Sexual selection on male body size, genital length and heterozygosity: consistency across habitats and social settings

Megan L. Head¹ *, Andrew T. Kahn¹, Jonathan M. Henshaw¹, J. Scott Keogh¹ & Michael D. Jennions¹

¹Division of Ecology and Evolution, Research School of Biology, The Australian National University, Canberra, Australia

* Corresponding author: megan.head@anu.edu.au

Summary

1. Spatial and temporal variation in environmental factors and the social setting can help to maintain genetic variation in sexually selected traits if it affects the strength of directional selection. A key social parameter which affects the intensity of, and sometimes predicts the response to, mating competition is the operational sex ratio (OSR; ratio of receptive males to females).

2. How the OSR affects selection for specific male traits is poorly understood. It is also unclear how sexual selection is affected by interactions between the OSR and environmental factors, such as habitat complexity, that alter key male-female interactions such as mate encounter rates.

3. Here, we experimentally manipulated the OSR and habitat complexity and quantified sexual selection on male mosquitofish (Gambusia holbrooki) by directly measuring male reproductive success (i.e. paternity).

4. We show that, despite a more equitable sharing of paternity (i.e. higher levels of multiple paternity) under a male-biased OSR, selection on focal male traits was unaffected by the OSR or habitat complexity. Instead, sexual selection consistently, and significantly, favoured smaller bodied males, males with higher genome wide heterozygosity (based on >3000 SNP markers), and males with a relatively long gonopodium (intromittent organ).
5. Our results show that sexual selection on male body size, relative genital size and heterozygosity in this system is consistent across environments that vary in ecological parameters that are expected to influence mate encounter rates.

Keywords: body size, environmental heterogeneity, HFC, poeciliid, reproductive success

Introduction

Variation in the strength and form of sexual selection has generated enormous diversity in morphology, behaviour and physiology between the sexes, across populations and among species (Pfennig & Pfennig 2010). Field studies have shown that sexual selection can vary across populations and over time (e.g., Kasumovic et al. 2008; Wacker et al. 2014). Spatio-temporal variation is thought to be important to maintain genetic diversity (e.g., Felsenstein 1976; Ellner & Hairston 1994) and, more specifically, it is thought to slow the erosion of additive genetic variation for sexual traits (Day 2000; Holman & Kokko 2014) that are usually under strong unidirectional selection (Hoekstra et al. 2001). Variation in the direction of sexual selection across heterogeneous environments, and genotype-by-environment effects, are often posited as a partial explanation for the persistence of additive genetic variation in sexual traits (Hunt & Hosken 2014), but relatively few studies address this question by taking an experimental approach to manipulating the environment (Cornwallis & Uller 2010).

Sexual selection on focal traits is known to vary among populations or between breeding cycles, but we rarely understand why (Janicke et al. 2015). Most research on variation in sexual selection involves long-term observational studies of wild populations that draw inferences based on annual or seasonal changes in ecological or social factors that can plausibly be linked to shifts in mate availability (e.g., Kasumovic et al. 2008; Wacker et al. 2014). Unfortunately, many ecological and social parameters covary in nature so it is difficult to identify which parameters actually drive
variation in sexual selection. Similar problems arise when comparing populations to try to identify factors that affect sexual selection. For example, population differences in male guppy (*Poecilia reticulata*) colouration are attributed to habitat differences in predation (e.g., Endler 1980), but predation also affects demography (e.g., Arendt *et al.* 2014), which could independently affect selection on male sexual signals. Ultimately, experiments are required to determine whether specific environmental or social factors moderate selection for specific traits.

