Evolutionary genetics and dynamics of transitions in sex determination

Schenkel, Martijn

DOI:
10.33612/diss.166344703

CHAPTER VI:

SEX-SPECIFIC FITNESS ESTIMATION IN THE HOUSEFLY *MUSCA DOMESTICA*

Martijn A. Schenkel
Julia I. Camacho
Tomás Cunha
Jean-Christophe Billeter
Ido Pen
Leo W. Beukeboom

Abstract

The housefly *Musca domestica* is a commonly used model system to test theories on sex determination and sex chromosome evolution. These studies are however impeded by our limited understanding of how sex-specific traits affect fitness in this species. Here, we identify candidate fitness proxies for male and female fitness in *M. domestica*. For females, we identified correlations between lifetime reproductive success and 13 fitness components or combinations of components. We found that a combination of early fecundity and lifespan has a slightly better predictive value as lifetime cumulative fecundity alone (though still highly similar; $\Delta LOO = 6.5 \pm 6.8$), which was previously found to be a strong predictor for female fitness in *M. domestica*. We discuss these findings in the context of previous work on female fitness in this species and its relatives. For males, in absence of novel data, we review past work to identify which processes are most likely to result in fitness variation. Mating success is a strong candidate predictor for male fitness, given that female remating is low or absent under laboratory conditions. However, female remating may occur frequently under natural conditions. It is necessary to assess how often female remating occurs in houseflies and its impact on male and female fitness to develop a definitive fitness estimation methodology. Additionally, to experimentally verify the adaptive value of early female fecundity, we performed artificial selection on early female fecundity (cumulative egg production during first six days) and found that selection led to only a slight increase in early female fecundity in a selected strain. We instead found that mothers tended to produce more daughters who successfully lay eggs, so that the proportion of non-reproductive females decreased. This suggests that the response to selection occurred not in the selected trait but rather in a correlated trait. We conclude that sex-specific fitness may be estimated by (early) fecundity in females and mating success in males, but additional research on the effect of female remating on female and male fitness is required.

Introduction

Theoretical research has been and continues to be fundamental to the field of evolutionary biology. Empirical confirmation, however, is generally infeasible given the scale (both in terms of time and populations) and complexity of many evolutionary processes. These issues have inspired the development of 'model systems', i.e., investing in developing tools and knowledge of a handful of specific organisms to derive generalizable findings. The general premise here is that a mechanistic (both ultimate and proximate) understanding of a phenomenon in one species not only provides a better insight into this phenomenon than piecemeal evidence from many different systems, but also that it may provide a solid foundation to study this phenomenon in other species afterwards.

The evolution of sex determination (SD) mechanisms and sex chromosomes are both phenomena that have been investigated extensively using theoretical approaches (van Doorn, 2014). Theory predicts that evolutionary transitions in SD mechanisms, which occur by the spread of a novel SD gene in a population, are driven by, amongst others, intralocus sexual conflict (van Doorn & Kirkpatrick, 2007, 2010; Muralidhar & Veller, 2018), parent-offspring conflict (Werren et al., 2002; Kuijper & Pen, 2014), and sex ratio selection (Kozielska et al., 2006; Uller et al., 2007). Modelling the evolution of sex chromosomes suggests that these chromosomes can arise from the appearance of an SD gene followed by a period of sex-specific adaptive evolution of chromosomal regions linked to the SD gene (Rice, 1987a; Charlesworth et al., 2005), followed by degeneration of these regions due to the absence of recombination (Bachtrog, 2008, 2013). Whether these predictions are accurate is unknown, partly due to the absence of suitable model systems to test them.

The housefly *Musca domestica* is a model system in which the evolution of SD mechanisms and sex chromosomes has been studied owing to its multifactorial SD system (Dübendorfer et al., 2002). Within this species, various SD mechanisms occur within and between populations (Hamm et al., 2015). This variation can be exploited experimentally to generate sex chromosomes *de novo*, which allows for real-time studies on early sex chromosome evolution (see Chapter 5). Alternatively, laboratory populations harbouring a combination of different SD genes may be set up and maintained to determine which SD genes can spread throughout the

population and which ones are purged by selection (Hamm & Scott, 2008; Kozielska, 2008; Hamm et al., 2009). Such approaches are however still rudimentary, in that they only assess the allele frequencies of different SD genes upon initiation and after several generations to assess the capacity for SD genes to spread without any attempt to determine which factors affect this process. An alternative approach is to experimentally induce sex ratio biases in this polymorphic system through e.g. meiotic drive (cf. Lyttle, 1981) and determining which and how variants respond. Such experiments on SD evolution require accounting for the role of fitness variation associated with different sex chromosomes or SD genes, and by extension how they are subject to adaptive evolution. Additionally, given the sex-specific nature of sex chromosomes and SD genes, their effect on individual fitness must obviously be assessed in a sex-specific manner. A crucial step in the development of *M. domestica* as a model system for the evolution of SD and sex chromosomes is therefore to develop a sex-specific methodology for fitness assessment.

Fitness assessment is a complex procedure, which faces both conceptual (see Box 1) as well as practical issues. Long-term databases on naturally-occurring, pedigreed populations have been used to determine individual fitness based on reproductive success. Such studies have been able to quantify fitness variation under natural conditions in species such as red deer (Foerster et al., 2007), lizards (Calsbeek et al., 2015) and Soay sheep (Hunter et al., 2019). The required databases may be generated for laboratory populations of houseflies too, and would allow developing fitness proxies by finding correlations between variation in fitness and variation in other traits (e.g., body size). This approach would largely circumvent the conceptual issues outlined in Box 1, but faces substantial practical issues in houseflies. Most importantly, genotyping houseflies is not (yet) feasible without sacrificing or substantially harming individuals, making it impractical to link variation in specific individual traits to their long-term reproductive success and therefore unsuitable to detect fitness proxies. With proper experimental design, the impact of these conceptual issues may be reduced so that small-scale fitness experiments may be suitable to identify components of fitness variation and correlations between fitness and other phenotypes that may then serve as fitness proxies.

In this chapter, we investigate the use of different fitness components as estimates for fitness in *M. domestica* males and females. We first assess how female lifetime reproductive success (LRS) is accrued over her lifetime, and how LRS

correlates with other traits to identify potential proxies; we consider here the original set of fitness components considered by Reed & Bryant (2004), who used age at first reproduction, number of clutches, total fecundity, lifespan, size of first successful clutch, hatch success of first successful clutch, and overall hatch success as fitness components. Additionally, we consider fecundity on the first day after onset of the experiment, total fecundity on the first six days of experiment, and interactions between several fitness components. Next, we carried out an artificial selection experiment in which we selected for early female fecundity as a proxy for female fitness to determine if such approaches can be used to induce sex-specific adaptation. We discuss our results in the context of past work on female fitness in *M. domestica*. Additionally, we review the reproductive biology of males in this species to identify candidate proxies for male fitness.

Box 1: Fitness definitions and conceptual issues

Although fitness is a keystone concept in evolutionary biology, different schools of evolutionary biology have adopted different definitions. A common thread in such definitions is that fitness entails a measure of the representation of genetic material in the future generations (i.e., reproductive success) or similarly the rate at which genetic variants will change in frequency over time. However, the level at which selection takes place varies. Fitness was initially defined at the level of individuals (Darwin, 1859), but alternative definitions have been proposed which consider fitness to be determined at the gene level (Dawkins, 1976). This view has subsequently been extended so that selection at the level of genes results in fitness maximization at the level of individuals via inclusive fitness effects (Hamilton, 1964). Here, traits that maximize a gene's transmission to the next generation may be favourable despite costs to its bearer due to increased transmission of gene copies borne by the bearer's relatives, who experience a fitness benefit from the trait expressed by the initial bearer. Yet another viewpoint considers fitness as the result of the interaction between individuals and their associated microbiome, resulting in the holobiont concept (Bordenstein & Theis, 2015).

Conceptual issues in measuring fitness

Quantifying fitness is essential to empirical research but has often proven difficult due to both practical issues, such as the inability to assign parentage in wild populations, and conceptual issues pertaining to the manner in which fitness is realized. Here, we discuss some of these conceptual issues, focussing on (1) the role of sex-specific function and the associated constraints on fitness maximization, (2) the role of life history traits and how different life history strategies may yield identical fitness, (3) the social nature of fitness under sexual reproduction and the potential role of interlocus sexual conflict, and (4) the composite nature of fitness.

In anisogamous species, fitness is a strongly sex-specific trait. Because males produce smaller gametes, they will be able to produce more gametes than females (Parker et al., 1972). This enables a single male to have a higher potential reproductive success than a single female (Bateman (1948), but see Gowaty et al. (2012)), although investment into reproductive success extends beyond gametes in males and females, e.g. in courtship and parental care (Trivers, 1972). The difference between gamete sizes leads to a sex difference in reproductive potential and concomitantly potential fitness, and therefore plays an important role in how males and females may seek to maximize their fitness in different ways. Males can be expected to reduce their investment in offspring development and care, as this would prevent them from making full use of their superior ability to generate gametes (Trivers, 1972); females instead are limited by their ability to generate costly gametes and may seek to increase investment by males into shared offspring. Fitness is however not solely maximized in males by minimizing investment, nor in females by maximizing the investment made by male partners. Males may stand to benefit by increasing female investment in parental care, an example of exploitation in the context of interlocus sexual conflict (as discussed in Chapter 2). Similarly, females may reduce investment into single eggs so as to be able to produce more eggs, ideally resulting in more surviving offspring. Both investments in parental care and egg production benefit offspring fitness and are examples of inclusive fitness effects, as the cost of

either investment to the male and/or female is offset by increased fitness in their offspring (Parker et al., 1972).

