QInfoMating: Sexual Selection and Assortative Mating Estimation Software

A. Carvajal-Rodríguez

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

email: acraaj@uvigo.es

Abstract

Background: Sexual selection theory is a multifaceted area of evolutionary research that has profound implications across various disciplines, including population genetics, evolutionary ecology, animal behavior, sociology, and psychology. It explores the mechanisms by which certain traits and behaviors evolve due to mate choice and competition within a species. In the context of this theory, the Jeffreys divergence measure, also known as population stability index, plays a key role in quantifying the information obtained when a deviation from random mating occurs for both discrete and continuous data. Despite the critical importance of understanding mating patterns in the context of sexual selection, there is currently no software available that can perform model selection and multimodel inference with mating data. Recognizing this gap, I have developed QInfoMating which provides a comprehensive solution for analyzing and interpreting mating data within the framework of sexual selection theory.

Results: The program QInfoMating incorporates a user-friendly interface for performing statistical tests, best-fit model selection, and parameter estimation using multimodel inference for both discrete and continuous data. A use case is presented with real data of the species Echinolittorina malaccana.

Conclusions:

The application of information theory, model selection, and parameter estimation using multimodel inference are presented as powerful tools for the analysis of mating data, whether quantitative or categorical. The QInfoMating program is one of the few (if not the only) tools designed to perform this type of analysis.

Contact: acraaj@uvigo.es

Keywords: quantitative traits, sexual selection, assortative mating, Jeffreys divergence, population stability index, infomating.

Background

By this point, there is no doubt about the importance of sexual selection theory and its impact in areas such as population genetics, evolutionary ecology, animal behavior, and even sociology and psychology. There are several strategies for measuring the impact of sexual selection, depending on the field of study and whether the available data pertain to discrete or continuous traits. In previous works [1–3], it has been demonstrated that the symmetric Kullback-Leibler divergence (Jeffreys divergence or population stability index, noted as J_{PTI} in [1]) corresponds to the increase in information when mating is not random. This divergence can be broken down additively into one component that measures sexual selection and another that measures assortative mating, and these statistics can be applied both to discrete and continuous data. Therefore, J_{PTI} unifies the measure of sexual selection for both discrete and continuous traits [3].

The program QInfoMating is a major update to the Infomating program [1, 2] that incorporates the new quantitative methods developed in [3]. The previous version (Infomating version ≤0.4) have already been utilized in empirical studies [4, 5]. While there are other software options available for computing sexual selection and assortative mating from mating tables, such as JMating [6], to the best of my knowledge, QInfoMating is the only software for applying model selection theory to mating data, enabling the estimation of sexual selection and assortative mating for both discrete and quantitative data.

Implementation

The program can be used via the user interface (Figure 1-A) or directly from the console under Windows, Linux and macOS. Detailed instructions for use are provided in the

program manual, specifying input formats, the meaning of parameters, models, and result files [see Additional file 1]. In summary, QInfoMating offers the capability to analyze both continuous and discrete data for sexual selection and assortative mating. It begins by conducting relevant tests and subsequently identifies the best-fit model. The program employs multi-model inference techniques to estimate parameter values.

Difference of QInfoMating with the previous version (Infomating)

Unlike the previous version (Infomating), which lacked a user-friendly interface and only analyzed discrete data, QInfomating now accepts continuous data as input (see the format specification in the program manual). When continuous data is provided, the new statistical tests for sexual selection and assortative mating detection, as developed in [3], are performed. In this case, model selection is only executed if any of the statistical tests yield a significant result. Following this, the program automatically calculates the frequency distribution (classes) for the continuous data and proceeds with multi-model estimation. The algorithm to obtain the frequency distribution is detailed in the program manual.

There are also differences in the set of models considered for discrete and continuous cases. In the discrete case, there is a default set of models that can be modified by the user. Conversely, for the continuous case, the program, rather than the user, selects models based on the statistical significance of various tests. Specifically, if there is a statistically significant pattern of assortative mating but no sexual selection, the program uses a set of models that incorporate mate choice and bias in conjunction with random mating and the saturated model, but it does not include competition models. On the other hand, if there is a statistically significant pattern of both assortative mating and sexual selection in both sexes, QInfoMating includes double-effect models among the models to be tested. These models are not included by default when the data is discrete.