Sexual selection studies often focus on the role of the social environment, especially the operational sex ratio (OSR: the ratio of sexually receptive males to females) (review: Kvarnemo & Ahnesjo 1996; Shuster 2016). The OSR is, by definition, a measure of the intensity of competition for mates, but its value in predicting the strength of sexual selection is much debated (e.g. Clutton-Brock & Parker 1992; Klug *et al.* 2010). Variance in male reproductive success (i.e. the opportunity for selection) is expected to increase when the OSR is more male-biased due to greater competition for mates (Emlen & Oring 1977; Kvarnemo & Ahnesjo 1996; Shuster 2016). This is not inevitable, however, as it depends on idiosyncratic features of each species’ biology (e.g., is harem defence easier or harder as the OSR becomes more biased?) (Jennions *et al.* 2012; Klug *et al.* 2010; meta-analysis: Moura & Cardoso Peixoto 2013). The OSR might affect the various components of sexual selection in different ways so that the observed effects depend on the type of mating system, and which traits were measured (Kokko & Rankin 2006). For example, Weir *et al.* (2011) noted that as the OSR becomes more biased the competitive sex tends to become more aggressive, but courts less. This will lead to different estimates of how the OSR affects selection on sexual traits depending on whether one measures traits involved in intra- versus intersexual selection (Fitze & Le Galliard 2008; Head *et al.* 2008 and references therein). Finally, inconsistent effects of the OSR on sexual selection might arise because its effects are context-dependent and modulated by other environmental factors (e.g., predation risk).
One ecological parameter that is of special interest is habitat complexity, because it can have profound effects on sexual selection (Myhre et al. 2013). For example, the transmission of mating signals depends on the habitat type: open habitats generally allow transmission of signals over greater distances. Such differences affect how females perceive and assess male sexual signals and alter selection on males (e.g. sensory bias/drive - Boughman 2002; Endler & Basolo 1998). Habitat complexity can also affect selection on traits that affect male fighting success. For example, recent studies report higher aggression between male sticklebacks (Gasterosteus aculeatus) in more open habitats that might generate stronger selection for larger body size than in closed habitats (Lackey & Boughman 2013). Finally, the OSR and habitat complexity are both likely to alter mate encounter rates, and habitat complexity can generate small-scale variation in local OSRs and densities (i.e., those directly experienced by a female) (Myhre et al. 2013). Such effects could alter the intensity of male-male competition and opportunities for female choice. For example, in guppies greater habitat complexity reduced interference competition between males and increased female mating receptivity (Hibler & Houde 2006).

Here we conduct an experiment to investigate how habitat complexity and the adult sex ratio (which for convenience we refer to as the OSR since male mosquitofish constantly mate with females) affect sexual selection on male mosquitofish Gambusia holbrooki. This species is well suited to test how these two factors affect sexual selection because it inhabits habitats that vary greatly in their structural complexity (Pyke 2005), and adult sex ratios show major shifts over the breeding season (Kahn et al. 2013). We used SNP-based paternity analysis to determine male reproductive success and thereby quantify sexual selection on three focal male traits that have previously been identified as putative targets of sexual selection (male body size - McPeek 1992; Pilastro et al. 1997: residual gonopodium length - Head et al. 2015; Kahn et al 2010: and genome wide heterozygosity - Vega-Trejo et al 2017).
In mosquitofish, males incessantly attempt to mate by approaching females from behind and thrusting their gonopodium (a modified anal fin used to transfer sperm) into her gonopore (Bisazza 1993; Bisazza & Marin 1995). Male size and residual gonopodium size are expected to be important determinants of reproductive success because large males can dominate access to females (Bisazza & Marin 1991), and female mate choice (in the form of longer association times) seems to favour larger males with a relatively long gonopodium (Bisazza et al. 2001; Head et al. 2015; Kahn et al. 2010; McPeek 1992). However, smaller males are more adept at sneaking copulations with females (Pilastro et al. 1997). These contrasting selection pressures suggest that sexual selection on male traits, resulting from the balance of female mate choice and male sexual coercion, might depend on the social setting. For instance, as populations become more male biased selection might tilt in favour of male ability to sneak copulate and favour smaller over larger males and weaken selection for a longer gonopodium, as female choice becomes less important. In contrast, complex habitats might make it easier for females to evade sexual coercion and to exert mate choice such that overall sexual selection might be expected to favour large males with a long gonopodium. Finally, experimental inbreeding studies show that heterozygosity is positively correlated with male reproductive success, (Vega-Trejo et al. 2016), but the effect of natural, standing variation in heterozygosity is unknown, as are the specific traits whose expression is negatively affected by a decline in heterozygosity (i.e. the proximate mechanisms driving sexual selection against inbred males). Here we aim to directly quantify how OSR and habitat complexity affect sexual selection on these traits, and to determine whether differences in habitat complexity mediate the potential effects of the OSR on sexual selection.

