Maximum likelihood estimation of fitness components in experimental evolution

Jingxian Liu[1,2*], Jackson Champer[1,2*+], Chen Liu[1,2], Joan Chung[1,2], Riona Reeves[1,2], Anisha Luthra[1,2], Yoo Lim Lee[1,2], Andrew G. Clark[1,2], Philipp W. Messer[1+]

[1]Department of Biological Statistics and Computational Biology,
[2]Department of Molecular Biology and Genetics,
Cornell University, Ithaca, NY 14853

*equal contribution
+corresponding authors: jc3248@cornell.edu or messer@cornell.edu.

Abstract

Estimating fitness differences between allelic variants is a central goal of experimental evolution. Current methods for inferring selection from allele frequency time series typically assume that evolutionary dynamics at the locus of interest can be described by a fixed selection coefficient. However, fitness is an aggregate of several components including mating success, fecundity, and viability, and distinguishing between these components could be critical in many scenarios. Here we develop a flexible maximum likelihood framework that can disentangle different components of fitness and estimate them individually in males and females from genotype frequency data. As a proof-of-principle, we apply our method to experimentally-evolved cage populations of *Drosophila melanogaster*, in which we tracked the relative frequencies of a loss-of-function and wild-type allele of *yellow*. This X-linked gene produces a recessive yellow phenotype when disrupted and is involved in male courtship ability. We find that the fitness costs of the yellow phenotype take the form of substantially reduced mating preference of wild-type females for yellow males, together with a modest reduction in the viability of yellow males and females. Our framework should be generally applicable to situations where it is important to quantify fitness components of specific genetic variants, including quantitative characterization of the population dynamics of CRISPR gene drives.

Introduction

The concept of fitness lies at the core of Darwin's theory of evolution by natural selection. If two allelic variants differ in fitness, the fitter allele should increase in population frequency over time at the expense of the less fit allele. The study of temporal changes in allele frequencies can therefore allow us to infer relative fitness differences. Various approaches have been developed to this end that can disentangle such selection from the stochastic frequency fluctuations generated by random genetic drift.[4, 15–18, 23–25, 27, 29, 35–37, 40–42]

The general principle of these methods is to devise a probabilistic model for expected allele frequency dynamics, often based on a Wright-Fisher process. Comparison of model predictions with empirical allele frequency measurements then allows inference of model parameters (which may include effective population size) using statistical approaches such as maximum likelihood (ML) or Bayesian inference. Applications of these methods have provided key insights into the strength of selection operating on particular alleles.[26, 30]

A common assumption in these methods is that selection between two allelic variants can be described by a single selection coefficient. While often reasonable, it will be critical in many scenarios to distinguish among the individual components that constitute fitness.[9, 33, 39] For example, preferential mate choice cannot typically be mapped onto a logistic growth model specified by a fixed selection coefficient, since it could result in frequency-dependent dynamics.[2] Individual fitness effects could also often differ between males and females, which would lead to systematic biases in genotype frequencies between sexes when selection is sufficiently strong.

Rigorous inference of selection can include four major selection components: zygotic selection (viability from zygote to adult), sexual selection, fecundity selection, and gametic selection that usually involves genetic elements with biased inheritance.[9, 33] Evaluation of these four selection components in a natural population

1

requires detailed population monitoring, as observations need to be recorded at four different life-cycle stages.[9, 33, 34, 38] Typically, the analysis of selection components in the laboratory involves performing a series of isolated experiments designed to individually quantify each component.[5] For example, progeny counts from controlled crosses and backcrosses can reveal differences in zygote-to-adult viability as measured by deviation from Mendelian proportions.[32] Egg counts from controlled crosses can reveal genotype-specific differences in fecundity. Mating arenas with either observed matings or subsequent scoring of progeny can allow for estimation of sexual selection.[12] The challenge with all of these methods is that it is extremely difficult to know if all attributes of a gene that may result in differential propagation have been considered, although it is possible to test the correspondence between individual fitness components estimates and allele frequency dynamics in cage populations.[10, 11]

In this study we develop a ML inference framework that can disentangle sexual, fecundity, and viability selection from genotype frequency time series data. Our analytical approach employs a continuous extension of the multinomial distribution, allowing us to infer effective population size simultaneously with selection parameters. As a proof-of-principle, we apply our inference framework to empirical data obtained from cage evolution experiments of *Drosophila melanogaster*, in which we tracked relative genotype frequencies of a wild-type and mutant version of *yellow*, an X-linked gene required for black pigment formation.[14] The mutant allele disrupts *yellow*, resulting in a recessive yellow body phenotype. The *yellow* gene is also required for the wing extension behavior that is part of male courtship in *D. melanogaster*.[3, 44] Consequently, disruption of *yellow* affects mating competitiveness in male flies.[13, 21]

We find that in cage experiments, our ML inference method can robustly distinguish between several selection models involving preferential mate choice, fecundity, and viability, and further distinguish between male and female fitness costs. Such measurements should become particularly important in scenarios where fitness costs are large and allele frequencies are expected to change rapidly, as will likely be the case for recently proposed CRISPR gene drive approaches.[6]

Methods

Plasmid construct design and generation of transgenic fly line

We designed a construct targeting the X-linked *yellow* gene in *D. melanogaster*. Disruption of this gene causes a recessive yellow phenotype, specified by a lack of dark pigment in the adult cuticle. Our construct additionally encodes a dsRed protein driven by a 3xP3 promoter, which produces an easily identifiable fluorescent phenotype. dsRed is not visible in wild-type eyes in the Canton-S background, but we observed dsRed expression in the abdomen of younger insects and in the ocelli.

The donor plasmid BHDyR was constructed by Gibson assembly of the restriction digest of IHDyV1[8] with *Stu*I and *Xho*I and PCR amplification of the pDsRed-attp plasmid[20] with the oligos dsRedY_F: GGGTT-TTGGACACTGGGAATTCTTGCATGGCTAGACGAAGTTATCGTACGGGATCTAAT and dsRedY_R: TTAGTGGTGGTATTGCCGATGCCCACGGACGCGCCGGTTAAGATACATTGATGAGTTTGG (IDT). Gibson assembly of plasmids was performed with Assembly Master Mix (New England Biolabs), and plasmids were transformed into JM109 competent cells (Zymo Research).

The transgenic line in the study was transformed at GenetiVision by injecting the BHDyR donor plasmid into Canton-S *D. melanogaster* embryos. Cas9 was provided by co-injection with plasmid pHsp70-Cas9[19] (a gift from Melissa Harrison, Kate O'Connor-Giles, and Jill Wildonger, Addgene plasmid 45945), and a gRNA plasmid (BHDyg1)[8] was also included in the injection. Concentrations of donor, Cas9, and gRNA plasmids were 98, 94, and 58 ng/μL, respectively, in 10 mM Tris-HCl, 23 μM EDTA, pH 8.1 solution. The injected plasmids were purified with ZymoPure Midiprep kit (Zymo Research). To obtain a homozygous fly line, the injected embryos were reared and crossed with wild-type Canton-S flies. The progeny with dsRed fluorescent protein in the eyes and abdomen, which indicated successful insertion of the construct, were selected and crossed with each other. The stock was considered homozygous when all male progeny had dsRed fluorescence for two consecutive generations.

Fly rearing and phenotyping

Flies were reared at 25 °C with a 14/10 hr day/night cycle. Bloomington Standard medium was provided as food every two weeks. During initial phenotyping, flies were anesthetized with CO_2 and examined with a stereo dissecting microscope, and the fluorescent red phenotype was observed using the NIGHTSEA system.

