplicative effect in the propensities. That is, if the normalized mutual propensity is the product of the relative marginal effects of each partner there is no sexual isolation (Carvajal-Rodríguez, 2018). However, in the general model the condition is affected by the mating round and it becomes

$$\sum_r \frac{Rm_{ij}}{M_R} q_{rij} = \sum_r \frac{m'_{fri}}{M_r} \frac{m'_{mrj}}{M_r} q_{rij}$$

so that, a sufficient condition for the absence of sexual isolation effects over the frequencies is

$$\frac{Rm_{ij}}{M_R} = \frac{m'_{fri}}{M_r} \frac{m'_{mrj}}{M_r} \forall i, j, r$$

Note that under constant frequencies $m'_{fi} = m'_{fi}$; $m'_{mrj} = m'_{mj}$; $M_r = M$ and $M_R = R \times M$ and then we recover the same condition as from formula (5) in (Carvajal-Rodríguez, 2018).

3. Example

The difference between the variable frequencies and the constant frequencies approaches can be shown by a toy-model based on a monogamous species. One of these species is the prairie vole (*Microtus ochrogaster*) that is a monogamous rodent that establishes long-term relationships between adult mates.

Concerning mate choice, there is no clear evidence of prairie vole mate choice behaviour in field studies. Indeed, some results indicate that prairie vole just mate with the first available partner (Getz et al., 2004; Keane et al., 2007). On the other hand, laboratory studies demonstrate that length polymorphism of microsatellite DNA related with genes within the vasopressin–oxytocin pathway (e.g. arginine vasopressin receptor 1a, *Avpr1a*, and oxytocin receptor, *Oxtr*) are correlated with female social and sexual preferences (Castelli et al., 2011; Keane et al., 2017). In addition, a potential for assortative mating at the *Oxtr* locus has recently been suggested in the bank vole (*Myodes glareolus*) species (Watts et al., 2017).

In consequence, we simulated three examples, first for a trait not linked with choice (random mating), and second and third, for a trait with positive assortative or sexual selection effects. For each example, we assayed both an individual and a mass encounter models (Gimelfarb, 1988). In the first, only one pair mate at each mating round while in the second up to 10 individuals can mate simultaneously. In both cases, the population frequencies were updated after each mating round.

The population consisted of 200 sexually mature adults with 50 individuals of each phenotype and sex. The species is monogamous, and assuming the whole population is involved in the breeding season, we can have up to 100 mating pairs. Therefore, we need 100 mating rounds ($R = 100$) to sample 100 matings under the individual model (just one mating per round) while we need at least 10 rounds ($R \geq 10$) for sampling the same number of matings under the mass encounter model (at least one mating and a maximum of 10 at each round).

For comparing the performance between the two types of information indices we ran 10,000 breeding seasons for each case.

3.1. Random mating

As a trait not related to choice, we considered the belly which in this species is yellow, ranging from pale to dark, and it has not known effect on mate choice. Thus, we may classify the matings by the belly color phenotype as pale yellow (y) or dark (d). The initial frequency of the phenotype was assumed to be equally distributed with intermediate frequency in both sexes so $p_{1y} = p_{2y} = 0.5$ and $p_{1d} = p_{2d} = 0.5$.

Because mating is at random (mutual propensities are equal) we simply performed the Mont Carlo sampling based on the phenotypic frequencies. That is, to incorporate a female into the mating pair, we checked if a uniform $U(0,1)$ random number was lower than p_{1y} and then a yellow female entered to the pair, if not, a dark female did it. For incorporating the male partner, we checked if a new random number was lower than p_{2y} then a yellow male also entered to the pair, if not, a dark male did it. After 1 (individual model) or 10 (mass model) pairs were formed, the frequencies were updated, and the process repeated until a total of 100 pairs was reached.

In Table 1, we can see the result of 10 mating rounds under the mass encounter model. During the mating rounds, the population frequencies were recorded jointly with the number of matings for the different classes. At the end of the breeding season (after round 10), the sum of the observed matings over rounds corresponded to the finally observed matings that we would have recorded if we do not distinguish among the different mating rounds.