In any case, QInfoMating allows us to generate a mating table from the quantitative data, enabling us to create an input file for analysis in discrete mode, employing the default set of models or those selected by the user.

Results

Real data use case

Echinolittorina malaccana, a marine intertidal snail commonly found on rocky shores in the Indo-West Pacific region [7], exhibits mate selection behaviors. Males have shown a clear preference for females slightly larger than themselves [8]. In a subsequent study conducted by Lau et al. [5], the continuous trait of size (specifically, shell length) was transformed into categorical variables by creating four distinct size classes. This transformation allowed for the creation of a mating table (see Table 1), which enabled the utilization of the advanced multimodel inference techniques available in the previous version of the program (Infomating). This approach provided a more profound insight into the intricate dynamics of mate choice. The analysis ultimately confirmed that the most suitable model for mate choice in Echinolittorina malaccana involves a size bias, resulting in both assortative mating and the presence of sexual selection favoring larger individuals.

In a recent study (Carvajal-Rodríguez 2023), the same onshore data, originally presented in [5], and accessible via [9], was directly used without converting the continuous shell length trait into discrete classes. This approach aimed to demonstrate the computation of Jeffreys divergence indices [1] for quantitative traits. The research revealed that the statistical tests for the quantitative data yielded similar conclusions as the previous analysis using the statistical tests for discrete data. However, the

quantitative data lacked the necessary structure for applying the model selection methodology.

Table 1. Mating table for the Wild data used in Lau et al [5]. The data is classified in four classes.

Premating males numbers (421)			
23	181	167	50
Premating female numbers (455)			
23	146	190	96
Matings (228, rows are females)			
0	2	0	0
3	30	14	1
2	42	45	9
1	8	43	28

In this current study, we will leverage QInfoMating to conduct a quantitative analysis of the same data set. The program will apply the quantitative statistical tests and automatically transform data into classes to facilitate model selection and multi-model inference. This analysis aims to compare the results with those obtained in [5]. For consistency, we will employ the Akaike Information Criterion (AIC), the same criterion used in Lau et al., and additionally, the more conservative Bayesian Information Criterion (BIC) to showcase the software's versatility and utility.

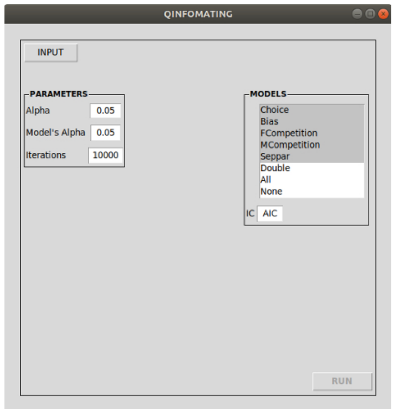
A CSV file containing shell length data for individual snails and their mate pairs served as the input for QInfoMating (referred to as QIM). Although the original study by Lau et al., used four classes, QIM segmented the data into three classes (Table 2), before proceeding with the model selection process.

Table 2. Mating table generated by QInfoMating for the Wild data in Lau et al [5]. The data is classified in three classes.

Premating males numbers (421)		
76	333	12
Premating female numbers (455)		
57	370	28
Matings (228, rows are females)		
4	6	0
18	163	2
1	30	4

If Figure 1-B the best-fit model from [5] (B-7P, bias with 7 parameters model) and the two best obtained by QIM (D-3P, double 3 parameters, and B-5P, bias 5 parameters models) are presented. The D-3P model is a model with double effect i.e., generates sexual selection and assortative mating patterns. From Figure 1-B we appreciate that the choice parameters (main diagonal) are c_1 and c_3. The parameter c_1 can also be considered a male sexual selection parameter jointly with c_2 since they change the marginal propensity for the second class males (second column). However, from the point of view of biased choice, the parameter c_2 can be seen as a bias (subdiagonal) in a model in which individuals in the shorter class have more difficulty mating. In fact, the second best model was B-5P, which is the equivalent model to B-7P when there are three classes instead of four.