Methods

Experimental design

We independently manipulated the adult sex ratio and habitat complexity in pools (1m diameter, 15cm depth) in a greenhouse using a 2 x 2 factorial design. We had two levels of habitat complexity.
In the ‘simple’ habitat the pool floor was lined with gravel, the pool walls were lined with white plastic, and there was no vegetation or cover. The ‘complex’ habitat was the same, but we added a network of white plastic baffles to create multiple, interconnected compartments (Fig. S1). Manipulating habitat complexity in this way ensured that the manipulation was applied evenly across the pool. The manipulation is similar to that used by Hibler & Houde (2006) who found that increased visual isolation in complex habitats altered sexual behaviour in guppies. We also had two levels of OSR. The female-biased OSR consisted of 10 males and 20 females in a pool. The male-biased OSR consisted of 10 males and 5 females in half a pool. We avoid confounding changes in fish density and the number of males by adjusting the pool size to keep the number of males and the overall density of fish constant. Fish density across all treatments was approximately 1 fish per 4 litres, which falls well within *G. holbrooki* densities in the wild (e.g. Jordan *et al.* 1998) and those used in previous studies of other poeciliid fishes (e.g. Marriette *et al.* 2010; Devigili *et al.* 2015; Hibler & Houde 2006). Any treatment differences are therefore due to the OSR and/or habitat complexity, and not to demographic parameters that often covary with OSR (see Head *et al.* 2008).

We set up six blocks: each comprised one replicate per treatment (*n* = 6 blocks x 4 treatments x 10 males = 240 males in 24 replicates). Our experimental design was sufficient to detect medium to large effect sizes (see results).

**Experimental protocol**

We used *G. holbrooki* from ponds in Canberra, Australia (35°14’27”S, 149°5’27”E and 35°14’13”S, 149° 5’55”E). These ponds are less than 2km apart and likely to be connected during periods of high rainfall. Experimental males were caught from the wild. Experimental females were lab-reared offspring of wild caught females. This ensured they were virgins at the beginning of our experiment. We collected the mothers of experimental females from the wild and allowed them to give birth. The fry were then placed in 3l aquaria in groups of up to five. From 4 weeks of age onward these fry
were checked weekly for signs of maturation. As soon as we could determine their sex (elongation of
the anal fin for males, development of eggs visible through the body wall for females) fish were
placed in single sex tanks. Elongation of the anal fin occurs well before males are ready to mate, so
we are sure that these fish were virgins at the commencement of the experiment. Virgin females
were 3-9 months old when used in our experiment. The use of virgin females ensured that all
offspring were sired by males from our experimental pools.

Importantly, prior to placement in experimental pools, both sexes underwent a priming period. This
mimicked the experimental conditions that fish would later experience to ensure that paternity
results reflected the treatments experienced and not a sudden change from stock to experimental
conditions. For priming, focal males and females were placed in experimental pools with the
appropriate number of individuals of the opposite sex. Focal males were placed with stock females,
and focal females were placed with stock males whose gonopodium tip had been removed to
prevent sperm transfer (Mautz 2011). After four days of priming, focal fish were placed directly into
their respective experimental treatments, and stock fish were returned to stock tanks.

Once in experimental pools focal fish had 14 days to interact and mate. They were fed thawed
frozen *Artemia* nauplii twice daily. The female-biased treatments were fed twice the amount of food
as the male-biased treatments since there were twice as many fish.

Males were euthanised after being removed from the experimental pools. We photographed their
left side alongside a microscale using a digital camera (Nikon Coolpix 5700) mounted to a dissecting
microscope (Leica Wild MZ8). Males were then preserved in absolute ethanol and stored at -20°C.
We later measured male standard length and gonopodium length in *Image J*. Since male Gambusia
have determinate growth (Zulian et al 1993) measuring their size after the experimental treatment gives an accurate measure of their size during the treatment when competing for mates.

Once females were removed from the experimental pools they were anaesthetised in ice slurry, photographed (see above) and then placed individually in 1 l tanks. Each tank contained a gravel substrate, plastic aquarium plants and a mesh divider to reduce maternal cannibalism. Tanks were checked twice daily for fry until the female had either produced two broods, or three months had passed. We collected fry from two broods to increase our sample sizes. Females were kept at 27°C±1°C on a 14:10 light:dark cycle and fed live Artemia twice daily. When a female gave birth she was placed in a new 1 l tank if it was her first brood. If it was her second brood she was euthanised and preserved for genotyping. All fry were euthanised (< 24h after birth), and preserved in family groups of up to 10 fry/vial.

Sampling for paternity analysis

To determine male reproductive success we took tissue samples from up to five mothers (on average 4.1 per pool, n = 100 in total), all possible sires (10 per pool, n = 240 in total) and all offspring from the selected mothers for each pool (mean: 35.2 per pool, n = 844 in total). In the male-biased OSR treatment we therefore sampled all mothers that gave birth, and in the female-biased OSR treatment we randomly sampled five females that gave birth. By so doing, we ensured that our power to detect multiple paternity and to detect selection on male traits was similar in the male-biased and female-biased treatments (i.e. in both treatments we collected data from five females). This sampling approach did not bias our estimates of selection (see Supplementary Information). DNA was extracted from the tail muscle/caudal fin for adults and from the whole body
(excluding head) for fry, using Qiagen DNeasy Blood and Tissue Kits (Qiagen, Victoria, Australia) following the manufacturer’s instructions.