With the recorded data, we can compute J_{RPTI} and J_{PTI} which for the specific case in the Table 1 gave non-significant values at 5% significance level ($J_{RPTI}=0.01$, $J_{PTI}=0.002$). Recall that each test is performed multiplying the corresponding index by the total number of matings. The obtained value follows a chi-square distribution with 3 degrees of freedom under the null hypothesis of random-mating.

This mating process was repeated 10,000 times and the obtained false positive rate was 4% for J_{RPTI} and 1% for J_{PTI} . The same percentages were obtained for the individual encounter model.

TABLE 1. Result of 10 mating rounds under mass encounter with random mating:

$$m_{yy}=m_{yd}=m_{dy}=m_{dd}=10.$$

Round r	p_{1ry}	p_{2ry}	$y \times y$	$y \times d$	$d \times y$	$d \times d$
1	0.5	0.5	1	4	4	1
2	0.5	0.5	3	3	4	0
3	0.487	0.475	5	2	1	2
4	0.457	0.457	5	2	1	2
5	0.417	0.433	0	2	2	6

6	0.46	0.48	4	1	3	2
7	0.45	0.425	1	0	5	4
8	0.567	0.367	1	5	1	3
9	0.55	0.45	4	3	0	3
10	0.4	0.5	2	2	3	3
Observed total			26	24	24	26
		q'_R	0.26	0.24	0.24	0.26
		q_R	0.219	0.26	0.24	0.281
		Δq_R	0.04	-0.02	0	-0.02
		$\ln(q'_R / q_R)$	0.172	-0.08	0	-0.078
J_{RPTI} (p-value)	0.01 (0.79)					
J_{PTI} (p-value)	0.002 (0.98)					

q'_R : observed mating frequencies; q_R : expected random mating frequencies computed following the model in equation (1) by taking equal propensities; Δq_R : $q'_R - q_R$. $\ln$: natural logarithm.

3.2. Non-random mating: Sexual isolation

As a putative assortative mating trait, we considered the length polymorphism of microsatellite DNA related with *Oxtr*. Thus, we may classify the matings by long (*L*) versus short (*S*) length polymorphism. The population size, mating rounds and initial frequencies were defined as in the previous example, so, $p_{1L} = p_{2L} = 0.5$ and $p_{1S} = p_{2S} = 0.5$. As before, there would be 100 mating rounds under the individual encounter model and at least 10 mating rounds under the mass encounter model. In the latter, there are 10 encounters per round but some of them may not succeed, so more than 10 mating rounds would be necessary for completing 100 matings.

To model assortative mating, the mutual propensities were different and we defined the homotypic matings having three times more propensity than the heterotypic, so $m_{LL} =$

$m_{SS} = 3m_{LS}$ and $m_{SL} = m_{LS}$. The specific value of the maximum propensity depended on the encounter model i.e. $m_{LL} = 1$ for the individual encounter and $m_{LL} = 10$ for the mass encounter model.

3.2.1 Mass encounter

In each mating round, 10 random encounters occurred based on the population frequencies. Once the encounter happened, the mating succeeds if a uniform $U(0,1)$ random number was lower than the pair propensity normalized by M_r , if so, the mating was recorded. The frequencies were updated after the 10 encounters. As in the random mating case, the simulation ended when 100 matings pairs were formed.

With the recorded frequencies we computed J_{RPTI} and J_{PTI} , and performed the corresponding tests, which for the specific mass encounter example shown in Table 2, were significant for J_{RPTI} ($J_{RPTI}=0.13$, p -value = 0.004) but non-significant for J_{PTI} ($J_{PTI} = 0.04$, p -value = 0.256).

TABLE 2. Result of 14 mating rounds under a mass encounter model with assortative mating: $m_{LL} = m_{SS} = 10 = 3m_{LS}$, $m_{SL} = m_{LS}$.