For cage studies, enclosures of internal dimensions 30x30x30 cm (Bugdorm, BD43030D) were used to house flies. At the start of an experiment, transgenic and Canton-S flies of approximately the same age were separately allowed to lay eggs in food bottles for two days. These food bottles (nine for cage 1, eight for cage 2, five for Cages 3-5) were then placed in cages at the desired starting ratio between transgenics and Canton-S flies. Eleven days later, bottles were replaced in the cage with fresh food (at a 1:1 ratio for cage 1 and 2:1 for other cages), leaving adult flies. Two days later, bottles were removed again from the cages and flies retained, and fresh food bottles were added (nine for cage 1, eight for cage 2, five for cages 3-5). One day later, flies were frozen for later phenotyping, and food bottles were retained for eleven days to allow the next generation to hatch. This cycle was repeated until the completion of the experiments.

Results

Cage evolution experiments

We created a transgenic *D. melanogaster* line based on the Canton-S strain, in which we inserted a construct expressing a dsRed fluorescent protein into the *yellow* gene. This insertion effectively disrupts the gene, producing the characteristic yellow phenotype, which is recessive. At the same time, expression of the dsRed protein produces an easily identifiable, co-dominant phenotype characterized by red fluorescence in the eyes, abdomen, and ocelli, though only the latter two could be observed in Canton-S flies due to eye pigmentation.

We will refer to the allele with the inserted construct as y, while we will denote the wild-type allele as $+$. Because *yellow* is located on the X chromosome, males of genotype y are expected to show both the yellow (body) and red (eye) phenotypes, whereas $+$ males are expected to show neither phenotype. In females, yy homozygotes are expected to show both the yellow and red phenotypes, $y+$ heterozygotes are expected to show only the red phenotype, and $++$ homozygotes are expected to show neither phenotype. Our system thereby allows us to unambiguously ascertain genotypes at the *yellow* locus from phenotypic assays of individual flies.

We evolved five laboratory cage populations, each initialized from a mixture of wild-type Canton-S flies and our transgenic yellow flies. Genotype frequencies at the *yellow* locus were then tracked over the course of several generations. Our first two cages were evolved over six (cage 1) and five (cage 2) generations, starting with initial y allele frequencies of ∼70%. Three additional cages were each evolved for just a single generation with starting frequencies of the y allele of ∼20% for cage 3, ∼50% for cage 4, and ∼80% for cage 5. Each cage comprised between several hundred and several thousand flies. The observed genotype frequency trajectories and changes in male and female population size over time are shown in Figure 1.

The frequency of the y allele decreased systematically in all five cages, confirming that *yellow* disruption is likely associated with a substantial fitness cost. In cage 1, for example, the y allele decreased from ∼70% to less than 10% over the course of just six generations, and a broadly consistent rate of decline was observed among the other cages as well.

Census population sizes fluctuated noticeably in our cages. However, we do not believe that these fluctuations are the result of selection at the *yellow* locus, but rather reflect unrelated factors, such as larval competition and variance in food preparation. We also usually observed a bias in sex ratio of phenotyped flies since females had higher mortality from laying eggs in the crowded food bottles, while males approached the food bottles less often.

Evolutionary model

The observation of rapidly declining y-allele frequency in all five population cages raises the possibility that we may have statistical power to estimate individual fitness components associated with *yellow* disruption. To explore this possibility, we will first establish an evolutionary model for allele frequency dynamics in our experiments that incorporates three specific components of fitness: viability, fecundity, and mating success.
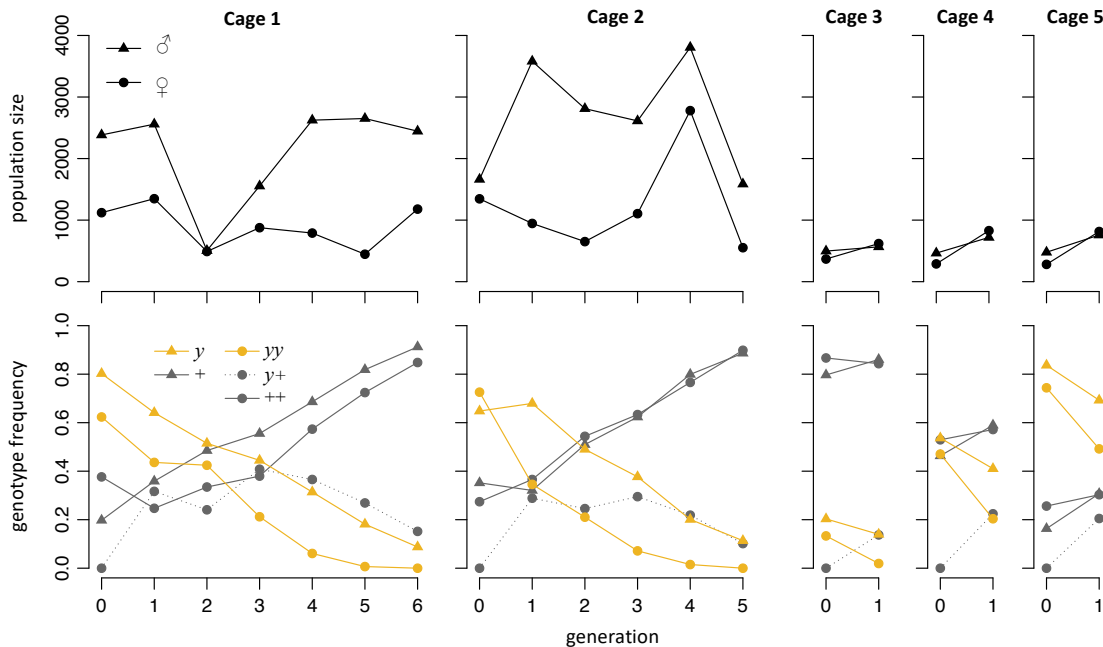
3

Figure 1: Population sizes and genotype frequencies in our five cage experiments. Note that we define genotype frequencies relatively among males, and relatively among females.

Consider an X-linked locus with two segregating alleles: wild-type (+) and yellow (y). The y allele produces the yellow phenotype, which we assume has lower fitness than the wild-type phenotype. These fitness costs could be due to reduced viability, fecundity, and/or mating success in individuals with yellow phenotype compared to wild-type individuals. In the following, we will describe how each of these three different fitness components is incorporated into our model. In all cases, we assume that the y allele is strictly recessive, so that wild-type and $y+$ heterozygous females have the same fitness. We will denote the counts of individuals observed in the present generation t, partitioned by genotype, as $N_y, N_+, N_{yy}, N_{y+}, N_{++}$ (the first two are males, the last three are females).

Viability specifies the probability that a zygote survives to reproductive age, which we assume will depend only on the phenotype and sex of the individual. Specifically, we define v_m to be the relative viability of yellow males compared to wild-type males, and v_f to be the relative viability of yellow females compared to wild-type females:

$$\frac{v(\male)}{v(\male)} = v_m \tag{1}$$

$$\frac{v(\female)}{v(\female)} = v_f. \tag{2}$$

Using relative viabilities among males and females, rather than absolute viabilities, can help prevent systematic biases due to the unequal sex ratios in our cages.