After extraction, DNA samples were sent to a commercial genotyping service - Diversity Arrays. The details of the process are described in the Supplementary Materials (see also Booksmythe et al. 2016). We obtained a data set of approximately 3171 SNPs with an average call rate of 97.7% and a reproducibility rate of 99.3%. From the selected SNPs we calculated a Hamming Distance Matrix of all 1185 individuals (potential sires, mothers, and offspring) to determine paternity. Recent studies show that as few as 30 optimized SNPs are sufficient to differentiate among 100 000 individuals using Hamming Distance values (HDV) (Hu et al. 2015). All fry were lined up against their mother and siblings and the HDVs evaluated to cross check for any sample mix ups. None were detected. HDVs were then compared against each of the 10 potential sires. The sire/fry with the lowest value was considered a match. We could assign paternities unambiguously for all 844 fry. Of these, 740 fry were from first broods and 104 from second broods.

Heterozygosity

We estimated heterozygosity (H) as the number of SNP loci that were scored as heterozygous divided by the total number of successfully classified loci (L) for each male who was a potential sire in the experiment ($F_{het}$: Vega-Trejo et al (2016). This is essentially a measure of genome wide heterozygosity. $F_{het}$ is identical to $1 - F_{hom}$ in Bérénos et al. (2016); and to H/L in Szulkin et al. (2010).

Data analysis

Sexual selection on males
To determine which male traits influenced his reproductive success and whether this varied across socio-environmental contexts we ran a Generalised Linear Mixed Model (GLMM). We treated the number of offspring each male sired as the response variable. OSR and habitat complexity were specified as fixed factors. Male standard length (logged), residual gonopodium length (i.e. residuals of the regression of log gonopodium length on log male length) and heterozygosity were included as covariates. Interactions between each of the three male traits and the two experimental factors were included in the model. We did not include interactions between traits, interactions with random effects or higher order interactions as these were not key to the hypotheses being tested and we did not want to over parameterise our models. We treated pool as a random effect to avoid pseudoreplication and specified a Poisson error structure. To account for overdispersion we included individual as a random effect (Harrison 2014). Following this correction our data was underdispersed (dispersion parameter = 0.13) and thus conservative. Although some of the male traits in the model were significantly correlated (see results) collinearity was not a problem for our model because these correlations were weak. A linear mixed effects model using power-transformed offspring number with the same model structure described above gave qualitatively similar results. Although the residuals from both models looked approximately normal, the linear model provided a worse fit to our data (log-likelihood test: $\chi^2 = 250.2$, $p<0.001$) and so the results presented here are from the poisson model.

To allow comparison between studies we also calculated the effect sizes for each of the parameters in our model. Effect sizes were calculated using the P values from the model following the formulae given in Lipsey & Wilson (2001) By convention we refer to $r=0.1$, 0.3, and 0.5 as small, medium, and large effect sizes respectively (Cohen 1988). To calculate effect sizes we used $N = \text{pools}$ (i.e. the number of independent replicates) rather than $N = \text{males}$, because the latter potentially
underestimates the effect size. The sign of the effect size is based on the direction of the estimate in the model.

Neither the OSR nor habitat complexity influenced the relationship between the number of offspring sired and any of the male traits (see Results) so we calculated experiment-wide selection gradients using a linear multiple regression (Lande & Arnold 1983). We treated the relative number of offspring a male sired (calculated within pools) as the response variable and log male length, residual gonopodium length and heterozygosity as predictor variables. All predictor variables were standardised across the experiment (mean = 0, s.d. = 1). Standardising traits within pools gave very similar selection gradient estimates. Significance values were obtained from the same model except that the relative number of offspring sired was power transformed to account for its non-normal distribution and pool identity was treated as a random effect to avoid pseudoreplication. We estimated non-directional selection on each trait using the recently developed method of Henshaw & Zemel (2016), which quantifies the total strength of non-directional selection of any kind (e.g. stabilising or disruptive selection: Brodie et al. 1995). We found no evidence for non-directional selection on any of the three traits (see Supplementary Information for details).