A second issue pertains to the relationship between life histories, individual fitness, and population dynamics. In age-structured populations (e.g. where generations overlap and parents may survive and coexist with their offspring), reproducing at different timepoints in an individual's life can affect fitness differently depending on population dynamics (Charlesworth, 1980; Brommer, 2000). Under these conditions, reproductive value rather than LRS may be a better indicator of individual fitness (Fisher, 1930). Whereas LRS is simply the cumulative number of offspring generated by an individual, its reproductive value takes into account how offspring generated at different times contribute differently to the overall rate at which an individual's genes can spread throughout the population (McGraw & Caswell, 1996; Engen et al., 2009). In growing populations, early reproduction is more advantageous as the value of each offspring will be larger since they represent a larger fraction of the overall gene pool. The inverse holds for shrinking populations: reproduction at later timepoints is considered advantageous as the overall population will have shrunk, and hence an offspring produced at this time represents a large fraction of the overall gene pool. In stable populations or those without age structure, timing of offspring production is considered to be neutral. Note that although we focus here on reproductive value in the context of timing, other factors may play into its calculation (Grafen, 2020). When offspring can vary for a certain trait which can affect fitness such as sex, then the production of sons and daughters must be weighed accordingly. To illustrate, consider a population with a male-biased sex ratio, where individuals that produce a surplus of daughters have higher reproductive value than those who do not because female offspring will themselves experience higher fitness.

Third, fitness under sexual reproduction is achieved via interaction with other parties that may have different interests in reproduction so that fitness is not an intrinsic individual quality but rather a socially-influenced trait (see Chapter 2). An individual's fitness can then be affected by traits exhibited by its mates (Moore & Pizzari, 2005). Such effects are generally termed indirect genetic effects (IGEs), and are used to indicate that an effect on an individual's trait

(fitness) is caused not by one's own genotype but rather by that of another individual via a phenotype that it encodes in this individual. In this case, an individual's mates may exhibit certain traits that can influence its fitness. Whether an individual exhibits high fitness or not then depends on how well it is adapted to the traits exhibited by its mates and vice versa (Rice, 1996b, 2000). By extension, individuals may exhibit high fitness with one type of mate but low fitness with another. To accurately assess fitness, it is necessary to (1) understand which part of fitness variation is caused by IGEs or (2) nullify the effects of IGEs on fitness variation. The first strategy requires fitness to be assessed with different types of mates (e.g. Chow *et al.*, 2010), whereas the second requires that fitness is always assessed with a standardized mating pool (i.e. an inbred/isogenic strain). However, even with a standardized mating pool, fitness remains a socially-affected trait, and specific mating pools may systematically favour or disfavour specific individuals, so that the first strategy is preferred.

Fourth and final is how to accurately empirically assess an individual's fitness. Fitness is accrued over an individual's entire (reproductive) lifespan and is the compound result of a sequence of biological processes including mate searching behaviour, mate choice, copulation, and fertilization. Experimental designs which fail to account for early and late reproduction or those that ignore variation in a specific process may not identify all fitness variation. For example, fitness measurements in forced crosses may ignore a male's performance with regard to searching for females, or a female's capacity to discern between high-quality and low-quality mates (Bluhm & Gowaty, 2004). Similarly, such approaches may fail to capture trade-offs between different components, such as when high mating success is offset by low offspring survival (Townsend, 1989). Multivariate approaches, where fitness is assessed for different components (e.g., mating success, fertilization success, offspring viability) remedy such issues and successfully identify causes of fitness variation. Fitness may also be assessed as the ability to invade into different populations (Metz et al., 1992; Mylius & Diekmann, 1995), where individuals of interest are introduced into a 'resident' population which is subsequently maintained for several generations. Fitness is then defined as the ability for focal individuals to

outcompete 'resident' individuals. By using an approach spanning multiple generations, all relevant fitness components are taken into consideration.

Methods

Fly strains & culturing procedures

Wildtype housefly stock strains were established from flies collected in the field in Spain (SPA1 to SPA5) (Figure 1); additionally two laboratory strains M^{III} and aabys were used. An overview of all stock strains is provided in Table 1. All housefly stock strains were maintained at 25°C, 14:10 LD in 3250-ml bottle cages (Semadeni, Ostermundigen, Switzerland; 24 × 13.5 × 13.5 cm; L × W × H). All cages contained two 15-ml vials of sugar water (20% wt/vol) and two vials of water, one 35-mm Petri dish with milk powder, and were shut off with tubular gauze. Sugar water and water was replaced three times per week on Mondays, Wednesdays, and Fridays. Egg-laying substrate was provided when flies in cages had become sexually mature (3-4 days after emergence, though mating may start earlier). Egg-laying substrate consisted of a mixture of wheat bran, flour, milk powder, and dry inactivated yeast (20: 3: 2.4: 1 ratios), of which approximately 200 g was mixed with a solution of Nipagin (Spruyt Hillen, IJsselstein, The Netherlands; 5 ml 10% wt/vol in 99% ethanol, mixed with demi water to a total of approximately 225 ml). Per cage, two 35-ml cups with egg-laying substrate were provided for 3-4 days, after which both cups were emptied into one 770-ml beaker. Approximately 150 g of egg-laying substrate was added to each beaker as larval feed. After this, beakers were again kept under 25°C, 14:10 LD until adults emerged between 7 to 10 days later. In addition to standard culturing procedures, the ITA-3 strain was also cultured in a large Plexiglas cage (30 × 35 × 40 cm; L × W × H) to generate offspring for use in the introgression experiment. Each cage contained 5 vials of sugar water and water each, as well as three Petri dishes with milk powder, and was set up using two beakers of newly-emerged adult flies. After sexual maturation, 6-12 egg-laying cups were provided which were further processed as per regular culturing conditions.

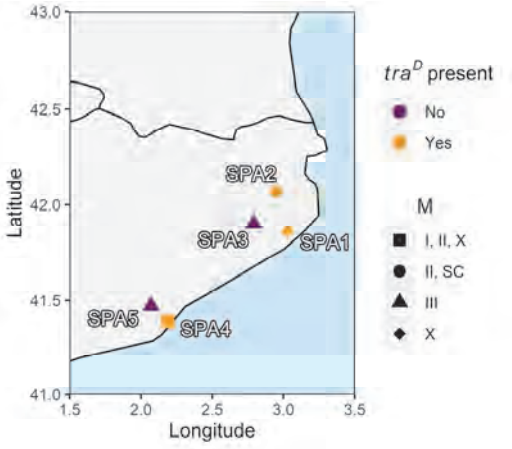

Figure 1: Geographic origin of wildtype strains SPA1 to 5 in Catalonia, Spain. M refers to the chromosome pair to which M was mapped in each strain; SC here refers to a sex-chromosomal M-factor located on either the X- and/or the Y-chromosome

Establishment of a genetically variable strain for use in fitness experiments

The strains SPA1 through SPA5 were used to generate a genetically variable strain SSM to be used in fitness experiments in December 2016. To this end, approximately 100 virgins (50 males, 50 females) were collected for each strain. Virgin flies were released in a single large plexiglass cage (30 × 35 × 40 cm; L × W × H; with a plastic mesh lid for ventilation) and allowed to interbreed. Conform standard culturing procedures, milk powder was available *ad libitum* and vials containing sugar water and water were provided and replaced three times per week. Egg-laying material was provided upon sexual maturation and processed conform standard culturing procedures to generate the F1 offspring. F1 offspring were maintained according to similar procedures as the preceding generation except that flies were released into the cage upon emergence rather than being collected as flies. The resulting SSM strain was maintained until the F6 generation before use in experiments to facilitate genetic admixture. Female fitness experiments were carried out starting in the F6 generation.

Table 1: Overview of housefly stock strains used in the experiments.

Strain	Origin	Coordinates	Collected	M location	*tra*^{D}	Phenotypic mutations[1]
SPA1	Calonge, Spain	41°86′N, 3°03′E	October 2015	X	Yes	No
SPA2	St. Jordi Desvalls, Spain	42°07′N, 2°95′E	October 2015	II, SC[2]	Yes	No
SPA3	Riudellots de la Selva, Spain	41°90′N, 2°79′E	October 2015	III	No	No
SPA4	Barcelona, Spain	41°39′N, 2°19′E	October 2015	I, II, X	Yes	No
SPA5	Sant Cugat, Spain	41°47′N, 2°07′E	October 2015	III	No	No
M^{III}	Laboratory strain	NA	NA	III	No	Females *bwb*/*bwb*, males *bwb*/+[3]
aabys	Laboratory strain	NA	NA	Y	No	*ac*; *ar*; *bwb*; *ye*; *snp*

[1] Recessive mutations are abbreviated as follows: *ac*: *alicurve* (autosome I, curly wings); *ar*: *aristapedia* (autosome II, antenna replaced by legs); *bwb*: *brown body* (autosome III, brown body coloration); *ye*: *yellow eyes* (autosome IV, yellowish eyes); *snp*: *snip wings* (autosome V; incision in wings near the tips).

[2] 'SC' refers to the sex chromosome pair, and is used to indicate that it is uncertain whether the M-factor is located on the X- or the Y-chromosome.

[3] Females have mutant brown body coloration, males have wildtype black body coloration. This is achieved by a linked M-factor to a wildtype allele at the *bwb* locus.

Female lifetime reproductive success

Female lifetime reproductive success (LRS) was defined as the total number of adult offspring produced by a single female. LRS was measured by tracking the number of eggs produced by individual females, and scoring the number of adult offspring that emerged for each batch of eggs. This procedure was carried out in the F0 and F1 generation of the artificial selection procedure (see below) to find covariates that most accurately predict LRS but are less cumbersome to assess (e.g., egg production in a given timeframe). The data reported here correspond to the results of the F1 generation of this experiment due to logistic issues encountered during the F0 iteration.