Fecundity specifies the reproductive success of a mating pair, as measured by the expected number of offspring. In contrast to viability, which is a function of an individual zygote, fecundity is therefore a function of a mother-father pair. In our model, we assume that fecundities are defined by the following mating table, specifying the relative fecundities of matings involving parents with yellow phenotype, compared to matings of wild-type mothers with wild-type fathers:

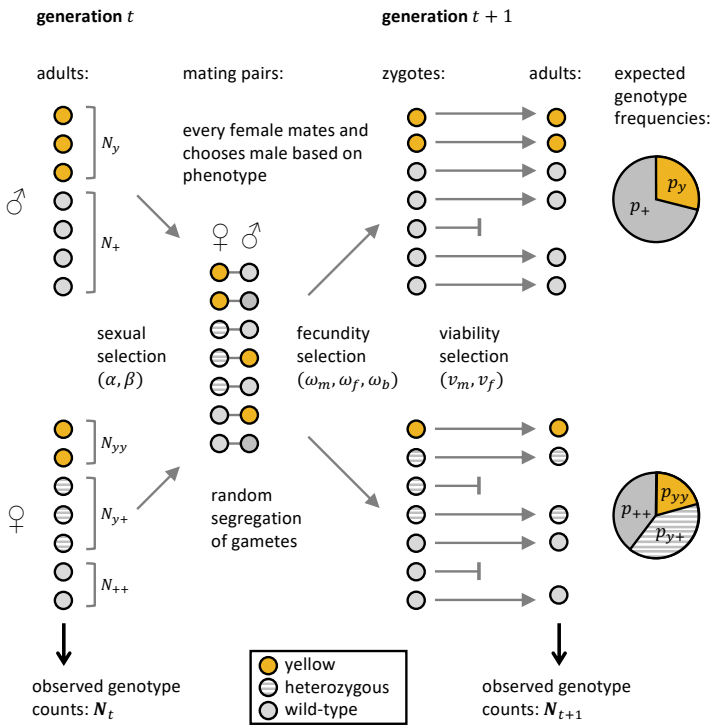|	$\female$	$\female$
$\male$	1	ω_f
$\male$	ω_m	ω_b

4

Figure 2: Illustration of the different stages of the life cycle during one generation in our model.

For our sexual selection model, we assume a scenario of female mate choice, in which females will choose mates with a potential preference based on phenotype. We define the actual probabilities that a female of a particular phenotype chooses a male of a particular phenotype to be:

$$
\begin{aligned}
P(\male|\female) &= N_+/(N_+ + \alpha N_y) \\
P(\male|\female) &= \alpha N_y/(N_+ + \alpha N_y) \\
P(\male|\female) &= N_+/(N_+ + \beta N_y) \\
P(\male|\female) &= \beta N_y/(N_+ + \beta N_y).
\end{aligned}
\tag{3}
$$

Here the parameters α and β specify the relative reduction in mating success of yellow males compared to wild-type males in mating with wild-type or yellow females, respectively.

Biologically there is no reason to assume that yellow phenotype males or females would actually have higher viability, fecundity, or mating success than their wild-type counterparts. Thus, we will assume that all of our selection parameters are constrained within $0 \leq v_m, v_f, \omega_m, \omega_f, \omega_b, \alpha, \beta \leq 1$, which will also help prevent overfitting in our ML inference approach. Finally, we assume discrete generations and random segregation of alleles within gametes due to a lack of any known mechanism for biased inheritance of y. Figure (2) illustrates the life cycle in this evolutionary model over the course of one generation, depicting where the different types of selection operate.

With these definitions at hand, we can calculate the expected values of genotype frequencies in generation $t+1$ as a function of their frequencies in generation t and the parameters of the evolutionary model. At our X-linked locus, male zygotes always inherit their X chromosomes from their mothers, while females inherit one copy from each parent. When assuming absence of any selection ($v_m = v_f = \omega_m = \omega_f = \omega_b = \alpha = \beta = 1$), the expected genotype frequencies in the next generation will then be given by:

$$
\begin{aligned}
p_y &= c_1(N_{yy} + N_{y+}/2) \\
p_+ &= c_1(N_{++} + N_{y+}/2) \\
p_{yy} &= c_2 N_y(N_{yy} + N_{y+}/2) \\
p_{y+} &= c_2[N_y(N_{++} + N_{y+}/2) + N_+(N_{yy} + N_{y+}/2)] \\
p_{++} &= c_2 N_+(N_{++} + N_{y+}/2).
\end{aligned}
\tag{4}
$$

5

Here p_y and p_+ denote the expected relative frequencies of y and $+$ genotypes among males, and p_{yy}, p_{y+}, and p_{++} denote the expected relative frequencies of yy, $y+$, and $++$ genotypes among females. The normalization coefficients c_1 and c_2 are defined by the conditions that gamete frequencies in males and females each have to sum to one: $p_y + p_+ = p_{yy} + p_{y+} + p_{++} = 1$. Note that we do not use absolute genotype frequencies in the population, but only relative frequencies in males and females, again to prevent any systematic biases due to unequal in sex ratios in our cages.

Viability selection is straightforward to incorporate into this framework: we can simply interpret the genotype frequencies derived in (4) as the expected frequencies of zygotes, which then just need to be multiplied by the corresponding viability coefficients (and re-normalized). To incorporate fecundity and mate choice, however, we also have to incorporate the specific probabilities that each mating that produces the respective zygote genotype actually occurs, and then weigh them by their fecundities. Altogether, this yields our full model:

$$
\begin{aligned}
p_y &= c_1 v_m ([\omega_b P(\male|\female) + \omega_f P(\male|\female)] N_{yy} + [\omega_m P(\male|\female) + P(\male|\female)] N_{y+}/2) \\
p_+ &= c_1 [\omega_m P(\male|\female)(N_{++} + N_{y+}/2) + P(\male|\female)(N_{++} + N_{y+}/2)] \\
p_{yy} &= c_2 v_f [\omega_b P(\male|\female) N_{yy} + \omega_m P(\male|\female) N_{y+}/2] \\
p_{y+} &= c_2 [\omega_f P(\male|\female) N_{yy} + \omega_m P(\male|\female)(N_{++} + N_{y+}/2) + P(\male|\female) N_{y+}/2] \\
p_{++} &= c_2 P(\male|\female)(N_{++} + N_{y+}/2).
\end{aligned}
\tag{5}
$$

The coefficients c_1 and c_2 are again defined by normalization. Note that although this model is specific for an X-linked gene, it is in fact easy to extend to other inheritance patterns, such as an autosomal locus. The general principle is always the same: to calculate the expected frequency of a given genotype, we have to sum over all matings that can produce this genotype, weight each by its respective mating probability and expected fecundity of the mating pair, then multiply all genotypes by their respective viabilities, and finally normalize all genotype frequencies. Other selection scenarios, such as different dominance effects, could be easily implemented as well by changing the genotypes to which fitness parameters are applied.

ML inference framework

The evolutionary model developed above allows us to calculate the expected changes in genotype frequencies across generations for a given set of model parameters. In turn, we may be able to use these relations for inferring the model parameters, given observed genotype frequency changes. To perform this inference in a ML framework, we require a probabilistic model that can specify the actual probabilities of observing a given set of genotype frequencies in the next generation, which will fluctuate around their expected values because of random genetic drift.

We will formally describe these fluctuations using a diploid Wright-Fisher model, where each individual in generation $t+1$ has its genotype sampled randomly from a multinomial probability distribution, specified by the expected genotype frequencies from Eq. (5) and an effective population size (N_e) controlling the level of random genetic drift in the population. Note that we expect the effective population size of our cages to be considerably smaller than their census sizes due to a variety of factors, such as a higher variance in offspring numbers among the flies in our cages (especially males) compared to a Wright-Fisher population.

There is an important complication when it comes to calculating probabilities in this model: In a Wright-Fisher population of effective size N_e, the probability of observing a given vector $(N_1, \cdots, N_d)$ of counts of all d possible genotypes is specified by a multinomial distribution, which is properly defined only when $\sum_{i=1}^{d} N_i = N_e$. However, this condition will typically be violated in our cages, as we expect sample sizes (constituted by all individuals present in a cage in our experiments) to be much larger than N_e.