Incomplete sampling of an individual’s mates or offspring can lead to systematic bias in estimates of sexual selection (e.g. the opportunity for sexual selection and the Bateman gradient: Mobley & Jones 2013; Jones 2015). It is therefore possible that our subsampling of females in the pools with female-biased OSR might introduce bias in our estimates of selection gradients on male traits (i.e. some males that sired offspring will not be noted as such if the female with whom they mated is not among the five sampled females). To investigate the extent of this problem we ran simulations to test the effects of subsampling. These simulations demonstrate that any bias is negligible (see Supplementary Information for details).
We ran a generalised linear mixed model (GLMM) with the number of sires per brood as the response variable and OSR, habitat and their interaction as fixed effects. Pool identity was treated as a random effect to avoid pseudoreplication, and we specified a poisson error distribution. This model gave qualitatively similar results to a linear mixed model on transformed data.

Since Bolker et al. (2009) do not recommend including random factors with fewer than 5-6 levels we report results without block as a random effect. Note, however, that including block as a random effect in addition to that of pool did not influence any of our results. All analyses were conducted in R version 3.2.0 (R core development team 2015). The lme4 package (Bates et al 2015) was used to construct models and p values were obtained using lmerTest.

All data are deposited in DRYAD (Head et al., 2017).

Results

Sexual selection on males

More heterozygous males were smaller (Pearson’s correlation: \( r=-0.164, t_{(234)}=1.354, p=0.012 \)), and had a relatively longer gonopodium for their body size (Pearson’s correlation: \( r=0.187, t_{(232)}=2.901, p=0.004 \)) (See Supplementary Information, Figure S2 and S3). Smaller males and more heterozygous males both had significantly greater reproductive success, but there was no effect of the OSR or habitat complexity on sexual selection on any of the three focal male traits (Table 1, Figure S4 and S5). When we examined net selection across all four treatments, the selection gradients were statistically significant for all three male traits (Table 2). It is worth noting that the patterns are very
clear: in 22 out of 24 pools selection favoured more heterozygous males; in 19 out of 24 pools selection favoured smaller males; and in 20 out of 24 pools selection favoured males with a relatively long gonopodium.

Number of sires per brood

The mean number of sires per brood was greater under a male-biased than female-biased OSR (estimate ± SE = 0.420 ± 0.192, Z = 2.181, p = 0.029), but it did not depend on habitat complexity (estimate ± SE = -0.095 ± 0.195, Z = 0.488, p = 0.626), nor was there an interaction between OSR and habitat complexity (estimate ± SE = -0.125 ± 0.280 Z = 0.446, p = 0.655) (Fig. 1). This finding is unlikely to be confounded by female fecundity depending on the OSR (see supplementary results).

DISCUSSION

Spatial and temporal environmental and/or social heterogeneity have long been invoked as factors that help to maintain variation in traits that are under directional sexual selection (Levins 1968; Cornwallis & Uller 2010. We experimentally tested how two key parameters – the operational sex ratio (OSR) and habitat complexity - influence sexual selection on male mosquitofish Gambusia holbrooki. Our estimates of sexual selection on focal male traits did not differ across environments. Small males, males with a relatively larger gonopodium, and more heterozygous males had greater reproductive success in all cases. This trend was remarkably consistent in direction across pools (19-22 of the 24 pools). Our results demonstrate that altering OSR and habitat complexity has little effect on sexual selection in G. holbrooki.

The operational sex ratio
Male-biased sex ratios increase competition for mates, and it is usually assumed that this will increase variation in male mating and reproductive success (Emlen & Oring 1977; Shuster 2016). Instead we found that the mean number of sires per brood was greater with a more male-biased sex ratio even though the number of males in our different sex ratio treatments remained constant. This could indicate that males are better able to monopolise females in female-biased environments. We cannot determine from our experiment whether this effect is due to changes in male or female density, however, because we deliberately kept the overall population density constant (meaning that the density of a given sex and the sex ratio covary perfectly). All else being equal, a greater sharing of paternity might be expected to reduce selection on sexual traits because it reduces variation in male reproductive success. However, the adult sex ratio had no detectable effect on our estimated selection gradients. This reveals that even though more sires contributed to each brood this did not affect the net distribution of paternity among males.