Round r	p_{1rL}	p_{2rL}	$L \times L$	$L \times S$	$S \times L$	$S \times S$
1	0.5	0.5	3	3	0	1
2	0.473	0.505	3	2	3	2
3	0.47	0.494	3	2	1	1
4	0.447	0.487	2	3	2	2
5	0.433	0.493	3	0	2	3
6	0.441	0.475	2	1	1	6
7	0.469	0.510	2	1	1	1

8	0.455	0.500	0	2	0	3
9	0.462	0.564	3	0	1	2
10	0.455	0.545	3	2	0	2
11	0.385	0.577	2	1	1	3
12	0.368	0.632	3	1	2	3
13	0.300	0.700	1	2	5	1
14	0.0	1.0	0	0	1	0
Observed total			30	20	20	30
		q'_R	0.3	0.2	0.2	0.3
		q_R	0.214	0.19	0.356	0.24
		Δq_R	0.086	0.01	-0.156	0.06
		$\ln(q'_R/q_R)$	0.338	0.051	-0.577	0.223
J_{RPTI} (p-value)	0.13 (0.004)	J_{RPSI} (p-value)	0.07 (0.007)			
J_{PTI} (p-value)	0.04 (0.256)					

q'_R : observed mating frequencies; q_R : expected random mating frequencies computed following the model in equation (1) by taking equal propensities; Δq_R : $q'_R - q_R$. $\ln$: natural logarithm.

In general, the power for detection of non-random mating was about 90% for J_{RPTI} and 73% for J_{PTI} (over 10,000 runs). In the case of detection of sexual isolation effects, the J_{RPSI} test had a power of 87% while a remaining 3% corresponded to the miss-detection of sexual selection (J_{RPSS}) instead of isolation. However, in the case of J_{PSI} all the non-random mating detected cases were identified as sexual isolation (73%).

3.2.2 Individual encounter

In the case of the individual encounter, the differences between the indices for detection of non-random mating were lower, with 81% of detection by J_{RPTI} against 74% by J_{PTI} .

However, in the case of the detection of sexual isolation, the J_{RPSI} test had 0% power while J_{PSI} had 74% (all the non-random mating detected cases).

As mentioned before, when we want to distinguish sexual isolation versus sexual selection effects, we need to identify the expectation P'_{Rij} of random mating from the frequencies in mated individuals. This expectation is formalized by equation (3) and P'_{Rij} is usually estimated by counting the different phenotypes within the matings.

However, a problem arises if we separate the counting in different mating rounds and there are very few mating per round. In such case, it is difficult to capture the marginal effect of each phenotype. Obviously, the individual encounter model is the extreme worst scenario since we only can assign a value of 1 for one of the matings at each round. Indeed, performing the estimation of P'_{Rij} in that way, under the individual encounter model, corresponded to the relative frequency of observed matings after the breeding season, i.e. q'_{Rij} . Thus, the value of RPSI_{ij} was 1 and so J_{RPSI} was 0, which explains the observed 0% power for the J_{RPSI} test under the individual encounter model.

There are different ways to avoid this problem. Ideally, if we have independent estimates of the mutual propensities we can compute directly P'_{Rij} by (3). When we do so, the 100% of the effects were detected as sexual isolation by J_{RPSI} (i.e. 81% power).

Nevertheless we may not have the mutual propensity estimates and so, we should group the information of different rounds to obtain reliable estimates of P'_{Rij} . As a rule of thumb, combining rounds so that the number of matings equals the product of the phenotypic classes ($k_1 \times k_2$) should be enough to provide more power than the constant frequency test. Alternatively, we could use the constant frequency test that still provides a power of 74% for detecting sexual isolation even under the individual encounter model.

Therefore, the constant frequency test (J_{PTI}) seems to be a good conservative control in this case, because it is not expected that the partition of J_{PTI} detect sexual isolation when the more powerful J_{RPTI} detects sexual selection instead. If that happens, we should concern about the J_{RPSI} estimate, especially if the number of matings per round is lower than $k_1 \times k_2$.