To address this problem, we will employ a continuous approximation for the multinomial distribution, where genotype counts are replaced by genotype frequencies and the discrete multinomial probabilities are replaced by a multi-dimensional probability density. We define this density by using the gamma function $\Gamma(x+1) = x!$ as a natural continuous extension for the factorials in the multinomial distribution. Specifically, consider a diploid Wright-Fisher model with effective population size N_e and d possible genotypes, where $\mathbf{p} = (p_1, \cdots, p_d)$ specifies their expected frequencies in generation $t+1$. The probability density of observing

a given set of genotype frequencies $\mathbf{x}_{t+1} = (N_1/N, \cdots, N_d/N)$ in generation $t+1$ is then defined by:

$$\rho(\mathbf{x}_{t+1}|\mathbf{p}, N_e) = \alpha\Gamma(N_e + 1)\prod_{i=1}^{d}\frac{p_i^{x_i N_e}}{\Gamma(x_i N_e + 1)}. \tag{6}$$

The constant α is to ensure normalization when ρ is integrated over the d-dimensional standard simplex defined by the constraint $\sum_i x_i = 1$. In this framework, the probability of observing genotype frequencies that fall inside a given area Ω of the d-dimensional frequency space is then obtained by integrating the probability density over this area: $\Pr(\mathbf{x} \in \Omega) = \int_{\Omega}\rho(\mathbf{x})\mathrm{d}\mathbf{x}$.

Unfortunately we are not aware of any closed-form analytic expressions for the normalization coefficient α, except for the special case $d = 2$, but we can derive a workable approximation by discrete partitioning. Specifically, we will count the number of possible frequency vectors assuming that frequencies x_i can still only take on discrete values $x_i \in \{0, 1/N_e, 2/N_e, \cdots, 1\}$ on the simplex $\sum_i x_i = 1$. This number is given by the binomial coefficient

$$\alpha = \binom{N_e + d - 1}{d - 1}. \tag{7}$$

Note that under this choice of α, the function $\rho(\mathbf{x})$ is technically no longer a proper probability density, as integration over the continuous frequency simplex does not always yield a value of exactly one. This is because our approximation of α is still based on a discrete partitioning of the frequency space, which should converge to the continuous limit only as N_e goes to infinity. However, these deviations become noticeable primarily for very small values of N_e, where the applicability of the Wright-Fisher model becomes questionable in general. Our simulations below also confirm that this discrete approximation works well in practice.

To allow for the possibility of biased sex ratios in our model, while avoiding specific assumptions about what caused these biases in our cages, we will further factorize the overall probability density of observed genotype counts into the individual probabilities of observing the male and female genotype counts independently:

$$\rho(\mathbf{x}_{t+1}) = \rho(x_y, x_+|p_y, p_+, N_{e,m}) \times \rho(x_{yy}, x_{y+}, x_{++}|p_{yy}, p_{y+}, p_{++}, N_{e,f}). \tag{8}$$

Here x_y and x_+ specify the observed genotype frequencies in generation $t+1$ in males, and x_{yy}, x_{y+}, x_{++} specify the observed genotype frequencies in females. For males, we have $d = 2$, and the normalization factor in ρ is therefore $\alpha = (N_{e,m} + 1)$. For females, we have $d = 3$, and therefore $\alpha = (N_{e,f} + 2)(N_{e,f} + 1)/2$.

The factorization employed in definition (8) provides a very flexible framework for defining effective population size, which could be assumed to differ between males and females or could vary based on the census population size. We will discuss this issue in more detail below. For now, we want to assume the simplest model of a constant effective population size with equal sex ratio:

$$N_{e,m} = N_{e,f} = N_e/2. \tag{9}$$

With these definitions in place, we can formulate the likelihood density function for the parameter vector $\theta = \{N_e, v_m, v_f, \omega_m, \omega_f, \omega_b, \alpha, \beta\}$, given the vectors of observed genotype counts in generations t and $t+1$ of a population cage. Multiplying these individual likelihoods across all generations $t = 0, 1 \cdots, n$ in an experiment (i.e., assuming independence across generations) then yields the log-likelihood density function:

$$\ln L(\theta) = \sum_{t=1}^{n}\ln\rho(\mathbf{x}_t|\theta). \tag{10}$$

We can easily extend this approach across multiple cages by simply summing the log-likelihoods of the individual cages. The resulting log-likelihood density function can then be used for inference of maximum likelihood estimates (MLEs) of the parameter values and their confidence intervals.

Several aspects of the above approach are worth further mention: First, N_e has now become just another parameter of the model and can thus be inferred the same way as we do with the other parameters. This way, N_e can also take on continuous values in our model. Second, our ML inference approach is very general and could be easily applied to other evolutionary models as well, so long as the model provides expected values for genotype frequencies, given their frequencies in the previous generation and model parameters. Finally,

7

we want to acknowledge that we have not explicitly incorporated sampling variance into our framework, since we assume that the observed genotype counts reflect the actual numbers of individuals that were present in a cage. While some sampling noise could of course have occurred in any given generation of an experiment, we expect that the resulting variance in genotype frequencies should be negligible compared to the variance due to drift. This is also supported by the fact that we typically infer effective population sizes that are an order of magnitude smaller than the census sizes of our cages.

Power evaluation

Even when combined over all five cages, the genotype trajectories in our experiments comprise a total of only 14 transitions between consecutive generations. In light of such limited data, can we still expect our ML framework to have sufficient power to infer the different parameters of our model? To explore this question and also provide a general proof-of-principle of our inference framework, we first tested it on simulated data.

Figure 3A shows the outcomes of three Wright-Fisher simulation runs, in which we modeled genotype frequency trajectories over 10 generations in a cage under three different selection scenarios, differing in whether fitness costs of an X-linked allele were due to reduced mating success, reduced viability, or reduced fecundity. These initial simulations were run without drift (in the limit of "infinite" N_e) to assess whether we can detect systematic differences between the three scenarios. We tuned the selection strength in each model to roughly conform with the observed decay-rate of y allele frequencies in our experimental cages.

A

(i) $\alpha = 0.2, \beta = 0.8$ (ii) $v_m = 0.3, v_f = 0.6$ (iii) $\omega_m = 0.4, \omega_f = 0.8, \omega_b = 0.2$

B   MLEs, estimated over 1000 simulated runs ($N = 2000$, $N_e = 200$):

(i)	mean	s.d.
$\widehat{N}_e$	250	73
$\hat{\alpha}$	0.201	0.042
$\hat{\beta}$	0.785	0.141
$\ln\hat{L}$	59.07	5.88

(ii)	mean	s.d.
$\widehat{N}_e$	271	78
$\hat{v}_m$	0.297	0.030
$\hat{v}_f$	0.599	0.082
$\ln\hat{L}$	68.71	5.97

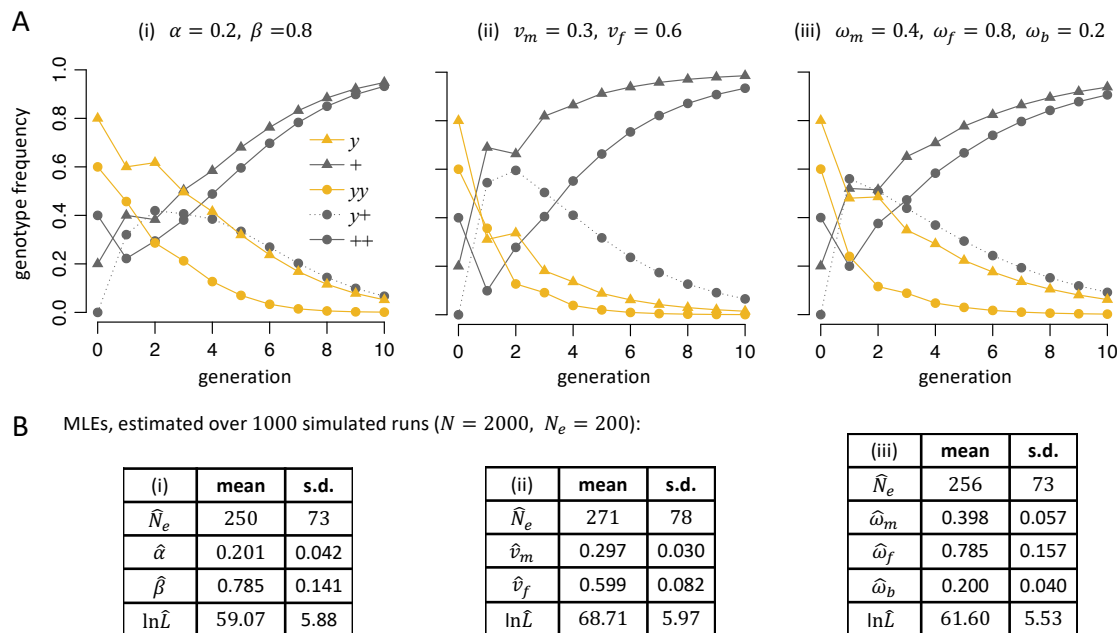| (iii)	mean	s.d.
$\widehat{N}_e$	256	73
$\hat{\omega}_m$	0.398	0.057
$\hat{\omega}_f$	0.785	0.157
$\hat{\omega}_b$	0.200	0.040
$\ln\hat{L}$	61.60	5.53