There was no evidence that the sex ratio, as manipulated here, influenced selection on male traits in G. holbrooki. Numerous studies have shown that the adult sex ratio and the OSR can both affect mating behaviour (e.g., Bretman et al. 2012; Holveck et al. 2015). However, few studies experimentally manipulate the OSR to test whether it affects sexual selection on specific male traits (most studies simply show that the OSR affects male trait expression: see Weir et al 2011). This is a surprising oversight as the evolution of male traits depends on how they affect fitness (i.e., relative number of offspring sired): for the OSR to affect evolution it must alter the trait-fitness relationship. Of the experimental OSR studies that measure selection on male traits based on actual reproductive success, the results are mixed. In guppies (Poecilia reticulata, Head et al. 2008) and bank voles (Clethrionomys glareolus, Mills et al. 2007) there was no effect of the OSR on selection on male sexual traits. In contrast, in two spotted gobies (Gobiusculus flavescensa, Wacker et al. 2013) and rough skinned newts (Taricha granulosa, Jones et al. 2004) selection on male traits was stronger
when the OSR was more male-biased. Finally, and contrary to some expectations, selection on male
traits was weaker when the OSR was more male-biased in bank voles (*Clethrionomys glareolus*,
Klemme *et al.* 2007) and common lizards (*Lacerta vivipara*, Fitze & Le Galliard 2011).

The conflicting results in previous studies of how the OSR affects sexual selection might be partly
due to confounding effects of other ecological parameters, especially those that determine how
often individuals interact (e.g., habitat complexity, or factors that influence population density (see:
Kokko & Rankin 2006)). We therefore tested for an interaction between the OSR and habitat
complexity that might affect sexual selection in *G. holbrooki*. There was no evidence that habitat
complexity, at least as manipulated in our study, affected sexual selection on the three measured
traits, either by moderating the effect of the OSR or by having a consistent effect irrespective of the
OSR. Another environmental parameter, that we did not manipulate in our experiment, that may be
important in mediating the effects of OSR on sexual selection is population density (but see Head *et
al* 2008; Wacker *et al* 2013). In our experiment fish densities were at the high end of what fish might
experience in the wild, thus it is possible that consistent sexual selection for small males with a long
gonopodium (two traits that might be expected to be favoured when there is more potential for
sexual coercion) is the result of high densities across all of our treatments. This hypothesis remains
to be tested.

**Habitat complexity and sexual selection**

Habitat complexity, as manipulated here, did not influence selection on male traits, even though
comparable variation in habitat complexity sometimes alters sexual behaviour in other poeciliid
fishes (e.g., Hibler & Houde 2006). Furthermore, habitat variation is important in shaping sexual
traits in many species, which is why ecological factors are often implicated in population variation in
sexual traits (e.g. Cornwallis & Uller 2010) and even in speciation (e.g. Maan & Seehausen 2011). For example, habitat differences in gravel size promote divergence in male colouration in guppies (Poecilia reticulata, Endler 1980). Similarly, variation in habitat complexity affects selection on male advertisement calls in cricket frogs (Acris crepitans, Ryan et al. 1990). The difference between these studies and ours might reflect the relationship between the traits being measured and the environment they are being measured in. In the studies mentioned above colouration and vocalisations are sexual signals whose transmission and detection is dependent upon the habitats they are being measured in. In contrast, in our study body size, residual gonopodium size and heterozygosity are not. In our study, a more likely mechanism by which habitat complexity would alter selection is via effects on mate encounter rates, which may affect how females assess males, or might shift the balance between different modes of sexual selection (e.g. mate choice vs coercion).

Traits under sexual selection in G. holbrooki

We detected strong directional selection on males for smaller body size, larger residual gonopodium length and higher heterozygosity. We consider each trait in turn.

Male body size

Smaller male G. holbrooki had greater reproductive success. This has long been assumed for Gambusia spp based on indirect behavioural evidence for insemination success (e.g., Pilastro et al. 1997), but until recently paternity data has been lacking. In a small paternity study, Deaton (2008) found a large male advantage based on 27 trials where a small and a large male competed freely for access to a female within small aquaria. In a much larger study of 180 males, Booksmythe et al. (2016) found no effect of male body size on paternity in 30 pools, in each of which 6 males freely competed for 8 females. As in many species of poeciliid fishes, male mosquitofish vary substantially
in size (range in this experiment: 19 – 32mm). Understanding how this type of variation persists
despite strong directional selection is a major challenge (Barton & Turelli 1989). This is true even if
there is no additive genetic variation in male size (but see Stearns 1983). Potential explanations
include context-dependent selection (Cornwallis & Uller 2010), genic capture (Tomkins et al. 2004),
fitness trade-offs between traits (Blows et al. 2003), and trade-offs between the effect of a given
trait under different modes of selection (Devigili et al. 2015; Johnston et al. 2013). Our current
results, while they clearly need to be replicated in a wider range of habitats, suggest that large size
variation in male *G. holbrooki* is not due to variation in habitat complexity, nor to ecological factors
that affect the OSR.