Individual SSM females were collected after emergence and stored individually for 2-3 days to allow for sexual maturation. After this, two virgin SSM

males of comparable age were added to each female (two males were added to reduce female mortality due to excessive harassment by the male as previously observed in single crosses). These triplets were then stored for 48-72 hours to allow for mating to occur. Next, a modified egg-laying substrate was provided in the form of 1/4th of a cotton pad (ca. 6.4 cm^2) soaked in milk powder solution (6.8 g milk powder per 100 ml water) on a small Petri dish (3.8 cm diameter). Egg-laying substrates were collected approximately every 18 hours, and new egg-laying substrates were provided approximately 6 hours later (i.e., 24 hours after the previous substrate was provided). This procedure was repeated daily until the female had died. Eggs from each egg-laying substrate were transferred to a Petri dish with a soft brush and counted using a stereomicroscope (ZEISS, 8× magnification). After counting, eggs were transferred to cups and standard substrate (see "Fly strains & culturing procedures") was added. Here, eggs were placed on a small layer of substrate and an additional layer of substrate was added on top of this to minimize mortality. Cups were transferred to 25°C and stored until hatching approximately 10 days later. Upon the onset of emergence, flies were counted daily until no flies had emerged for 5 subsequent days. The resulting dataset comprises the number of clutches laid by each female (N=31 excluding females that laid no eggs), and for each of these clutches (N=134) the number of eggs laid in that clutch, and how many adult offspring emerged from that clutch over time.

Artificial selection for increased female fecundity

To assess the efficacy of artificial selection in establishing enhanced female fitness, we carried out a small-scale selection experiment over 6 generations. In the F0 and F1 generations, we simultaneously carried out assays for female lifetime reproductive success (see above) which were used as the selection criterion. In subsequent generations, we used the cumulative number of eggs laid during the first six days after onset of the fecundity assay (CE). Similar to the procedure for the LRS assays, females were collected as virgins shortly after emergence and stored in individual cups (day 0). On day 2-3, when females had reached sexual maturity and were reproductively active, two virgin males were added and mating was allowed to occur for 48-72 hours. After mating, 1/4th of a cotton pad soaked in milk powder solution was provided as egg-laying substrate. The egg-laying substrate was

removed on the following day (typically 18-24h after being initially provided), the number of eggs laid were counted, and a fresh cotton pad was provided. This procedure was repeated six times in total per female.

The selection regime was then designed as follows. In the first generation (F0) the fitness of 40 females (20 in the control group and the selected group each) was assayed as described above. In the control group, we randomly picked 4 females that were allowed to reproduce, whereas in the selected group we picked those 4 females that had the highest CE to be allowed to reproduce. Note that for the control strain, we could only make use of females that managed to reproduce, and therefore sampling here is not fully random in the strict sense, but instead may be slightly biased towards fitter females. Per reproducing female, we assayed the fitness of 5 F1 daughters so that in both the control and the selected group we again assayed 20 females each (and thus a total of 40 females). This procedure was repeated until the F6 generation, in which we did assay 4×5 females per treatment but did not perform any further selection. However, due to issues with incomplete data in the F6 generation, it is omitted from the analysis here, but the results including this generation are provided in the supplementary material. In addition to CE, we also assessed female lifespan throughout the experiment, which is defined as the number of days between emerging as an adult and the day at which a female was first observed to no longer be alive. All females were checked daily to assess their survival. Altogether, we have a pedigree of females and their daughters, granddaughters, etc. along with their fitness scores (CE and lifespan) and further associated metadata.

Statistical analysis

All data analysis was carried out in R (v.4.0.2, R Development Core Team, 2020) using RStudio (v. 1.2.5033, RStudio Team, 2020). We used the "brms" package (Bürkner, 2017) to fit statistical models and analyse them with Bayesian MCMC. For these analyses, we assumed flat prior probability distributions unless specified otherwise. Data wrangling and visualisation was carried out using the "cowplot" (Wilke, 2019), "maps" (Becker et al., 2018), "mapproj" (McIlroy et al., 2020), "tidyverse" (Wickham et al., 2019) and "viridis" packages (Garnier, 2018).

We analysed female LRS by fitting different generalized linear models (GLMs) to data on offspring emergence rates and the number of offspring that emerged. Offspring emergence rates were fitted using binomial GLMs with a logit link, so that the emergence rate P_i is given by:

$$P_i = \frac{e^{\boldsymbol{X_i\beta}}}{1+e^{\boldsymbol{X_i\beta}}} \quad (1)$$

Here, $\boldsymbol{X_i}$ refers to row i of $\boldsymbol{X}$, which represents the design matrix containing the predictor variables and $\boldsymbol{\beta}$ the coefficient vector. For analyses per daily fecundity assay, we assume:

$$\boldsymbol{X_i\beta} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \quad (2)$$

where β_0 equals the intercept for the log of the odds of an offspring emerging, β_1 and β_2 are the partial regression coefficients for time (days since onset of fecundity assays) and number of eggs laid on that day, and x_{1i} and x_{2i} indicate the time and clutch size for assay i so that P_i represents the emergence rate for this assay.

For analyses based on entire lifetime data, we assume:

$$\boldsymbol{X_i\beta} = \beta_0 + \beta_1 x_{1i} \quad (3)$$

Here β_0 again equals the intercept for the log of the odds of an offspring emerging from a single egg, β_1 now represents the regression coefficient for lifetime egg production, and x_{1i} the cumulative number of eggs laid by a female i across her lifetime so that P_i represents the offspring emergence rate for this female; $P_i x_{1i}$ then represents the expected LRS for a female given a lifetime fecundity x_{1i}.

In addition to these confirmatory analyses, we explored correlates of female fitness by fitting 13 different models to the LRS scores of females (for details see Table 2); we include here all models considered by Reed & Bryant (2004) who previously studied female fitness proxies. We assumed a zero-inflated negative binomial distribution for each model. Here, zero inflation is modelled using a logit link (similar to equation 1) and assumed to be independent of the predictor variables

Table 2: Models for female LRS using various fitness proxies as predictor variables. Fecundity refers here to the egg production of a female in a given timeframe (see details below). Each model was fitted with a zero-inflated negative binomial distribution, and assumed the listed predictor variables for the negative binomial component. Zero inflation rates were assumed to be constant within each model.

Model number	**Predictor variable(s)[1]**	**Origin**
1	Age at first reproduction	(Reed & Bryant, 2004)
2	Number of clutches	(Reed & Bryant, 2004)
3	Total fecundity (lifetime egg production)	(Reed & Bryant, 2004)
4	Lifespan	(Reed & Bryant, 2004)
5	Size of first successful clutch	(Reed & Bryant, 2004)
6	Hatch success of first clutch	(Reed & Bryant, 2004)
7	Hatch success (overall)	(Reed & Bryant, 2004)
8	Fecundity on first day * lifespan	This study
9	Fecundity on first six days	This study
10	Fecundity on first day * number of clutches	This study
11	Number of clutches * lifespan	This study
12	Fecundity on first six days * lifespan	This study
13	Size of first successful clutch * lifespan	This study

[1] Models with multiple predictor variables include both the separate terms as well as the interaction effect, e.g. model 7 features egg productivity on first day, lifespan, and the interaction between these two predictor variables.

within each analysis. The negative binomial component was modelled using a log link, so that the expected value μ_i is given by:

$$\mu_i = e^{\boldsymbol{X_i\beta}} \tag{4}$$

Each model has its own regression coefficient vector $\boldsymbol{\beta}$ and a matrix of predictor variables $\boldsymbol{X}$. The dispersion parameter θ which affects the variance of the observed LRS y is assumed to be constant and is given by (assuming $y > 0$):

$$Var(y|y > 0) = \mu + \frac{\mu^2}{\theta} \tag{5}$$

We compared different models using leave-one-out (LOO) cross validation (Vehtari et al., 2017), and calculated a Bayesian r^2 (Gelman et al., 2019) to estimate the proportion of variation explained by each model.

To assess the evolvability of a female fitness proxy (early fecundity), we used CE data to fit an animal model with a zero-inflated negative binomial distribution with generation, treatment, lifespan, and the generation × treatment interaction as predictor variables. Generation and lifespan were standardized by subtracting the mean value and dividing by the standard deviation. Zero inflation was modelled as a binomial process with logit link and the negative binomial component with a log link. The binomial process was fitted as the probability P_i of observing a false zero for individual i is given by:

$$P_i = \frac{e^{\boldsymbol{X_i\beta_{ZI}+Z_a a_{ZI_i}}}}{1+e^{\boldsymbol{X_i\beta_{ZI}+Z_a a}_{\mathrm{ZI}_i}}} \tag{6a}$$

$$\boldsymbol{X_i\beta_{ZI}} + \boldsymbol{Z_a a}_{\mathrm{ZI_i}} = \beta_{\mathrm{ZI}_0} + \beta_{\mathrm{ZI}_1}x_{1i} + \beta_{\mathrm{ZI}_2}x_{2i} + \beta_{\mathrm{ZI}_3}x_{3i} + \beta_{\mathrm{ZI}_4}x_{1i}x_{2i} + a_{\mathrm{ZI}_i} \tag{6b}$$

Here, $\boldsymbol{X_i\beta_{ZI}}$ represents the fixed effect similar to those used in the previous binomial GLMs (see equation 1), and $\boldsymbol{Z_a}a_{ZI_i}$ represents the animal effect term, which is composed of an incidence matrix $\boldsymbol{Z_A}$ and a_{ZI}, which is a vector containing the breeding values. β_{ZI_0} is the intercept for the log of the odds of observing a false zero and β_{ZI_1} through β_{ZI_4} are the partial regression coefficients for generation, treatment, lifespan, and the generation × treatment interaction respectively. x_{1i} through x_{3i} represent the values for the three predictor variables generation, treatment, and lifespan associated with individual i. The component a_{ZI_i} is the additive genetic effect term or breeding value of individual i for the zero inflation component and is the essential component of the animal model that can be used to estimate the additive genetic variance for a trait (reviewed in Kruuk, 2004).