Figure 3: (A) Simulated genotype frequency trajectories under different selection scenarios in a deterministic model without drift: (i) sexual selection only ($\alpha = 0.2, \beta = 0.8; v_m = v_f = \omega_m = \omega_f = \omega_b = 1$); (ii) viability selection only ($v_m = 0.3, v_f = 0.6; \omega_m = \omega_f = \omega_b = \alpha = \beta = 1$); (iii) fecundity selection only ($\omega_m = 0.4, \omega_f = 0.8, \omega_b = 0.2; v_m = v_f = \alpha = \beta = 1$). (B) Accuracy of our ML inference framework for the three selection scenarios. Tables show the average parameter MLEs and standard deviations estimated over 1000 simulation runs of the specific scenario. These simulations were run with drift, assuming populations of 2000 individuals with $N_e = 200$. For each scenario, inference was performed by only inferring the parameters of the given model, while keeping the other parameters fixed at one. In all three scenarios, the parameter MLEs converge to their true values when averaging over individual runs.

The simulations reveal that there are indeed systematic differences in the expected genotype frequency trajectories between the three scenarios, even though the overall reduction in y-allele frequencies after 10 generations were similar. While differences are most pronounced in the early generations of a run, where

		Simulated model:		
		(i) sexual sel. only $\alpha = 0.2$ $\beta = 0.8$	(ii) viability sel. only $v_m = 0.3$ $v_f = 0.6$	(iii) fecundity sel. only $\omega_m = 0.4$ $\omega_f = 0.8$ $\omega_b = 0.2$
Model used for inference:	sexual sel. only	$\ln\hat{L} = 59.1$ $\hat{\alpha} = 0.20$ $\hat{\beta} = 0.79$	$\ln\hat{L} = 34.1$ $\hat{\alpha} = 0.78$ $\hat{\beta} = 0.37$	$\ln\hat{L} = 56.7$ $\hat{\alpha} = 0.35$ $\hat{\beta} = 0.20$
	viability sel. only	$\ln\hat{L} = 46.2$ $\hat{v}_m = 0.97$ $\hat{v}_f = 0.65$	$\ln\hat{L} = 68.7$ $\hat{v}_m = 0.30$ $\hat{v}_f = 0.60$	$\ln\hat{L} = 45.9$ $\hat{v}_m = 0.87$ $\hat{v}_f = 0.39$
	fecundity sel. only	$\ln\hat{L} = 58.0$ $\hat{\omega}_m = 0.24$ $\hat{\omega}_f = 0.58$ $\hat{\omega}_b = 0.42$	$\ln\hat{L} = 42.7$ $\hat{\omega}_m = 0.96$ $\hat{\omega}_f = 0.02$ $\hat{\omega}_b = 0.35$	$\ln\hat{L} = 61.6$ $\hat{\omega}_m = 0.40$ $\hat{\omega}_f = 0.79$ $\hat{\omega}_b = 0.20$

Table 1: Disentangling sexual, viability, and fecundity selection. The columns specify the three simulation models from Fig. 3. The rows specify three different inference models, assuming either a model of sexual selection only (free parameters α, β), viability selection only (free parameters v_m, v_f), or fecundity selection only (free parameters $\omega_m, \omega_f, \omega_b$). The cells show the resulting parameter MLEs and maximum log-likelihood values when the particular inference model is applied to the particular simulation model, averaged over 1000 simulation runs in each case.

frequency changes are the largest, they also extend throughout the run, as can be seen in the different frequency ratios of y males and yy females. These results suggest that, at least in principle, different fitness components could be disentangled by our approach.

Figure 3B demonstrates that even in a model with substantial drift ($N_e = 200$), our ML approach can still infer the individual model parameters from a single 10-generation experiment if selection is sufficiently strong. The average MLEs for the selection parameters converge closely to the true values, with standard deviations among individual simulation runs typically on the order of $10 - 15\%$ of the parameter values. However, the N_e estimates are consistently higher then the simulated values. This is likely due to the fact that in any single experiment, drift alone should already result in an upward or downward shift in yellow allele frequency, which our method would likely infer as "additional" selection, thereby underestimating the true amount of drift. Consider, for example, a completely neutral scenario. In this case, our method would still likely infer some positive or negative selection in any given run, leading to systematically lower inferred levels of drift, even though the direction of selection would be random and average out over many runs.

So far, we have only tested our inference method under the assumption that we know which type of selection has acted. In this case, we validated that the method can reliably infer the respective selection parameters, but can it also distinguish the different types of selection from each other? To test this, we applied our method to each of the three selection scenarios, using either the correct inference model or models in which a different type of selection was assumed. This can be easily implemented in our framework by fixing specific selection parameters. For example, to devise a sexual-selection-only inference model, the parameters α and β would be allowed to vary, while all other selection parameters would be set to one. The resulting parameter MLEs and maximum log-likelihoods values for each simulation model / inference model combination are shown in Table 1. As expected, the highest log-likelihoods were always achieved when the simulation model matched the inference model, and the decreases in likelihood when the inference model was misspecified were significant in all cases ($p < 0.05$) according to likelihood ratio tests. This indicates that at least for situations when model parameters have distinctly different values, we should have sufficient power to determine the type of selection operating on a particular system.

Parameter estimation in cage experiments and model comparison

We applied our ML inference method to the combined data from our five cage populations by adding the individual log-likelihoods of each of the 14 generational transitions observed in our experiments. Table 2 shows the resulting parameter MLEs and maximum log-likelihood values for several different inference models.

We first tested our full inference model with all three types of selection possible (eight free parameters total, including N_e). The resulting parameter MLEs for this model suggest that all three types of selection were indeed acting in the cages, as indicated by the fact that at least one parameter for each type of selection was inferred to be different from one. However, we suspected that there could still be substantial overfitting

model description	$\hat{N}_e$	$\hat{v}_m$	$\hat{v}_f$	$\hat{\omega}_m$	$\hat{\omega}_f$	$\hat{\omega}_b$	$\hat{\alpha}$	$\hat{\beta}$	$\ln \hat{L}$	p	AICc
full model	167	0.84	0.76	0.98	0.93	1.00	0.07	1.00	69.1	8	-93
no sexual sel.	104	0.99	1.00	0.11	0.39	0.29	1*	1*	59.1	6	-94
no viability sel.	156	1*	1*	0.82	0.87	0.73	0.09	1.00	67.7	6	-111
no fecundity sel.	166	0.83	0.78	1*	1*	1*	0.06	1.00	69.1	5	-121
sexual sel. only	134	1*	1*	1*	1*	1*	0.03	0.71	64.5	3	-121
viability sel. only	16	0.83	0.62	1*	1*	1*	1*	1*	16.0	3	-24
fecundity sel. only	104	1*	1*	0.11	0.39	0.29	1*	1*	59.1	4	-106
simplistic model	31	1*	1*	0.50	$= \omega_m$	$= \omega_m^2$	1*	1*	34.6	2	-64
minimal model 1	165	0.81	$= v_m$	1*	1*	1*	0.06	1*	68.9	3	-129
minimal model 2	156	1*	1*	0.86	$= \omega_m$	$= \omega_m^2$	0.09	1*	67.7	3	-127