3.3. Non-random mating: Sexual selection

Here we considered the same length polymorphism trait so that females with long length polymorphism had more mating vigor. The population size, mating rounds and initial frequencies were defined as before.

To model female sexual selection, the mutual propensities were different so that L females had three times more propensity than the S ones, so $m_{LL} = m_{LS} = 3m_{SL}$ and $m_{SS} = m_{SL}$. The specific value of the maximum propensity depended on the encounter model i.e. $m_{LL} = 1$ for the individual encounter and $m_{LL} = 10$ for the mass encounter model.

In Table 3 we see a specific mass encounter example, with significant J_{RPTI} ($J_{RPTI} = 0.123$, p -value = 0.007) but non-significant J_{PTI} ($J_{PTI} = 0.006$, p -value = 0.887). Sexual selection was also detected.

TABLE 3. Result of 15 mating rounds under mass encounter model with female sexual selection: $m_{LL} = m_{LS} = 10 = 3m_{SL}$, $m_{SS} = m_{SL}$.

Round r	p_{1rL}	p_{2rL}	$L \times L$	$L \times S$	$S \times L$	$S \times S$

1	0.5	0.5	4	2	1	0
2	0.478	0.489	1	3	1	2
3	0.471	0.506	4	0	1	1
4	0.456	0.481	3	3	2	0
5	0.423	0.465	4	2	1	3
6	0.393	0.459	2	2	1	2
7	0.370	0.463	2	2	3	1
8	0.348	0.435	3	2	0	3
9	0.289	0.447	2	1	3	1
10	0.258	0.387	0	1	1	2
11	0.259	0.407	2	0	2	2
12	0.238	0.333	0	1	2	2
13	0.250	0.312	0	1	1	5
14	0.333	0.444	0	3	1	1
15	0.0	0.750	0	0	3	1
Observed total			27	23	23	27
		q'_R	0.27	0.23	0.23	0.27
		q_R	0.152	0.186	0.307	0.355
		Δq_R	0.118	0.044	-0.077	-0.085
		$\ln(q'_R/q_R)$	0.575	0.212	-0.289	-0.274
J_{RPTI} (p-value)	0.123 (0.007)	J_{RPSS} (p-value)	0.075 (0.023)			
J_{PTI} (p-value)	0.006 (0.887)					

q'_R : observed mating frequencies; q_R : expected random mating frequencies computed following the model in equation (1) by taking equal propensities; Δq_R : $q'_R - q_R$. $\ln$: natural logarithm.

Noteworthy, the generalized indices had a lot of more power for the detection of the sexual selection non-random mating effects than the constant frequency indices. Under

the mass-encounter model, the percentage of non-random mating (J_{RPTI}) was 73%, with 59% of female sexual selection, 3% male sexual selection, and 5% of sexual isolation. The constant frequency methods however were not able of detecting non-random mating effects (1% of detection by J_{PTI}). The results under the individual model were similar, the constant indices had no power (0.8%) and the new ones detected non-random mating with female sexual selection 60% of the time.

4. Discussion

In a previous work we had described the change in the mating phenotypes due to mate choice as the gain of information with respect to random mating. That model is valid for polygamous species, or even for monogamous when only a fraction of the population is involved in the breeding season. In such cases, the processes of the encounter and mating correspond to sampling with replacement of the mating phenotypes from the population (Carvajal-Rodríguez, 2018).

In the current work we have generalized the previous model to incorporate the case of monogamous species with small population size, so that the mating process resembles a sampling without replacement.

Notably, the change in the phenotypes can still be described by the flow of information between the random and non-random mating cases and it can be partitioned in the sum of the information due to sexual isolation and sexual selection, plus a mixed effect term. The new formulas become the same as for the corresponding constant ones as soon as the population frequency values are set to constant.

Before obtaining the new information indices we previously defined a generalized version of the PTI, PSI and PSS pairwise statistics (Rolán-Alvarez and Caballero, 2000). The new statistics measure the same as the previous ones (total change, sexual isolation and sexual selection, respectively) without requiring that the population frequencies are constant.