We used the same predictor variables for the negative binomial component to model the mean (expected) CE, μ_i, for individual i. μ_i may depend on the independent variables and can be defined as:

$$\mu_i = e^{\boldsymbol{X\beta_{NB}+Z_a a}_{\mathrm{NB}}} \tag{7a}$$

$$\boldsymbol{X\beta_{NB}} + \boldsymbol{Z_a}a_{NB} = \beta_{\mathrm{NB}_0} + \beta_{\mathrm{NB}_1}x_{1i} + \beta_{\mathrm{NB}_2}x_{2i} + \beta_{\mathrm{NB}_3}x_{3i} + \beta_{\mathrm{NB}_4}x_{1i}x_{2i} + a_{\mathrm{NB}_i} \tag{7b}$$

in which β_{NB_0} is the intercept (i.e. the value of $\log(\mu_i)$ when all predictor variables equal zero and for an average additive genetic effect) and β_{NB_1} through β_{NB_4} are

the regression coefficients for generation, treatment, lifespan, and the generation × treatment interaction. x_{1i} through x_{3i} again denote the values of the predictor variables generation, treatment, and lifespan for individual i. a_{NB_i} denotes the additive genetic effect for the negative binomial component. We assume the dispersion parameter θ to be constant, i.e., independent of predictor variables.

In the animal model, the vectors a contains the animal breeding values and has a covariance matrix $\mathbf{G}$ which can be derived based on the expected covariance in additive genetic effects between related individuals (Kruuk, 2004). For two individuals i and j, the additive genetic covariance is given by $2\Theta_{ij}\sigma_A^2$, where Θ_{ij} is the coefficient of coancestry (i.e. the probability that a allele drawn at random from individual i is identical by descent to a randomly-drawn allele from individual j); σ_A^2 is the additive genetic variance. The covariance matrix $\boldsymbol{G}$ is given by $\boldsymbol{G} = \boldsymbol{A}\sigma_A^2$, with $\boldsymbol{A}$ being the additive genetic relationship matrix where $A_{ij} = 2\Theta_{ij} = r_{ij}$, i.e. the relatedness between individuals i and j. The relatedness r_{ij} for a trait can also be calculated using the covariance-variance ratio $\frac{cov(i,j)}{var(i)}$, where $var(i) = \sigma_A^2$ so that:

$$\frac{cov(i,j)}{var(i)} var(i) = A_{ij}\sigma_A^2 = cov(i,j) \tag{8a}$$

If we substitute $\boldsymbol{A}$ for A_{ij}, we retrieve the covariance matrix $\boldsymbol{G}$:

$$\boldsymbol{A}\sigma_A^2 = \boldsymbol{G} \tag{8b}$$

In our experiment, paternity is always unknown, and therefore the matrix $\boldsymbol{A}$ cannot be reconstructed with full certainty. We therefore analysed the animal model based on two versions of $\boldsymbol{A}$, one where all daughters of a given female are assumed to be sired by different fathers (half sib model), and one where all daughters are assumed to be sired by the same male (full sib model). Given that female remating in *M. domestica* has been described but only on rare occasions, the latter is likely to be more appropriate. For the animal models, we assumed $Normal(0{,}10)$ as prior distribution for the partial regression coefficients, as well as for the genetic standard deviation ($\sqrt{\sigma_A^2}$) of both the zero inflation as well as the negative binomial component of the animal model.

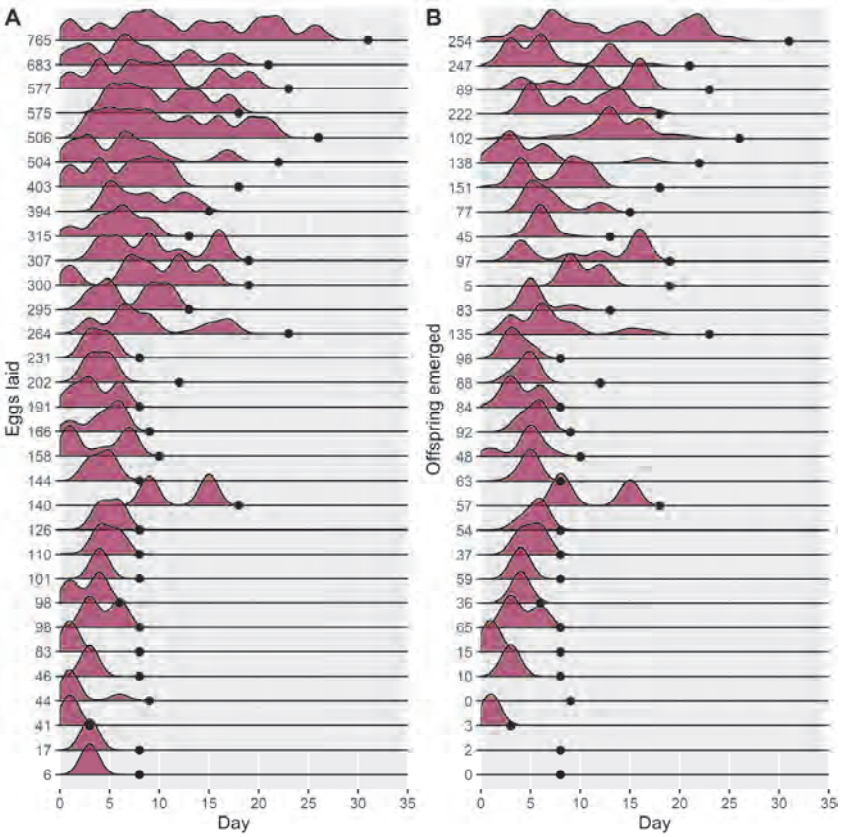

Figure 2: Clutch sizes and offspring emergence rates across female lifetime. Depicted are the daily number of eggs produced (A) and offspring emerged (B) for 31 females relative to the highest number of eggs laid/offspring emerged for each individual female. Day refers to the number of days since the first possible egg-laying opportunity; in (A), this indicates the number of eggs laid on said day, whereas in B it indicates the emergence from days laid on this day. Black dots indicate the time of death. Numbers on the Y-axis denote the cumulative number of eggs laid (A) or offspring emerged (B) by each female over her lifetime. Females were sorted from highest to lowest cumulative egg production in (A) and (B).

Results

Female lifetime reproductive success

To explore the suitability of various traits as proxies for female fitness, we analysed a dataset on female LRS in relation to fecundity-related phenotypes. We found that overall fecundity (lifetime number of eggs laid) varied substantially between females

Table 3: Probability of emergence of offspring from F1 females in response to days since onset of reproduction. Emergence rates were used to fit a zero-inflated negative binomial GLM with time (days since onset of reproduction) as the sole predictor variable and the log of the time the female had access to the egg-laying substrate in hours as an offset.

Parameter	Name	Posterior mean	Posterior SD	95% CI	Probability of direction
β_0	Intercept	1.203	0.12	0.972, 1.440	1.000
β_1	Time	-0.0147	0.012	-0.038, 0.0096	0.887
P_0	Zero inflation rate	0.690	0.022	0.646, 0.732	-
θ	Dispersion parameter	1.658	0.216	1.262, 2.112	-

Table 4: Probability of emergence of offspring from F1 females. Emergence rates were used to fit a binomial GLM with time (days since onset of reproduction) and clutch size as the predictor variables for the analysis per clutch. For whole lifetime emergence rates, we fitted a binomial GLM with lifetime fecundity (total egg count) as the sole predictor variable.

Parameter	Name	Posterior mean	Posterior SD	95% CI	Probability of direction
Per clutch model					
β_0	Intercept	-2.011	0.106	-2.215, -1.802	1.000
β_1	Time	0.007	4.9×10^{-3}	-0.0032, 0.0096	1.000
β_2	Clutch size	0.015	1.1×10^{-3}	0.0124, 0.0167	0.903
Lifetime model					
β_0	Intercept	-0.701	4.33×10^{-2}	-0.787, -0.617	1.000
β_1	Total egg count	-1.13×10^{-3}	5.73×10^{-5}	-1.25×10^{-3}, -1.02×10^{-3}	1.000

(median = 191.0, interquartile range = 99.5 – 354.5, excluding individuals with zero eggs), and within-individual variation was similarly high (median difference between smallest and largest batch = 43, interquartile range = 5 – 79.5, excluding batches with zero eggs). Although a female's fecundity varied between days, this variation did not consistently cluster on any specific time range (Figure 2A), and per-day fecundity was not correlated with time (days since onset of reproduction; Table 3). In the per-clutch analyses, we found that the proportion of offspring emerged was positively correlated with batch size (Figure 3A, 3B; Table 4), but was not correlated with days since onset of reproduction (Figure 3C, Figure 3D; Table 4). In contrast, in the whole-lifespan analysis, the proportion of offspring that emerged was found