A



B

B-7P model

$\mathbf{C}_1$	1	1	1
$\mathbf{B}_1$	$\mathbf{C}_2$	1	1
1	$\mathbf{B}_2$	$\mathbf{C}_3$	1
1	1	$\mathbf{B}_3$	$\mathbf{C}_4$

Best model in Lau et al.

0 ± 0	1 ± 0	1 ± 0	1 ± 0
2 ± 1	**2** ± 0.4	1 ± 0	1 ± 0
1 ± 0	**2** ± 0.3	**3** ± 0.4	1 ± 0
1 ± 0	1 ± 0	**5** ± 0.7	**12** ± 2

D-3P model

1	1	1
1	$\mathbf{C}_1$	1
1	$\mathbf{C}_2$	$\mathbf{C}_3$

QIM best model (28%)

1 ± 0	1 ± 0	1 ± 0
1 ± 0	**3** ± 0.1	1 ± 0
1 ± 0	**6** ± 1	**23** ± 11

B-5P model

$\mathbf{C}_1$	1	1
$\mathbf{B}_1$	$\mathbf{C}_2$	1
1	$\mathbf{B}_2$	$\mathbf{C}_3$

QIM model (21%)

3 ± 1	1 ± 0	1 ± 0
2 ± 0.4	**4** ± 0.1	1 ± 0
1 ± 0	**9** ± 2	**35** ± 17

Fig. 1. A. QInfoMating interface. **B.** Best fit models after re-analysis of Lau et al data. QIM: QInfoMating. The standard errors when minor than one were rounded to 1 decimal place.

The best model by Lau et al. had a weight of 99%, so it is almost exact to the multi-model inference parameter matrix. In the case of QIM, the best model (D-3P) had only 28% of the weight and other models such as B-5P (21%), D-5P (21%), D-4P (18%) and D-6P (8%), are necessary to capture more than 95% of the weight. The pattern obtained using the Bayesian Information Criterion (BIC) was similar to that obtained using the AIC criterion but with greater weight concentrated in the D-3P model (Table 3).

Table 3. Models that accumulate up to 95% of the weight. The parameter notation corresponds to the B-7P model in Figure 1B where C_i represent the ith cell on main diagonal and B_i on the subdiagonal.

Analysis	**Models (>95%)**	**Weight**	**Parameter estimation**						
			C_1	C_2	C_3	C_4	B_1	B_2	B_3
Lau et al 2021	B-7P	99%	0	2	3	12	2	2	5
QInfoMating-AIC	D-3P	28%	1	3	22	-	1	6	-
QInfoMating-AIC	B-5P	21%	3	4	35	-	2	9	-
QInfoMating-AIC	D-5P	21%	3	4	35	-	2	9	-
QInfoMating-AIC	D-4P	18%	2	3	24	-	1	6	-
QInfoMating-AIC	D-6P	8%	3	4	36	-	2	10	-
QInfoMating-BIC	D-3P	85%	1	3	23	-	1	6	-
QInfoMating-BIC	D-4P	10%	2	3	24	-	1	6	-
QInfoMating-BIC	B-5P	2%	3	4	35	-	2	9	-
QInfoMating-BIC	D-5P	2%	3	4	35	-	2	9	-

However, the structure of all these models is very similar: large males (rightmost columns) choose females larger (closest to the bottom rows) than themselves, especially for the largest classes (B_2 and B_3 parameters in the bias models, or c_2, c_3 in the D-3P model) .

The pattern behind the QIM models aligns with the findings reported by Lau et al. [5] for the Multiple-design and Wild data, as depicted in their Figure 4. However, there is a notable distinction: Lau et al. observed that the shorter class exhibited considerably lower mating ability. This difference may be attributed to the presence of their first class with a minimal number of individuals for both females and males, as documented in (Table 1).

However, one might wonder why QInfoMating did not clearly favor the B-5P model, which is most akin to the B-7P model chosen in the original analysis by Lau et al. The explanation lies in how QInfoMating operates when handling quantitative data. As explained before, the program selects models based on the statistical significance of various tests and have included double-effect models (as D-3P, D-4P and D-6P) among the set of models to be tested. These particular models were not considered in the original Lau et al. study.