Male genital size

Male *G. holbrooki* with a relatively long gonopodium for their body size had higher reproductive
success. Similar positive directional selection on gonopodium length has been shown previously in
guppies, *P. reticulata* (Devigili et al. 2015; Evans et al. 2011), and in a second study on *G. holbrooki*
from our laboratory (Vega Trejo et al. 2016). This could be due to female choice for males with a
long gonopodium (Kahn et al. 2010; Langerhans et al. 2005), or a greater ability to inseminate
females coercively (Evans et al. 2011). Interestingly, Booksmythe et al. (2016) recently showed no
increase in reproductive success for males from lines artificially selected for greater residual
gonopodium length. This suggests that although residual gonopodium length is heritable (i.e., it
evolved under artificial selection) and there is also directional *selection for* males with a relatively
long gonopodium for their body size, this might not be due to *selection of* a relatively long
gonopodium (see Morrissey (2014) for a discussion on the distinction between ‘selection for’ and
‘selection of’ a trait). That is, an unmeasured factor might cause both greater residual gonopodium
length and higher reproductive success. A likely candidate is body condition (see also Kruuk et al.
2002). Alternatively, the evolution of gonopodium length might be constrained by genetic
covariance with other traits that affect fitness, so that the response to artificial selection is largely orthogonal to the unmanipulated direction of selection in multivariate trait space (Blows et al. 2004; Hine et al. 2004). This is a reminder of the easily overlooked fact that estimates of selection gradients can only truly estimate direct selection on traits if all relevant covarying traits are measured (Lande & Arnold 1983).

**Heterozygosity**

Male *G. holbrooki* with higher heterozygosity had greater reproductive success. This is a finding that we have recently replicated in a second paternity analysis study after using a formal breeding design to systematically manipulate male heterozygosity (Vega Trejo et al. 2016). Studies of heterozygosity fitness correlations (HFCs) show that homozygosity negatively affects fitness-enhancing traits (reviews: Chapman et al. 2009; Coltman & Slate 2003; Szulkin et al. 2010). There are, however, relatively few HFC studies that link heterozygosity to male reproductive success under sexual selection (i.e. control for male mortality). Of these, several studies show that lower heterozygosity decreases male reproductive success (e.g., water dragons, *Intellagama lesueurii* (Frere et al. 2015); Black rhinoceros, *Diceros bicornis michaeli* (Cain et al. 2014); zebra finches, *Taeniopygia guttata* (Forstmeier et al. 2012); house mice, *Mus musculus musculus* (Thoss et al. 2011); blue tits, *Cyanistes caeruleus* (Olano-Marin et al. 2011)), although this is not always true (e.g., Great tits, *Parus major* (Chapman & Sheldon 2011)).

We observed a strong positive relationship between heterozygosity and male reproductive success ($r=0.267$) compared to a mean value for HFCs of $r=0.05$ (meta-analysis: Chapman et al. 2009). There are several reasons why we might see a strong relationship in *G. holbrooki*. First, we had a better estimate of genome wide heterozygosity (Balloux et al. 2004). Although microsatellite markers are
generally 4-10 times more variable than SNPs (Mariette et al. 2002; Morin et al. 2004), the 3171 SNP
markers we used is equivalent to using over 300 microsatellite markers. To date, most HFC studies
use <20 microsatellite markers (Chapman et al. 2009). Second, traits that are more closely related to
actual fitness are more likely to suffer inbreeding depression (Kristensen et al. 2010). The studies in
Chapman et al. (2009) mainly report HFC correlations for morphological, physiological and life-
history traits; very few studies provide direct fitness estimates such as reproductive success. As such,
the average HFC in Chapman et al. (2009) is likely to be an underestimate of the true link with fitness
(Chapman & Sheldon 2011). Third, the HFC is likely to depend on a population’s demographic
history, with relationships being weaker in highly outbred or inbred populations where variation in
heterozygosity is lower. In our study population there is a relatively low mean heterozygosity, but it
is within the range of natural populations (Vera et al. 2016). Crucially, however, there is substantial
variation in heterozygosity (17-36%) which makes the population particularly conducive to
quantifying the HFC. Interestingly, another recent study that used a large number of SNPs to
estimate inbreeding also found strong inbreeding depression when looking at fitness traits in red
deer (Cervus elaphus) (including lifetime reproductive success) despite relatively low variation in
heterozygosity (Huisman et al. 2016).