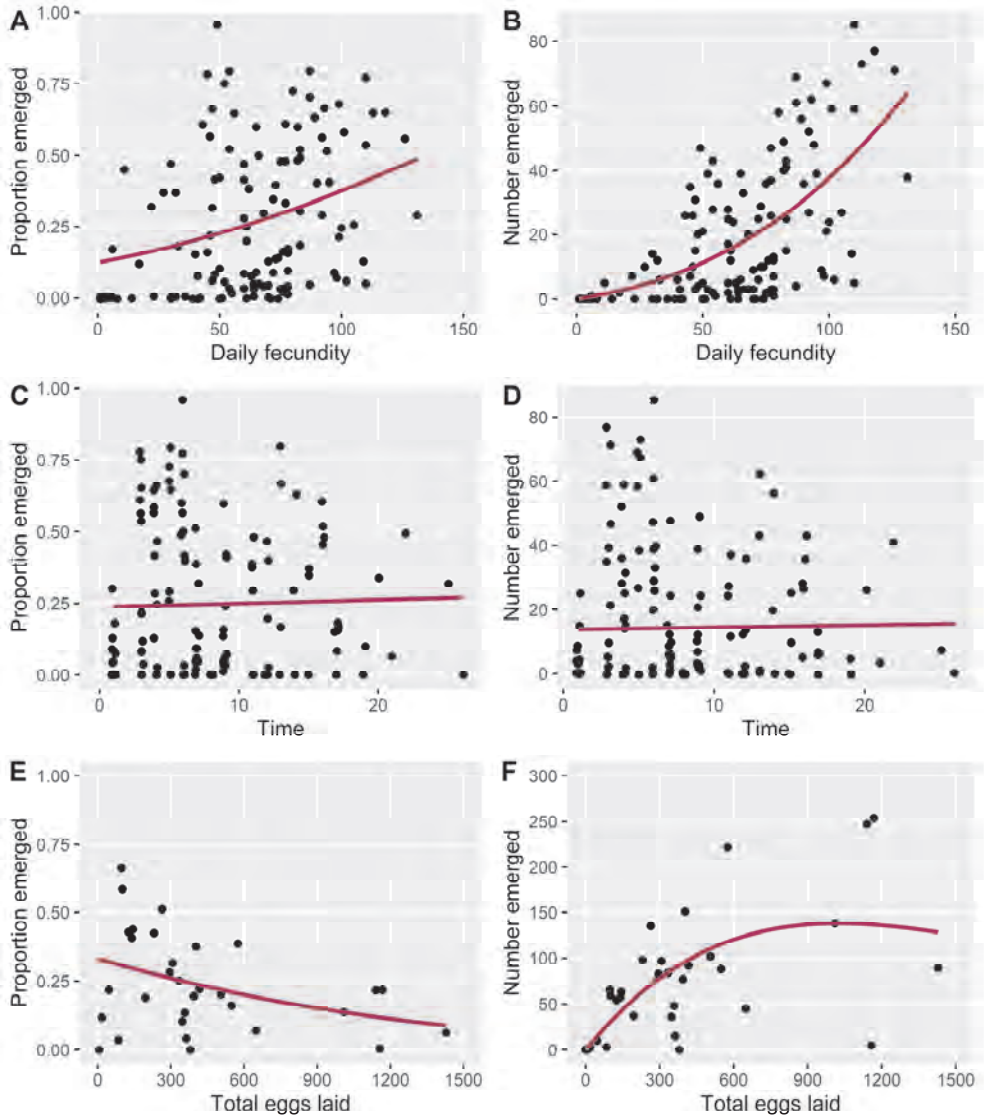

Figure 3: Offspring emergence. Left and right column shows proportion and number of offspring emerged relative to daily fecundity (A & B), time (days since onset of reproduction; C & D), and total number of eggs a female has laid in her lifetime (E & F). Note that A-D denote proportion or number of individuals emerged in a given clutch whereas E & F depict total number of offspring versus total number of eggs laid per female. The pink line indicates the model fit (A, C, E: binomial GLM; B, D, F: binomial GLM multiplied by (B) clutch size, (D) day, or (F) total eggs laid; A-D used clutch size and clutch number as independent variables; E & F used total number of eggs laid).

to decrease with increasing fecundity (Figure 3E), so that in terms of number of offspring emerged females that laid between 900 and 1000 eggs were expected to have the highest LRS (Figure 3F).

We fitted several exploratory models to the female LRS data (Table 2). We found that the model with fecundity during the first six days, lifespan, and the interaction between these two variables had the lowest LOO (model number 11, Table 5). This model did marginally better than a model with only total fecundity (cumulative lifetime egg production), which was previously found to be the strongest predictor for female LRS (Reed & Bryant, 2004). Following these two models, we find that models using either number of clutches (i.e. days on which eggs were laid) or lifespan did equally well, whereas a model featuring both these variables did slightly worse. Similar to (Reed & Bryant, 2004), we find that models using only predictor variables based on the first clutch only (size, hatch rate and female age at first clutch) performed least well.

Table 5: Model fits for female fitness proxies. In each model, we fit a zero-inflated negative binomial GLM to the lifetime reproductive success scores of 31 F1 females. Parameter estimates indicate the (partial) regression coefficients for the negative binomial component. We assumed zero-inflation was constant within each model. For each model, we list the Bayesian r^2, the intercept (β_0) and the predictor variables coefficients, along with the dispersal parameter θ and the zero inflation rate $P(0)$. 'Interaction' refers to the interaction effect term of the two preceding predictor variables (we consider at most 2 predictor variables and their interaction within each model). Models are ordered based on leave-one-out (LOO) cross validation scores (mean difference ± SE).

Model & $\Delta LOO \pm SD$	**Trait**	**Posterior mean**	**Posterior SD**	**95% CI**	**Probability of direction**
12 0	Bayesian r^2	0.619	0.09	0.414, 0.745	-
	Intercept	-0.962	1.192	-3.393, 1.349	0.795
	Fecundity (days 1-6)	0.029	0.009	0.011, 0.047	0.999
	Lifespan	0.181	0.047	0.088, 0.277	1
	Interaction	-0.001	0	-0.002, 0	0.995
	Dispersal coefficient (θ)	2.818	0.915	1.36, 4.909	-
	Zero inflation rate ($P(0)$)	0.081	0.049	0.01, 0.199	-
3 6.5 ± 6.8	Bayesian r^2	0.678	0.136	0.274, 0.782	-
	Intercept	3.401	0.246	2.911, 3.891	1
	Total fecundity	0.003	0.001	0.002, 0.005	1
	Dispersal coefficient (θ)	2.14	0.629	1.107, 3.554	-
	Zero inflation rate ($P(0)$)	0.086	0.049	0.015, 0.203	-

2 10.7 ± 5.7	Bayesian r^2	0.538	0.147	0.148, 0.666	-
	Intercept	3.446	0.277	2.912, 4	1
	Number of clutches	0.187	0.052	0.091, 0.294	1
	Dispersal coefficient (θ)	1.845	0.548	0.946, 3.094	-
	Zero inflation rate ($P(0)$)	0.086	0.05	0.014, 0.203	-
4 11.9 ± 6.3	Bayesian r^2	0.48	0.169	0.091, 0.661	
	Intercept	2.606	0.517	1.582, 3.627	1
	Lifespan	0.075	0.022	0.034, 0.119	1
	Dispersal coefficient (θ)	1.764	0.505	0.91, 2.894	-
	Zero inflation rate ($P(0)$)	0.086	0.049	0.014, 0.202	-
10 13.1 ± 6.2	Bayesian r^2	0.561	0.106	0.244, 0.658	-
	Intercept	3.622	0.386	2.87, 4.398	1
	Fecundity (day 1)	-0.01	0.01	-0.029, 0.01	0.86
	Number of clutches	0.174	0.087	0.011, 0.355	0.981
	Interaction	0.001	0.002	-0.002, 0.004	0.757
	Dispersal coefficient (θ)	1.802	0.547	0.908, 3.044	-
	Zero inflation rate ($P(0)$)	0.085	0.05	0.013, 0.205	-
11 13.9 ± 5.3	Bayesian r^2	0.518	0.135	0.194, 0.678	-
	Intercept	2.386	1.081	0.299, 4.602	0.987
	Number of clutches	0.365	0.231	-0.097, 0.82	0.943
	Lifespan	0.05	0.062	-0.074, 0.174	0.799
	Interaction	-0.007	0.007	-0.02, 0.007	0.87
	Dispersal coefficient (θ)	1.803	0.549	0.896, 3.032	
	Zero inflation rate ($P(0)$)	0.085	0.05	0.013, 0.202	
13 13.9 ± 6.6	Bayesian r^2	0.538	0.154	0.163, 0.707	
	Intercept	2.767	1.545	-0.343, 5.802	0.962
	Size of first clutch	0.001	0.023	-0.043, 0.047	0.507
	Lifespan	0.052	0.073	-0.09, 0.2	0.764
	Interaction	0	0.001	-0.002, 0.002	0.597
	Dispersal coefficient (θ)	1.715	0.519	0.874, 2.879	-
	Zero inflation rate ($P(0)$)	0.085	0.049	0.013, 0.202	-
8 -15.1 ± 6.3	Bayesian r^2	0.524	0.139	0.158, 0.66	-
	Intercept	2.998	0.82	1.379, 4.604	1
	Fecundity (day 1)	-0.014	0.019	-0.05, 0.024	0.775
	Lifespan	0.06	0.036	-0.008, 0.134	0.958
	Interaction	0	0.001	-0.001, 0.002	0.763
	Dispersal coefficient (θ)	1.665	0.497	0.849, 2.787	-
	Zero inflation rate ($P(0)$)	0.086	0.049	0.014, 0.203	-

9	Bayesian r^2	0.438	0.161	0.061, 0.601	-
15.3 ± 7.3	Intercept	3.31	0.383	2.554, 4.065	1
	Fecundity (days 1-6)	0.008	0.003	0.003, 0.013	0.999
	Dispersal coefficient (θ)	1.532	0.444	0.806, 2.527	-
	Zero inflation rate ($P(0)$)	0.08	0.05	0.008, 0.198	-
7	Bayesian r^2	0.279	0.141	0.009, 0.485	-
19.4 ± 7.9	Intercept	3.268	0.57	2.153, 4.392	1
	Overall hatchability	3.433	1.659	0.297, 6.829	0.984
	Dispersal coefficient (θ)	1.254	0.355	0.679, 2.056	-
	Zero inflation rate ($P(0)$)	0.07	0.049	0.004, 0.188	-
5	Bayesian r^2	0.207	0.143	0.003, 0.494	-
19.6 ± 7.6	Intercept	3.764	0.359	3.086, 4.505	1
	Size of first clutch	0.011	0.005	0, 0.021	0.979
	Dispersal coefficient (θ)	1.315	0.378	0.696, 2.15	-
	Zero inflation rate ($P(0)$)	0.081	0.05	0.009, 0.199	-
6	Bayesian r^2	0.096	0.118	0, 0.428	-
23.2 ± 7.9	Intercept	4.308	0.234	3.862, 4.791	1
	Hatchability of first clutch	0.704	0.756	-0.699, 2.299	0.834
	Dispersal coefficient (θ)	1.178	0.331	0.63, 1.912	-
	Zero inflation rate ($P(0)$)	0.08	0.05	0.008, 0.198	-
1	Bayesian r^2	0.072	0.089	0, 0.324	-
23.4 ± 7.9	Intercept	3.83	0.915	2.036, 5.665	1
	Age	0.06	0.086	-0.107, 0.231	0.76
	Dispersal coefficient (θ)	1.179	0.325	0.636, 1.9	-
	Zero inflation rate ($P(0)$)	0.082	0.05	0.01, 0.201	-