Table 2: Parameter MLEs and model comparison. Each row shows the parameter MLEs and associated maximum log-likelihood value when estimated from the combined data across all cages (1-5). An entry of 1* indicates that the parameter was fixed at a value of one in the particular inference model. "=" indicates that the value of a parameter is fixed in relation to another. p denotes the number of free parameters of the model, and AICc specifies its corrected Akaike information criterion value, AICc $= 2p - 2 \ln \hat{L} + (2p^2 + 2p)/(n - p - 1)$, where $n = 14$ is the sample size. Smaller AICc values indicate better support for the model.

in this model, given the rather large number of parameters for the amount of data available. Furthermore, various parameters could be intertwined with each other and difficult to disentangle in practice. For example, it should generally be possible to approximate many scenarios of preferential mate choice to some extent by appropriately tuning a model with only fecundity selection, especially since the latter has more free parameters in our framework.

To assess which fitness components are indeed most essential to incorporate for accurately describing the genotype dynamics observed in our cages, we tested several inference models, systematically reducing model complexity. The quality of these models was then compared by calculating the corrected Akaike information criterion (AICc) values,[1,22] which provide an estimator for the goodness-of-fit of a given model while also penalizing an increase in the number of model parameters. While the correction term is strictly accurate only for linear models, it serves as an adequate approximation for more complex scenarios in most cases.[22]

We first evaluated inference models in which one type of selection was completely eliminated (by setting all parameters describing that component to a value of one). In this case, the largest reduction in maximum log-likelihood was observed when sexual selection was excluded. Inference models without viability or fecundity selection also had decreased maximum log-likelihoods compared to the full model, but reductions were less dramatic. All of these models yielded better AICc values than the full model because of their reduced number of parameters.

We next tested models with only one type of selection acting. Consistent with the above result, we found that a model with only sexual selection achieved a higher maximum log-likelihood compared to models with only viability or only fecundity selection, and the AICc value of this model was again better than the full model. The model with only viability selection was the least-supported model we tested overall.

A simplistic model with only a single fecundity parameter $\omega_m = \omega_f$, assumed to be equal in males and females and multiplicative ($\omega_b = \omega_m \omega_f$), also produced a very poor fit to the observed data, emphasizing the need for more complex models for understanding the fitness costs of the yellow phenotype.

Finally, we sought to identify a minimal model of low complexity, based on a limited number of biological assumptions, which can still describe the observed dynamics reasonably well. Motivated by the observation that the β parameter was often estimated to be one, we assumed for this minimal model that sexual selection acts only by lowering the mating success of yellow phenotype males when wild-type females choose their mate. In addition, we assumed that yellow phenotype individuals have either reduced viability (minimal model 1), or reduced fecundity (minimal model 2), with equal costs between the sexes and multiplicative costs in the case of fecundity. Both of these minimal models have only three free parameters (including N_e), yet yielded better AICc values than all other models. In fact, our minimal model 1 yielded the best AICc value of all models tested, suggesting that the data can be explained well with this simple model in which wild-type females show severe mating-bias against males of yellow phenotype ($\hat{\alpha} = 0.06$) and both sexes with the yellow

phenotype experience a modest reduction in viability ($\hat{v}_m = \hat{v}_f = 0.81$). Note, however, that the inclusion of preferential mate choice is crucial in this model, given that the model with only viability selection produced the worst fit among all models tested.
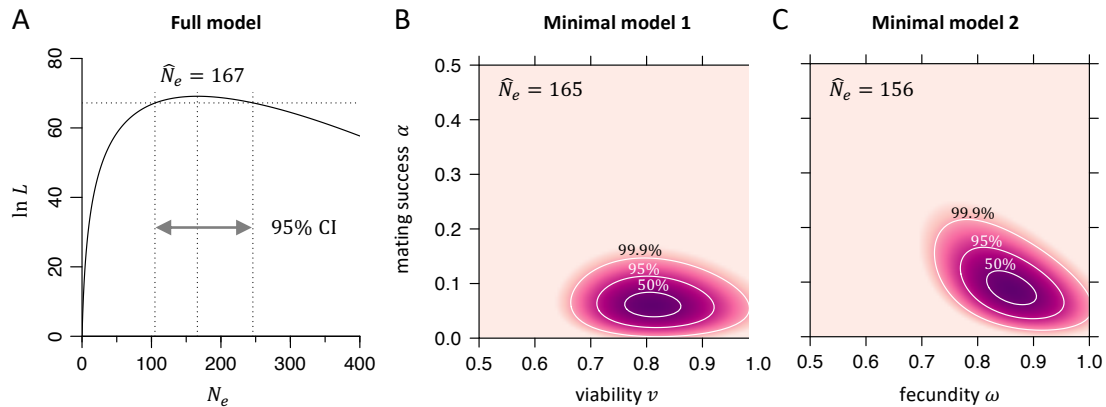

Figure 4: Log-likelihood surfaces and confidence intervals. (A) The log-likelihood curve of our full model varying N_e, while keeping other parameters fixed at at their maximum likelihood estimate. The vertical dashed lines indicate the maximum likelihood estimate $\hat{N}_e = 167$ and the corresponding 95% confidence interval (CI), estimated from a likelihood ratio test with one degree of freedom. (B) The log-likelihood surface for minimal model 1 with the mating success parameter α, viability parameter $v_m = v_f$, and fixed $\hat{N}_e$. (C) The log-likelihood surface for minimal model 2 with the mating success parameter α, fecundity parameter $\omega_m = \omega_f$, $\omega_b = \omega_m\omega_f$, and fixed N_e. Darker colors correspond to higher likelihoods. The three contour lines specify confidence intervals according to a likelihood ratio test with two degrees of freedom.

One key advantage of a ML-based approaches is that confidence intervals of parameter estimates can be easily calculated by likelihood ratio tests. Figure 4A demonstrates this for the effective population size parameter N_e in our full model, which we inferred to be $N_e = 167$ with a 95% confidence interval of approximately $100 - 250$, or about 5-12% of the census population size. This estimate is consistent with previous cage evolution experiments in *D. melanogaster*, where N_e values were observed ranging from 4%-25% of population census sizes.[28,31] Importantly, in our framework, the inference parameter N_e will also depend on the overall goodness-of-fit of the inference model. A misspecified model would be expected to result in lower N_e estimates, because the ML approach would have to assume higher amounts of drift to explain the larger deviations between the predicted and observed frequency trajectories.

Figures 4B,C show the likelihood surface plots and confidence intervals for the selection parameters in each of our two minimal models, demonstrating that our ML approach has good power in inferring these parameters. The diagonal elongation of the contour areas in minimal model 2 suggest that the two parameters are in fact partially dependent on one another in this model.

Discussion

In this study, we developed a ML approach for estimating selection parameters from time series data of genotype frequencies. The key advancement of our approach over existing methods is the ability to explicitly distinguish among the different components that constitute fitness, including mating success, fecundity, and viability. While our framework can be applied to a wide range of scenarios, we demonstrated it in this study for the specific example of a recessive X-linked allele potentially undergoing sex and phenotype-based sexual, fecundity, and viability selection. Our analysis indicates that these forms of selection indeed can result in distinct genotype frequency trajectories, which can be reliably distinguished from each other using our inference method.