When the frequencies vary over different mating rounds, the tests performed by the new indices have more power than the constant ones which are more conservative. The difference becomes extreme for the sexual selection scenarios. In this case, the indices that assume constant frequency have almost no detection power, while the generalized indices have up to 70%. The reason is that the variation of population frequencies over the mating rounds, hides the sexual selection effect at the end of the breeding season. This happens because in the first rounds most of mating favour some phenotypes (e.g. females *L*) so that the encounters and mating for such phenotypes diminish at the end of the season.

We have tested the extreme case of monogamous species with low population size where all adults perform the mating. As the proportion of mating individuals decreases, the differences between the two kinds of indices disappear.

Last but not least, the application of the new statistics requires the estimation of frequencies at each mating round, which can be difficult especially under individual encounter-mating models. Simulations show that in the sexual isolation scenario, the previous indices still perform well under variable population frequencies and so, the use of one or another type of statistics will depend on the biological scenario as well as in the availability and easiness to split the sampling in more than one mating round.

Acknowledgement

I would like to thank Carlos Canchaya for their valuable comments on the manuscript.

This work was supported by Xunta de Galicia (Grupo de Referencia Competitiva, ED431C2016-037), Ministerio de Economía y Competitividad (CGL2016-75482-P) and by Fondos FEDER (“Unha maneira de facer Europa”).

450

References

- 451 Carvajal-Rodríguez, A., 2018. Non-random mating and information theory. Theoretical
452 Population Biology. 120, 103-113.
- 453 Carvajal-Rodríguez, A., Rolan-Alvarez, E., 2006. JMATING: a software for the analysis of sexual
454 selection and sexual isolation effects from mating frequency data. BMC Evol Biol. 6, 40.
- 455 Castelli, F. R., et al., 2011. Female prairie voles show social and sexual preferences for males
456 with longer avpr1a microsatellite alleles. Animal behaviour. 82, 1117-1126.
- 457 Getz, L. L., et al., 2004. Does mate choice take place in free-living prairie voles *Microtus*
458 *ochrogaster*? Evidence from field data. Acta Zoologica Sinica 50, 527-534.
- 459 Gimelfarb, A., 1988. Processes of Pair Formation Leading to Assortative Mating in Biological
460 Populations: Encounter-Mating Model. The American Naturalist. 131, 865-884.
- 461 Keane, B., et al., 2007. No effect of body condition at weaning on survival and reproduction in
462 prairie voles. Canadian Journal of Zoology. 85, 718-727.
- 463 Keane, B., et al., 2017. Effects of avpr1a length polymorphism on male social behavior and
464 reproduction in semi-natural populations of prairie voles (*Microtus ochrogaster*).
465 Ethology. 123, 675-688.
- 466 Rolán-Alvarez, E., Caballero, A., 2000. Estimating sexual selection and sexual isolation effects
467 from mating frequencies. Evolution. 54, 30-6.
- 468 Rosenthal, G. G., 2017. Mate choice: the evolution of sexual decision making from microbes to
469 humans. Princeton University Press.
- 470 Wang, H., et al., 2013. Histone deacetylase inhibitors facilitate partner preference formation in
471 female prairie voles. Nature Neuroscience. 16, 919-924.
- 472 Watts, P. C., et al., 2017. Stabilizing selection on microsatellite allele length at arginine
473 vasopressin 1a receptor and oxytocin receptor loci. Proceedings of the Royal Society B:
474 Biological Sciences. 284.
- 475 Williams, J. R., et al., 1992. Development of partner preferences in female prairie voles
476 (*Microtus ochrogaster*): The role of social and sexual experience. Hormones and
477 Behavior. 26, 339-349.
- 478 Young, K. A., et al., 2011. The neurobiology of pair bonding: Insights from a socially
479 monogamous rodent. Frontiers in Neuroendocrinology. 32, 53-69.