Artificial selection for increased female fecundity

Once a suitable fitness estimation framework is established, it can be put to use in various ways such as for artificial selection experiments. It is however not sufficiently clear if artificial selection can efficiently establish sex-specific increases in fitness. We therefore carried out an artificial selection on a fitness proxy in females (total fecundity during first six days after onset of fecundity assays). Female fecundity data were used to fit zero-inflated negative binomial GLMs (see methods), where both the zero inflation and the negative binomial components were modelled using lifespan, generation, treatment and the generation × treatment interaction as predictor variables. For zero inflation, lifespan and generation × treatment

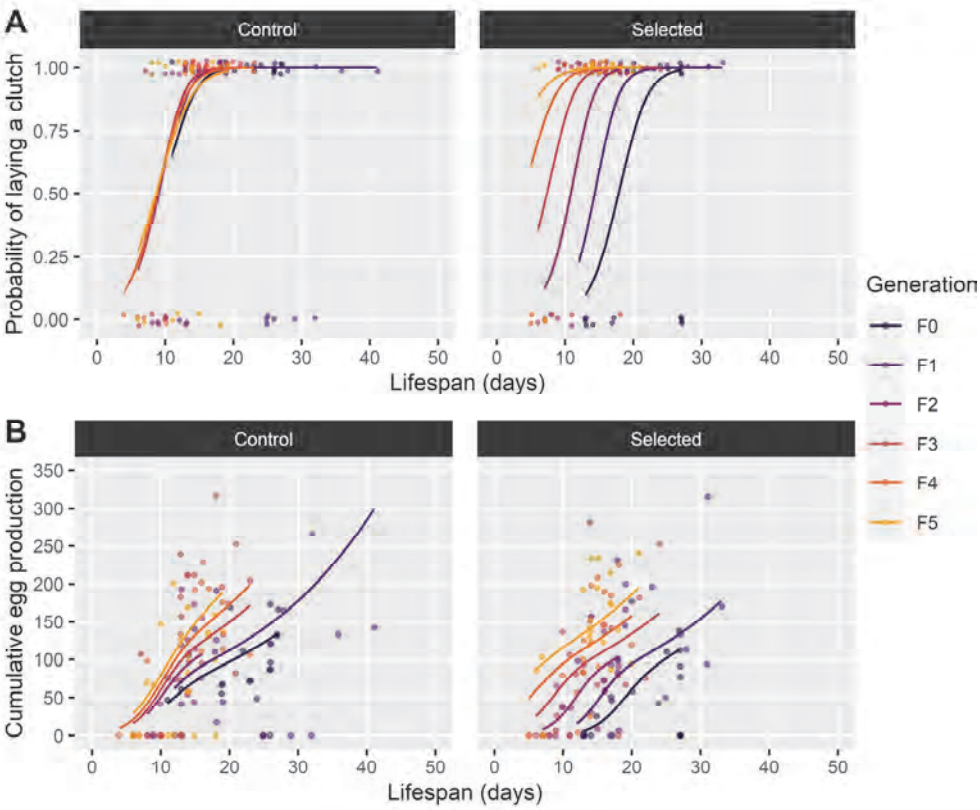

Figure 4: Egg production is positively correlated with lifespan in terms of both the probability that a female will lay at least one clutch (A) and the cumulative number of eggs laid during her lifetime (B). Points indicate individual observations (jittered to reduce overlapping data points) denoting either (A) whether a female laid one or more clutches (1) or not (0), or (B) absolute counts of the cumulative number of eggs laid (B). Lines indicate model predictions generated using parameter estimates from a full sib animal model with generation, treatment, lifespan and the generation × treatment interaction as predictor variables.

interaction were the only negative predictor variables (Table 6). This results in a positive effect on the probability that a female will lay eggs (lifespan: Figure 4A; generation × treatment: Figure 5A). For the negative binomial component, there is a positive effect of both lifespan and generation; these therefore have a positive effect on the number of eggs produced (Table 6; lifespan: Figure 4B; generation: Figure 5B). The generation × treatment interaction effect for the negative binomial component, which was hypothesized to be positive, did not appear to differ from zero as indicated by the overlapping 95% CI and the low probability of direction

Table 6: Model parameter estimates for zero inflation (ZI) and negative binomial (NB) component. Listed are the posterior means and standard deviations (SD), the associated 95% credible intervals, and the probability of direction. Generation and lifespan were standardized by subtracting the mean and dividng by the standard deviation.

Parameter	Name	Posterior mean		Posterior SD		95% CI		Probability of direction	
		Half	Full	Half	Full	Half	Full	Half	Full
Zero inflation component									
β_{ZI_0}	**Intercept**	-8.834	-9.099	3.974	4.418	-17.881, -2.437	-19.02, -1.923	0.999	0.996
β_{ZI_1}	**Generation**	-0.847	-0.41	2.978	3.176	-7.09, 4.923	-7.006, 5.8	0.612	0.548
β_{ZI_2}	**Treatment**	-3.389	-1.556	4.209	4.713	-12.014, 4.871	-10.961, 8.083	0.805	0.641
β_{ZI_3}	**Lifespan**	-8.396	-8.349	2.954	3.019	-14.932, -3.517	-15.028, -3.328	1	1
β_{ZI_4}	**Generation × treatment**	-8.626	-8.602	4.211	4.399	-17.843, -1.376	-18.098, -0.944	0.991	0.986
$\sigma^2_{A_{ZI}}$	**Genetic variance**	252.719	245.307	25.215	26.62	50.939, 715.028	44.33, 713.405	-	-
Negative binomial component									
β_{NB_0}	**Intercept**	4.804	4.8	0.069	0.073	4.668, 4.938	4.654, 4.941	1	1
β_{NB_1}	**Generation**	0.286	0.286	0.08	0.081	0.129, 0.443	0.126, 0.445	1	1
β_{NB_2}	**Treatment**	-0.111	-0.11	0.093	0.099	-0.297, 0.07	-0.307, 0.08	0.886	0.87
β_{NB_3}	**Lifespan**	0.287	0.288	0.057	0.057	0.176, 0.4	0.176, 0.401	1	1
β_{NB_4}	**Generation × treatment**	0.034	0.035	0.103	0.105	-0.168, 0.236	-0.171, 0.241	0.628	0.631
$\sigma^2_{A_{NB}}$	**Genetic variance**	0.009	0.01	0.005	0.005	0.004, 0.254	0.004, 0.261	-	-
θ	**Dispersion parameter**	3.317	3.325	0.392	0.394	2.627, 4.164	2.633, 4.172	-	-

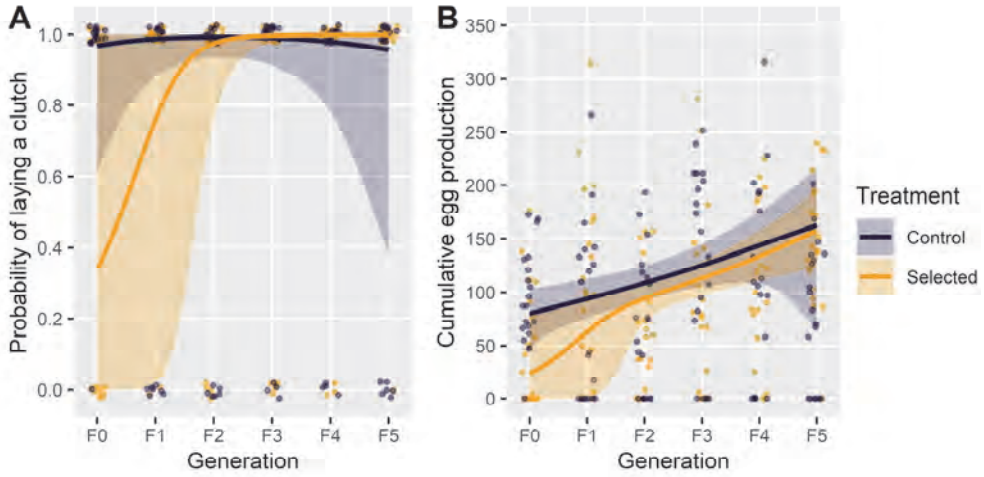

Figure 5: Probability of successful reproduction and cumulative egg production during the selection phase. (A) Proportion of females that laid a clutch during their lifetime (cumulative egg production exceeds 0). (B) Cumulative egg production by females in clutches 1 through 6. Points indicate individual observations (jittered to reduce overlapping data points), lines represent mean predictions generated based on full sib animal model with generation, treatment, lifespan and the generation × treatment interaction as predictor variables. Areas indicate 95% credible intervals for the estimated means. Predicted values were generated assuming lifespan is equal to the mean lifespan over all observations and excluding genetic effects.

(Table 6). The genetic variance for the zero inflation component was estimated to be 245.3 (95% CI: $[44.33, 713.405]$). For the negative binomial component, the genetic variance was very low (σ_A^2: 0.01; 95% CI: $[0.004, 0.261]$). It therefore seems that the genetic variation for laying eggs (i.e. as a yes/no decision) is substantially larger than the genetic variation for the number of eggs that are laid.