As a proof-of-principle, we applied our approach to study the fitness costs associated with a disrupted *yellow* allele in *D. melanogaster*, using data from cage evolution experiments. Consistent with previous

studies,[13,21] we found that yellow phenotype males indeed experience a large fitness cost compared to their wild-type counterparts. However, our ML approach allowed us to further show that this primarily results from wild-type females strongly preferring to mate with wild-type males over yellow males. Additionally, yellow phenotype males and females may both experience a moderate reduction in viability according to our minimal model 1 that yielded the highest statistical support (Table 2).

Figure 5 shows that dissecting individual fitness components is crucial for an accurate description of the genotype frequency dynamics at the *yellow* locus. A simplistic model that seeks to approximate this with a single selection parameter would predict significantly different dynamics from those observed in the cages, whereas our minimal model captures the true dynamics with substantially higher accuracy.
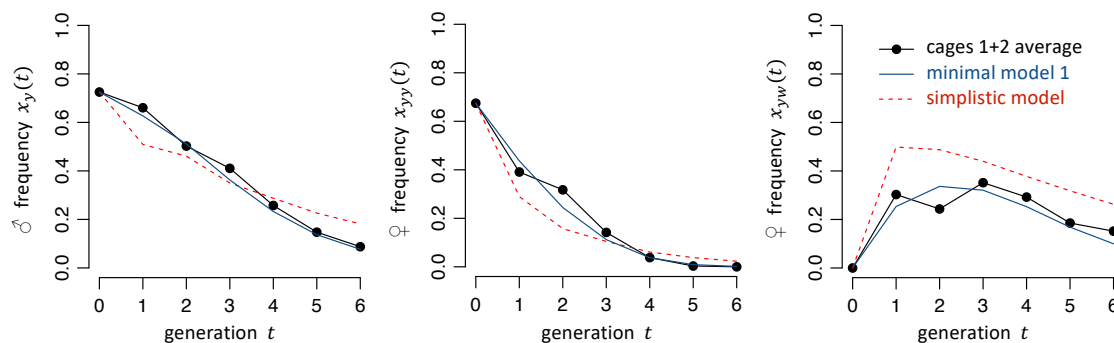

Figure 5: Comparison of the observed genotype trajectories in our population cages (averaged over cages 1 and 2) with the predicted trajectories of the simplistic one-parameter fecundity-selection-only model ($\omega_m = \omega_f = 0.50$, $\omega_b = \omega_m\omega_f$) and our minimal model 1 ($v_m = v_f = 0.81$, $\alpha = 0.06$), using the parameter MLEs inferred in Table 2. Simulations were started from the same initial genotype frequencies as the cages and were run without drift.

Despite its generality, there are a number of important limitations to our method. In contrast to many existing methods, which are able to analyze sparse allele frequency data obtained from distant time-points, our method requires frequency estimates over discrete, consecutive generations. This requirement allowed us to establish a fully analytical model based on the Wright-Fisher process, which does not rely on simulations or numerical approximations of the transition probabilities across multiple generations, as are typically invoked by other methods. However, it does likely limit the applicability of our approach to experimental studies where discrete generations can be enforced and data can be obtained for each generation.

There are also potential issues regarding overfitting in our model. For example, in our implementation of a mate choice model, we restricted mate choice selection to males, while females could also potentially undergo such selection. Additionally, with the large number of possible parameters, certain combinations of values can create very similar outcomes. It is likely that such parameters can only be distinguished in complex scenarios involving major sex-based differences between these parameters. This is apparent, for instance, when examining sexual vs. fecundity parameters in Table 2. Similarly, even though our minimal models were alike in that both had mate choice selection against yellow males, the first model had a viability cost for yellow phenotype individuals while the second had a fecundity cost. These results may also be influenced by small sample sizes, so it remains unclear whether or how the yellow phenotype induces a modest reduction in viability, fecundity, or perhaps both.

For calculating the probabilities of specific allele frequency changes between consecutive generations in our model, it was necessary to incorporate an effective population size. We defined this parameter to be equal for males and females and constant across generations, yet other valid possibilities exist. For example, the effective population size could be defined as a fixed percentage of the total population for each generational transition, using either the current generation size, the previous generation size, or some combination of both. Furthermore, male and female likelihoods could be combined, whereas we chose to factorize them in our experimental situation because of a skewed sex ratio likely due to experimental factors. Another possibility would be to treat male and female effective population sizes separately, which would also allow for a lower effective proportion of males compared to females. In some scenarios, these considerations may

yield a model that is more representative of the level of stochasticity in each generational transition, though we found that most methods yielded similar results in our limited data set.

Disrupted *yellow* alleles are classic genetic markers in *Drosophila* genetics and have long been known to affect the mating success of male flies due to less effective courtship displays. Specifically, wing extension display is known to be defective in yellow males, resulting in their securing only approximately 10% of matings with a wild-type female when in 1:1 competitions with wild-type males,[13] which is consistent with our results. Interestingly, in long-term yellow phenotype stocks, yellow females were thought to have evolved to be more receptive to their male counterparts,[3] but it was later determined that females with the yellow phenotype were themselves more receptive to yellow males, irrespective of their genetic background.[12] Our study corroborates this result. While it is possible that yellow females do show a modest preference for wild-type males over yellow males, our study lacked the statistical power to determine this.

Many evolutionary scenarios will still be well served by classical models invoking only a single, fixed selection coefficient. However, certain scenarios will likely require more detailed fitness models such as the one we presented here. Perhaps the primary application for such models involve the spread of selfish genetic elements. The potential use of engineered gene drives,[6,43] for example, demands accurate and detailed models for assessing their performance in natural populations, which will likely have to be based on cage trials. For instance, while an existing *D. melanogaster* gene drive[8] that targeted the *yellow* gene would not be able to spread, a multiplexed-gRNA drive[7] at this locus could be expected to spread in a frequency-dependent fashion in either of our minimal models involving sexual selection, a feature not commonly associated with such homing-type gene drives. Thus, determination of individual fitness components is not only essential for understanding the spread of gene drives, but could potentially inform their design as well.

Acknowledgements

References

[1] AKAIKE, H. Information Theory and an Extension of the Maximum Likelihood Principle. In *2nd Int. Symp. Inf. Theory*, B. Petrov and F. Csaki, Eds. Akadémiai Kiadó, Budapest, Hungary, 1998, pp. 199–213.

[2] BARTON, N. H., AND SERVEDIO, M. R. The interpretation of selection coefficients. *Evolution 69*, 5 (2015), 1101–1112.

[3] BASTOCK, M. A Gene Mutation Which Changes a Behavior Pattern. *Evolution 10* (1956), 421–439.

[4] BOLLBACK, J. P., YORK, T. L., AND NIELSEN, R. Estimation of 2Nes from temporal allele frequency data. *Genetics 179*, 1 (2008), 497–502.

[5] BUNDGAARD, J., AND CHRISTIANSEN, F. B. Dynamics of polymorphisms. I. Selection components in an experimental population of Drosophila melanogaster. *Genetics 71*, 3 (1972), 439–460.

[6] CHAMPER, J., BUCHMAN, A., AND AKBARI, O. S. Cheating evolution: Engineering gene drives to manipulate the fate of wild populations. *Nat. Rev. Genet. 17*, 3 (2016), 146–159.

[7] CHAMPER, J., LIU, J., OH, S. Y., REEVES, R., LUTHRA, A., OAKES, N., CLARK, A. G., AND MESSER, P. W. Reducing resistance allele formation in CRISPR gene drive. *Proc. Natl. Acad. Sci. 115*, 21 (2018), 5522–5527.

[8] CHAMPER, J., REEVES, R., OH, S. Y., LIU, C., LIU, J., CLARK, A. G., AND MESSER, P. W. Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. *PLoS Genet. 13*, 7 (2017), e1006796.

[9] CHRISTIANSEN, F. B., AND FRYDENBERG, O. Selection component analysis of natural populations using population samples including mother-offspring combinations. *Theor. Pop. Biol. 4* (1973), 425–445.