To further elucidate, I excluded the double-effect models and repeated the analysis using exactly the same set of models as in Lau et al. Upon conducting this analysis, the best-fitting model becomes the B-5P, with a weight of 94%. The program's preference for the D-3P model, when included, stems from its ability to explain the data with two fewer parameters, thus aligning with the principle of parsimony. In cases where both models are closely weighted, the choice between them is influenced by the researcher's experience and the biological plausibility of the results. Nonetheless, it's worth noting that the fundamental structure of these models remains quite similar: larger males tend to prefer females larger than themselves, whenever such a pairing is possible.

Consistent with the result of models, the statistical tests performed by QInfoMating detect assortative mating and sexual selection both in females and males (Table 4).

Table 4. QInfoMating tests.

Test type	**Statistic**	**H_0**	**p-value**
Quantitative	$QJ_{PSI} = 0.49$	no assortative mating	$<10^{-6}$
Quantitative	$QJ_{PS1} = 0.24$	no female sexual selection	$<10^{-6}$
Quantitative	$QJ_{PS2} = 0.07$	no male sexual selection	0.0004
Discrete	$J_{PSI} = 0.08$	no assortative mating	0.0009
Discrete	$J_{PS1} = 0.17$	no female sexual selection	$<10^{-6}$

Discrete	$J_{PS2} = 0.06$	no male sexual selection	0.002

Discussion

Assortative mating based on size and other quantitative traits is common in many species, from gastropods [8] to humans [10]. The processes leading to assortative mating may not solely rely on similarity-based mechanisms and may not affect different phenotypic classes equally. This is where the power of model selection methodology becomes evident, as it doesn't require proposing a single ad hoc model. Instead, it allows best fit of data for a wide range of models, enabling researchers to study non-trivial mating patterns, some of which may not have been obvious before the experiment.

In conclusion, the application of information theory, model selection, and parameter estimation using multimodel inference are presented as powerful tools for the analysis of mating data, whether quantitative or categorical. QInfoMating is software designed to perform this type of analysis.

Availability and requirements

Project name: QInfomating

Project home page: https://acraaj.webs.uvigo.es/InfoMating/QInfomating.htm

Operating system(s): Windows, Linux, macOS

Programming language: C++, Python.

License: GNU GPL.

Declarations

- Ethics approval and consent to participate

  Not applicable

- Consent for publication

  Not applicable

- Availability of data and materials

  The datasets generated and/or analysed during the current study are available in the DRYAD repository, http://doi.org/10.5061/dryad.h214h8t

- Competing interests

  The authors declare that they have no competing interests

- Funding

  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”)

- Authors' contributions

  ACR is the sole author of this work.

- Acknowledgments

  Dedicated to my sister Paci. I thank E. Rolán-Alvarez for comments on the manuscript.

References

1. Carvajal-Rodríguez A. Non-random mating and information theory. Theor Popul Biol. 2018;120:103–13.

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

3. Carvajal-Rodríguez A. The information theory formalism unifies the detection of the patterns of sexual selection and assortative mating for both discrete and quantitative traits. 2023;:2023.08.14.552693.

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

5. Lau SLY, Williams GA, Carvajal-Rodríguez A, Rolán-Alvarez E. An integrated approach to infer the mechanisms of mate choice for size. Anim Behav. 2021;175:33–43.

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

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

8. Ng TPT, Rolán-Alvarez E, Dahlén SS, Davies MS, Estévez D, Stafford R, et al. The causal relationship between sexual selection and sexual size dimorphism in marine gastropods. Anim Behav. 2019;148:53–62.

9. Rolán-Alvarez E, Williams G, Ng T, Estévez-Barcia D, Davies M, Saltin S, et al. Data for: The causal relationship between sexual selection and sexual size dimorphism in marine gastropods. 2021;1.

10. Robinson MR, Kleinman A, Graff M, Vinkhuyzen AAE, Couper D, Miller MB, et al. Genetic evidence of assortative mating in humans. Nat Hum Behav. 2017;1:1–13.

Additional files

Additional file 1.

- PDF file.
- QInfoMating Manual v. 1.0.
- Complete manual for the QInfomating program with detailed explanations about its use, arguments, input files, tests and statistical models, etc.