Discussion

The housefly *M. domestica* is a model system in which the evolution of SD mechanisms and sex chromosomes can be studied owing to its multifactorial sex determination system. However, assessing fitness variation in this species has received relatively little attention. As a result, there is as of yet no proper methodology to assess or estimate fitness in this species. To understand whether and how adaptive evolution occurs during SD transitions and sex chromosome

evolution, it is necessary to develop a fitness assessment methodology for both males and females. In this chapter, we have presented how different proxies for fitness may be derived in female and male houseflies.

Female sexual function and fitness in houseflies

In houseflies, females have an archetypical sex role, where they produce larger gametes than males but invest less into other aspects of reproduction such as mate searching or competition with other females, which are instead more commonly observed in males. Female houseflies are outwardly passive in terms of courtship behaviour, whereas males engage actively in seeking out females and initiating courtship (Murvosh et al., 1964; Colwell & Shorey, 1975). Females may locate males via male pheromones (Schlein & Galun, 1984), but female pheromones are a stronger attractant to males (Silhacek et al., 1972; Noorman & Den Otter, 2002). In many insect species, females benefit from multiple mating either directly via increased fecundity or indirectly via more genetically variable offspring (Arnqvist & Nilsson, 2000; Birkhead & Pizzari, 2002). Direct effects appear to be absent in houseflies, though transfer of accessory seminal proteins during the first mating was found to increase female fecundity (Arnqvist & Andrés, 2006). Remating could then be advantageous to females if additional doses of seminal products result in additionally increased fecundity. Female remating in *M. domestica* is however considered rare, with its prevalence estimated to be approximately 3% under laboratory conditions; other estimates are however higher and have reached 21% (Leopold, Terranova, & Swilley, 1971; Andrés & Arnqvist, 2001). Remating inhibition is achieved by a seminal product transferred from males to females so that the potential fecundity benefits of remating are largely inaccessible to females (Riemann et al., 1967; Leopold, 1970; Leopold, Terranova, & Swilley, 1971).

However, female remating under natural conditions is undocumented, and may occur more frequently than it does under laboratory conditions (Leopold, 1976). If remating is so strongly inhibited, the effect of a female's mating behaviour on her fitness must involve very high sexual selectivity and resistance to male attempted copulations. In *M. domestica*, fecundity in both sexes is correlated with body size, and males and females alike prefer to copulate with larger mates (Goulson et al., 1999). Adaptive female mate choice would then revolve around her ability to gauge

male size and subsequently to reject small mates c.q. accept large mates. Female fitness was however identical between females engaged in mate choice experiments versus those in no-choice matings (Carrillo et al., 2012), with size of first clutch and egg-to-adult survival of this clutch identical between these two treatments. A different study, however, finds that size of first clutch is a poor predictor for female fitness in *M. domestica*, explaining only 1.5% of the total variation in LRS (Reed & Bryant, 2004). This suggests that the absence of fitness effects in mate choice versus forced mating trials may be an artefact; a more rigorous test of these effects is therefore necessary to determine how mate choice may influence a female's fitness. This may also include a reassessment of the frequency of female remating under natural conditions to test the assumption that female remating is indeed low.

Although many aspects may affect a female's fitness, ultimately her realized fitness is generated by the eggs she herself produces and the yield of emerging offspring from these eggs. A previous study on correlations between fitness and other traits in *M. domestica* found that fecundity, i.e. the cumulative number of fertilized eggs produced by a female during her lifetime, explained 64% of the variation in LRS between females (Reed & Bryant, 2004). Fecundity is therefore a strong candidate proxy for female fitness, particularly if only a single trait is to be assessed. However, because they do not consider interactions between different fitness components, it is not fully possible to determine if other sets of traits may be equally suitable to explain fitness variation. For example, although they find that size of the first clutch is a poor predictor of female fitness, it is a strong predictor of total fecundity (Francuski et al., 2020). If clutch sizes remain constant throughout a female's life, the product of first clutch size and longevity or first clutch size and number of clutches laid could be a less labour-intensive substitute for total fecundity. We find that models with lifespan and either size of first successful clutch (which is always nonzero; model 13) or fecundity on first day after onset of the assay (which may be zero; model 8) are less appropriate for describing variation in LRS. However, a model with fecundity on the first six days and lifespan performed best, and was marginally better than a model using only total fecundity (Table 5, model 12 versus model 3). The two models using a single fecundity assay may not describe variation in LRS as accurately as a model where fecundity over several assays is taken as a predictor variable (the use of the first six days here being somewhat arbitrary), as daily fecundity is simply too variable (Figure 2A). By using multiple fecundity assays,

the effect of the stochasticity of daily fecundity is reduced, resulting in a more accurate fit of the model to the data. As an added benefit, fecundity over the first six days and individual lifespan is more easily assessed than lifetime fecundity, which effectively would require lifespan to be assessed regardless of whether it was used in the statistical model. Taken together, these two fitness components therefore provide the most suitable known fitness proxy for female fitness in *M. domestica*.

Male sexual function and fitness in houseflies

Contrary to female houseflies, males generally have a more overt sexual function and play a more active role in sexual behaviours. Male houseflies indeed engage far more actively in mate searching and courtship initiation than females (Murvosh et al., 1964; Colwell & Shorey, 1975), exhibit more locomotor activity (which presumably aids in mate finding) (Meffert & Bryant, 1992), and are generally more promiscuous than females (Riemann et al., 1967). Male sexual behaviour obviously has a partial genetic basis as suggested by the existence of variation in sexual behaviour between populations (Bryant, 1980), but is also socially affected as a male's mating success, mating speed and copulation duration all were found to be affected by the size of his competitor (though these effects varied between male-male pairs of different populations) (Baldwin & Bryant, 1981). However, male courtship is rather indiscriminate in that males may attempt to mate with other males in absence of females (Murvosh et al., 1964), as well as previously-mated and therefore non-receptive females (Ragland & Sohal, 1973; Meffert & Hagenbuch, 2005). Altogether, this suggests male sexual behaviour is quite variable and therefore is likely to result in fitness variation between males. This is further supported by the notion that females exhibit little remating under laboratory conditions (Arnqvist & Andrés, 2006), which would cause male fitness to be chiefly determined by courtship performance.

Female remating inhibition is primarily induced by males by transferring an accessory seminal product (Riemann & Thorson, 1969; Leopold, 1970; Leopold, Terranova, Thorson, et al., 1971). Presumably, this increases his chances of siring her offspring as female remating is associated with reduced paternity rates of males that mated earlier in many species (Parker, 1970; Gromko & Pyle, 1978; Birkhead et al., 2002). A male's ability to inhibit remating diminishes as he engages in

subsequent matings because its seminal product becomes depleted (Riemann et al., 1967; Leopold, 1970; Leopold, Terranova, & Swilley, 1971), although some *de novo* synthesis may occur between copulations (Leopold, 1970). Moreover, female remating was found to be virtually absent when mated with virgin males (3%), whereas females who mated with multiple-mated males were found to remate at a much higher rate (21%) (Leopold, Terranova, & Swilley, 1971). This suggests the assumption that females do not remate may not always hold; a male that has recently mated a female may not successfully inhibit remating in subsequent female partners, who may instead remate later on. Male fitness is thus not determined only by mating success, but extends to the postcopulatory phase in which female remating must be inhibited to avoid suffering a cost in the form of reduced fertilization success. Interestingly, sperm transfer occurs only during the initial phase of copulation (approximately 10-20 minutes) whereas accessory products continue to be transferred up to approximately 40 minutes after onset of mating (Murvosh et al., 1964; Leopold, Terranova, Thorson, et al., 1971), with disruption of mating at this time point resulting in reduced efficacy of female remating inhibition (Arnqvist & Andrés, 2006). Under natural conditions, matings may be more frequently disrupted than under laboratory conditions, leading to female remating and males to potentially lose paternity depending on seminal product transfer efficiency. The efficacy with which sperm and accessory products are transferred during mating may be an interesting proxy for male fitness. Here, the fitness proxy is effectively the number of successful fertilizations over time, with higher fitness being attained by males that achieve identical fertilization success in less time or those that achieve higher success in equal time.

Female remating allows for postcopulatory sexual selection to occur (Parker, 1970; Birkhead & Pizzari, 2002), which favours the evolution of traits that act after copulation to affect fitness. Such traits include those that inhibit female remating (e.g. via the transfer of a seminal product described above; Leopold, Terranova, Thorson, et al., 1971) or that bias the rate at which sperm from different mates are used (i.e. sperm precedence (Parker, 1970) or cryptic female choice; reviewed in Miller & Svensson, 2014). Such effects are becoming increasingly well-documented in other species such as *D. melanogaster*, in which both males and females are promiscuous. When a female *D. melanogaster* mates with multiple males, males that mate first tend to lose fertilization success as latter males may induce the female to

shed sperm from previous mates to increase the proportion of offspring sired by these latter males (though to different success depending on how many males mate with a given female; Laturney et al., 2018). In *M. domestica*, postcopulatory selection has received considerably less attention, but is nonetheless inferred to occur as females exhibit different remating rates and oviposition rates depending on their own genotype as well as that of their mates (Andrés & Arnqvist, 2001). In addition to variation in a male's ability to inhibit remating, sperm precedence and cryptic female choice may play a yet unknown role in *M. domestica*. Many studies on housefly male fitness make use of single male-female pairs or similar small-scale setups (e.g. (Goulson et al., 1999)), whereas under natural conditions houseflies generally occur in large, aggregated groups where both sexes may mate multiply. Considering male fitness in a social context may elucidate hitherto-unknown but important factors by which it is affected, such as in *D. melanogaster* where social context affects male and female reproductive behaviour (Krupp et al., 2008; Billeter et al., 2012; Garbaczewska et al., 2013; Laturney & Billeter, 2016).