[10] CLARK, A. G., AND FELDMAN, M. W. The estimation of epistasis in components of fitness in experimental populations of drosophila melanogaster II. Assessment of meiotic drive, viability, fecundity and sexual selection. *Heredity (Edinb). 46*, 3 (1981), 347–377.

[11] CLARK, A. G., FELDMAN, M. W., AND CHRISTIANSEN, F. B. The estimation of epistasis in components of fitness in experimental populations of *Drosophila melanogaster* I. A two-stage maximum likelihood model. *Hered. 46*, 3 (1981), 321–346.

[12] DOW, M. A. The genetic basis of receptivity of yellow mutant *Drosophila melanogaster* females. *Behav. Genet. 6*, 2 (1976), 141–143.

[13] DRAPEAU, M. D., CYRAN, S. A., VIERING, M. M., GEYER, P. K., AND LONG, A. D. A cis-regulatory sequence within the yellow locus of *Drosophila melanogaster* required for normal male mating success. *Genetics 172*, 2 (2006), 1009–1030.

[14] DRAPEAU, M. D., RADOVIC, A., WITTKOPP, P. J., AND LONG, A. D. A gene necessary for normal male courtship, yellow, acts downstream of fruitless in the *Drosophila melanogaster* larval brain. *J. Neurobiol. 55*, 1 (2003), 53–72.

[15] FEDER, A. F., KRYAZHIMSKIY, S., AND PLOTKIN, J. B. Identifying signatures of selection in genetic time series. *Genetics 196*, 2 (2014), 509–522.

[16] FERRER-ADMETLLA, A., LEUENBERGER, C., JENSEN, J. D., AND WEGMANN, D. An Approximate Markov Model for the Wright-Fisher Diffusion and Its Application to Time Series Data. *Genetics 203*, 2 (2016), 831–846.

[17] FOLL, M., SHIM, H., AND JENSEN, J. D. WFABC: A Wright-Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. *Mol. Ecol. Resour. 15*, 1 (2015), 87–98.

[18] GOMPERT, Z. Bayesian inference of selection in a heterogeneous environment from genetic time-series data. *Mol. Ecol. 25*, 1 (2016), 121–134.

[19] GRATZ, S. J., CUMMINGS, A. M., NGUYEN, J. N., HAMM, D. C., DONOHUE, L. K., HARRISON, M. M., WILDONGER, J., AND O'CONNOR-GILES, K. M. Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease. *Genetics 194*, 4 (2013), 1029–1035.

[20] GRATZ, S. J., UKKEN, F. P., RUBINSTEIN, C. D., THIEDE, G., DONOHUE, L. K., CUMMINGS, A. M., AND OCONNOR-GILES, K. M. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in *Drosophila*. *Genetics 196*, 4 (2014), 961–971.

[21] HEISLER, I. L. Inheritance of female mating propensities for yellow locus genotypes in *Drosophila melanogaster*. *Genet. Res. 44*, 2 (1984), 133–149.

[22] HURVICH, C. M., TSAI, C.-L., AND JUN, N. Regression and Time Series Model Selection in Small Samples. *Biometrika 76*, 2 (1989), 297–307.

[23] ILLINGWORTH, C. J., PARTS, L., SCHIFFELS, S., LITI, G., AND MUSTONEN, V. Quantifying selection acting on a complex trait using allele frequency time series data. *Mol. Biol. Evol. 29*, 4 (2012), 1187–1197.

[24] ILLINGWORTH, C. J. R., AND MUSTONEN, V. Quantifying selection in evolving populations using time-resolved genetic data. *J. Stat. Mech. Theory Exp.*, 1 (2013), P01004.

[25] LACERDA, M., AND SEOIGHE, C. Population genetics inference for longitudinally-sampled mutants under strong selection. *Genetics 198*, 3 (2014), 1237–1250.

[26] LINNEN, C. R., AND HOEKSTRA, H. E. Measuring natural selection on genotypes and phenotypes in the wild. *Cold Spring Harb. Symp. Quant. Biol. 74* (2009), 155–168.

[27] MALASPINAS, A. S., MALASPINAS, O., EVANS, S. N., AND SLATKIN, M. Estimating allele age and selection coefficient from time-serial data. *Genetics 192*, 2 (2012), 599–607.

[28] MALPICA, J. M., AND BRISCOE, D. A. Effective population number estimates of laboratory populations of *Drosophila melanogaster*. *Experientia 37*, 9 (1981), 947–948.

[29] MATHIESON, I., AND MCVEAN, G. Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. *Genetics 193*, 3 (2013), 973–984.

[30] MESSER, P. W., ELLNER, S. P., AND HAIRSTON, N. G. Can Population Genetics Adapt to Rapid Evolution? *Trends Genet. 32*, 7 (2016), 408–418.

[31] MUELLER, L. D., JOSHI, A., SANTOS, M., AND ROSE, M. R. Effective population size and evolutionary dynamics in outbred laboratory populations of *Drosophila*. *J Genet 92*, 3 (2013), 349–361.

[32] MUKAI CARDELLINO, R. A., WATANABE, T. K., & CROW, J. F., T., AND MUKAI, T. The genetic variance for viability and its components in a local population of Drosophila melanogaster. *Genetics 78* (1974), 1195–1208.

[33] NADEAU, H. J., AND BACCUS, R. Selection components of four allozymes in natural populations of Peromyscus maniculatus. *Evolution (N. Y). 35*, 1 (1981), 11–20.

[34] NADEAU, J. H., DIETZ, K., AND TAMARIN, R. H. Gametic selection and the selection component analysis. *Genet. Res. 37*, 3 (1981), 275–284.

[35] NISHINO, J. Detecting Selection Using Time-Series Data of Allele Frequencies with Multiple Independent Reference Loci. *G3 (Bethesda) 3*, 12 (2013), 2151–2161.

[36] SCHRAIBER, J. G., EVANS, S. N., AND SLATKIN, M. Bayesian inference of natural selection from allele frequency time series. *Genetics 203*, 1 (2016), 493–511.

[37] SHIM, H., LAURENT, S., MATUSZEWSKI, S., FOLL, M., AND JENSEN, J. D. Detecting and Quantifying Changing Selection Intensities from Time-Sampled Polymorphism Data. *G3 (Bethesda) 6*, 4 (2016), 893–904.

[38] SIEGISMUND, H., AND FRYDENBERG, O. Selection component analysis of natural populations using population samples including mother-offspring combinations III. *Theor. Pop. Biol. 27* (1985), 268–297.

[39] SOBER, E., AND LEWONTIN, R. C. Artifact, Cause and Genic Selection. *Philos. Sci. 49*, 2 (1982), 157–180.

[40] STEINRÜCKEN, M., BHASKAR, A., AND SONG, Y. S. A novel spectral method for inferring general diploid selection from time series genetic data. *Ann. Appl. Stat. 8*, 4 (2014), 2203–2222.

[41] TERHORST, J., SCHLÖTTERER, C., AND SONG, Y. S. Multi-locus Analysis of Genomic Time Series Data from Experimental Evolution. *PLoS Genet. 11*, 4 (2015), e1005069.

[42] TOPA, H., JÓNÁS, Á., KOFLER, R., KOSIOL, C., AND HONKELA, A. Gaussian process test for high-throughput sequencing time series: Application to experimental evolution. *Bioinformatics 31*, 11 (2015), 1762–1770.

[43] UNCKLESS, R. L., MESSER, P. W., CONNALLON, T., AND CLARK, A. G. Modeling the manipulation of natural populations by the mutagenic Chain reaction. *Genetics 201*, 2 (2015), 425–431.

[44] WILSON, R., BURNET, B., EASTWOOD, L., AND CONNOLLY, K. Behavioral Pleiotropy of Yellow Gene in *Drosophila melanogaster*. *Genet. Res. 28*, 1 (1976), 75–88.