Altogether, male fitness in houseflies seems to be largely determined at the mating phase, as the perceived absence of female remating would lead a male that mates with a female to have full paternity over her offspring. Suitable male fitness proxies therefore must reflect variance in mating success. However, female promiscuity may be more prevalent in houseflies than previously considered, and female remating inhibition by males may be imperfect. Pending further investigation of the role of female remating, male fitness proxies that describe male fitness in terms of acquired matings per unit of time are likely to best reflect the nature in which male fitness is determined in this species under natural conditions. Should female remating be prevalent in *M. domestica*, then these fitness proxies must be adapted to reflect not the number of matings per unit of time, but the number of fertilizations per unit of time. Additionally, they must incorporate both "defensive" traits against losing fertilization success (e.g. by inhibiting their female mates from remating) as well as "offensive" traits to acquire such success (e.g. by inducing previously-mated females to mate with them).

Evolutionary response of a female fitness proxy under artificial selection

Artificial selection has previously been employed in numerous ways; applications to fitness in a sex-specific context have however proven troublesome, not in small part due to the complexity of establishing a sex-specific fitness framework. Establishing strains with increased fitness in one sex may benefit fundamental evolutionary biological studies on topics such as intra- and interlocus sexual conflict (conform Rice (1996, 1998), though such approaches are more adequately described as experimental evolution) but also in an applied context. For example, female fecundity may be a bottleneck in the production of various biological goods, such as in feed production (discussed for *M. domestica* in e.g. Francuski et al. (2020).

To assess whether fitness proxies can be used in artificial selection experiments to establish sex-specific increases in fitness, we set up an artificial selection experiment in which we selected females for increased fecundity during the first six days after onset of egg production. We find that the selection regime was partly able to positively affect this trait; however, in both the control and selected strain, we observe an increase over time, rather than only in the selected strain. Contrary to our expectation, the generation × treatment interaction term, which describes the difference in the effect of generation between the control strain and the selected strain, was found to be small (Table 6). We therefore find that the control strain and the selected strain both showed increased fecundity over the course of the experiment, but they did not differ in the rate at which this change occurs (Figure 5). One explanation is that in the control strain, we cannot randomly sample from all females but instead can only use females that manage to produce clutches and thereby generate offspring. Moreover, small clutches tend to suffer from reduced emergence rates (Figure 3A), so that a small bias towards higher-fecundity females exists in the control group similar to the selected group. An additional selection regime for reduced fecundity might provide a stronger contrast to the selection regime for increased fecundity applied here, and may thereby provide stronger evidence for the effect of artificial selection.

Although we observed a relatively weak response in terms of absolute number of eggs laid by females, we found a stronger change in the proportion of females that laid at least one clutch, which increases under artificial selection for increased clutch sizes (Table 6, Figures 4A and 5A). In the selected group, we initially

observed a higher prevalence of females that failed to produce any clutch whatsoever, but over 6 generations of selection this proportion drops close to zero so that virtually all females lay at least one clutch. In the control group, only a slight increase in the proportion of females that lay at least one clutch is observed. One explanation for this might be that the conditions under which female fecundity is assessed in our experiment are actually unfavourable. Females then put off investing into egg production for the time being in hopes of encountering more favourable conditions later on. Selection for increased fecundity here would then reflect selection for an increased tendency to lay eggs, even when conditions are deemed unfavourable. This would also explain the difference in the estimated genetic variance for the zero inflation component versus that for the negative binomial component.

Synthesis

Given our interest in using *M. domestica* as a model for sex determination and sex chromosome evolution, we focus here on fitness proxies that incorporate sex-specific function because these may be principally affected by these processes and may have direct influence on an individual's fitness. In females, fitness appears to be chiefly determined by her own capacity to produce eggs, and therefore fitness proxies may primarily involve fecundity-related traits (Reed & Bryant, 2004). Additionally, a female's ability to assess mate quality and acquiring mating with preferred males (or inversely preventing mating with unpreferred males) may indirectly affect her fitness via ensuring that her offspring receive good genes from their fathers. In males, mating success is likely to be the strongest determinant of fitness, and accordingly courtship performance and similar traits may be suitable proxies. However, for both sexes it is assumed that male inhibition of female remating is strong and therefore that there is little scope for female promiscuity and subsequent postcopulatory sexual selection to occur. Given that repeated matings reduces male efficacy of remating inhibition in females (Leopold, Terranova, & Swilley, 1971), it seems likely that under natural situations as well as less simplified laboratory situations (e.g. setups involving multiple males and females) this assumption will be invalid. Instead, female promiscuity may occur, and variance in postcopulatory performance may contribute to fitness variance in males as well as in females.

Aside from sex-specific traits, certain aspects of *c.q.* proxies for fitness may however be non-sex-specific or be relevant to reproductive function in both sexes such as lifespan, development time, locomotor activity, and body size. Including such traits in a definitive fitness estimation methodology may consolidate fitness estimates which involve traits that are sex-specific. For example, body size has been found to positively affect reproductive success in many species including houseflies (Black IV & Krafsur, 1987), where it is positively correlated with fitness in both sexes (Baldwin & Bryant, 1981).

Alternatively, although we focus here on fitness assessments by measuring specific traits of individuals, fitness effects may also be evaluated with experimental evolution, i.e. using population-scale experiments to assess invasibility (Metz et al., 1992; Mylius & Diekmann, 1995). For example, Hamm et al. (Hamm & Scott, 2008; Hamm et al., 2009) assessed the fitness of III^M-bearing males relative to Y^M-bearing *M. domestica* males by establishing populations with known initial frequencies of both types of males. These populations were maintained for some time (varying from several generations to several years) after which the frequencies of the different types of males were assessed again. Although the results differed between the two studies in the sense that in one study III^M-bearing males were found to decrease in frequency (Hamm & Scott, 2008) whereas in the other these males increased in frequency (Kozielska, 2008; Hamm et al., 2009), such approaches may still be applied to investigate fitness differences. An intrinsic benefit of this type of approach is that no specific fitness assessment needs to be carried out as fitness is evaluated as part of the ability of individuals to successfully reproduce under (quasi-)natural conditions, and many aspects that can affect fitness are evaluated in the process. Care must however be taken as each distinct population only provides one estimate of potential fitness differences between the type of males so that the statistical power of these approaches is generally low (Kawecki et al., 2012); for example the use of 4 replicates in Hamm et al. (2009) is equivalent to sample size of 4. Additionally, the use of small populations may lead to a relatively large influence of genetic drift or sampling error. Provided that these issues are accounted for, the use of population-scale fitness assessment may nonetheless proof a powerful tool for detecting fitness differences between different populations, strains, or genotypes.

The housefly *M. domestica* holds promise as a model system for studies on sex determination and sex chromosome evolution, but substantial efforts are still required to realize this potential, particularly in the context of fitness estimation. Here, we have discussed the development of and illustrated the use of sex-specific fitness proxies which take into account the different roles females and males assume in reproduction in this species. Although further research will be necessary to determine the extent to which female promiscuity occurs and how this affects the manner in which individuals of both sexes may maximize their fitness, the principles laid out in this chapter provide a foundation for the development of fitness estimation methodologies in future studies.

Acknowledgements

We thank Ljubinka Francuski for assistance in the introgression experiment, Xuan Li for providing data on the sex determination genes in the Spanish housefly strains (SPA1 to SPA5), and the Evolutionary Genetics cluster for fruitful discussion. MAS was supported by an Adaptive Life grant awarded to J-CB, IP, and LWB; TC was supported by an Erasmus+ scholarship awarded by the European Commission.

Supplementary Table 1: Model parameter estimates for zero inflation (ZI) and negative binomial (NB) component including the F6 data. Listed are the posterior means and standard deviations (SD), the associated 95% credible intervals, and the probability of direction. Generation and lifespan were standardized by subtracting the mean and dividng by the standard deviation.

Parameter	Name	Posterior mean		Posterior SD		95% CI		Probability of direction	
		Half	Full	Half	Full	Half	Full	Half	Full
Zero inflation component									
β_{ZI_0}	Intercept	-9.656	-9.982	3.976	4.28	-18.5, -3.159	-19.509, -2.888	1	0.999
β_{ZI_1}	Generation	k-1.06	-0.902	2.736	2.964	-6.681, 4.337	-7.001, 4.95	0.66	0.626
β_{ZI_2}	Treatment	0.467	2.236	4.117	4.593	-7.723, 8.789	-6.619, 11.822	0.543	0.691
β_{ZI_3}	Lifespan	-9.617	-9.653	3.257	3.276	-16.746, -4.189	-16.889, -4.177	1	1
β_{ZI_4}	Generation × treatment	-4.405	-5.051	3.615	3.949	-12.204, 2.1	-13.556, 1.999	0.907	0.919
$\sigma^2_{A_{ZI}}$	Genetic variance	308.587	292.456	28.948	28.223	66.132, 841.363	60.043, 807.011	-	-
Negative binomial component									
β_{NB_0}	Intercept	4.735	4.726	0.066	0.07	4.602, 4.864	4.584, 4.859	1	1
β_{NB_1}	Generation	0.175	0.175	0.07	0.071	0.037, 0.313	0.035, 0.315	0.993	0.993
β_{NB_2}	Treatment	-0.071	-0.069	0.09	0.094	-0.251, 0.104	-0.256, 0.115	0.786	0.771
β_{NB_3}	Lifespan	0.279	0.281	0.058	0.058	0.167, 0.393	0.168, 0.396	1	1
β_{NB_4}	Generation × treatment	0.086	0.086	0.093	0.095	-0.099, 0.269	-0.101, 0.273	0.824	0.817
$\sigma^2_{A_{NB}}$	Genetic variance	0.01	0.013	0.005	0.005	0.004, 0.261	0.005, 0.277	-	-
θ	Dispersion parameter	3.081	3.102	0.349	0.355	2.464, 3.828	2.48, 3.874	-	-