There are many mechanisms whereby lower heterozygosity could reduce fitness due to negative
effects on the ability to acquire mates and fertilizations. Previous studies have shown that inbred
(i.e., more homozygous) males can be: less attractive (review: Pusey & Wolf 1996); produce less
competitive ejaculates (e.g., Michalczyk et al. 2010); and have poorer locomotion performance (e.g.,
Manenti et al. 2015) or poorer cognition (Fareed & Afzal 2014) that could reduce the ability to locate
females. Regardless of the proximate mechanism, however, the reported strong effect of
heterozygosity on paternity is likely to have wider implications. For example, by changing the
effective population size and identity of successful males it might affect the persistence and recovery
of small populations (Keller & Waller 2002) (i.e., fewer sires reduces population genetic diversity, but their being more genetically diverse increases population genetic diversity).

Conclusion

Sexual selection in the mosquitofish was consistent across populations that differed in two parameters that are expected to affect mate encounter rates. Persistent ecological differences between habitats is clearly important for generating divergence in sexual traits between species (reviewed in: Maan & Seehausen 2011). However, the extent to which temporal and spatial habitat variation generates variation among extant populations of a single species is less clear (Cornwallis & Uller 2010). Experimental studies like ours that quantify sexual selection in different environments and over different time/spatial scales are needed to understand better how ecological variation affects the strength and form of sexual selection and, by extension, how organisms might respond to changing environments.

DATA ACCESSIBILITY

Data will be made available through DRYAD upon acceptance.

ACKNOWLEDGEMENTS

We thank the ANU animal services team and Regina Vega-Trejo for help with fish maintenance. This research was approved by the Australian National University Animal Ethics Committee (Approval # A2011/64). This work was supported by an Australian Research Council Grant awarded to MDJ (DP160100285)

DATA ACCESSIBILITY

Data are available on DRYAD doi:10.5061/dryad.93351
AUTHOR CONTRIBUTIONS

MLH, ATK, JSK and MDJ designed the experiment, MLH & ATK collected the data, MLH, JMH, JSK and MDJ analysed the data, MLH wrote the first draft of the manuscript and all authors helped with revisions.

REFERENCES


22


Chapman JR, Sheldon BC (2011) Heterozygosity is unrelated to adult fitness measures in a large, noninbred population of great tits (Parus major). Journal of Evolutionary Biology, 24, 1715-1726.


Mautz BS (2011) Sexual selection in animals, Australian National University, Canberra.


Evolution 44:1869-1872.


Table 1. The effects of OSR and habitat complexity on the relationship between male traits and the number of offspring sired. Effect sizes (r) and their 95% confidence intervals were calculated from P following the formula in Lipsey & Wilson (2001).

<table>
<thead>
<tr>
<th>Trait</th>
<th>Term</th>
<th>Estimate</th>
<th>SE</th>
<th>z</th>
<th>P</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Offspring</td>
<td>Intercept</td>
<td>16.139</td>
<td>12.603</td>
<td>1.370</td>
<td>0.181</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OSR (m)</td>
<td>-2.790</td>
<td>15.927</td>
<td>-0.191</td>
<td>0.861</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>Habitat (s)</td>
<td>0.301</td>
<td>16.600</td>
<td>0.016</td>
<td>0.986</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>Log Standard length (SL)</td>
<td>-18.414</td>
<td>8.034</td>
<td>-2.335</td>
<td>0.022</td>
<td>0.465</td>
</tr>
<tr>
<td></td>
<td>Residual gonopodium length (Gono)</td>
<td>0.479</td>
<td>0.331</td>
<td>1.420</td>
<td>0.148</td>
<td>0.305</td>
</tr>
<tr>
<td></td>
<td>Percent heterozygosity (Het)</td>
<td>31.336</td>
<td>10.341</td>
<td>3.072</td>
<td><strong>0.002</strong></td>
<td>0.599</td>
</tr>
<tr>
<td></td>
<td>OSR(m)*Habitat (s)</td>
<td>2.131</td>
<td>21.874</td>
<td>0.092</td>
<td>0.922</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>OSR(m)*SL</td>
<td>1.091</td>
<td>10.576</td>
<td>0.116</td>
<td>0.918</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>OSR(m)*Gono</td>
<td>-0.395</td>
<td>0.507</td>
<td>-0.704</td>
<td>0.436</td>
<td>0.167</td>
</tr>
<tr>
<td></td>
<td>OSR(m)*Het</td>
<td>4.456</td>
<td>14.971</td>
<td>0.308</td>
<td>0.766</td>
<td>0.064</td>
</tr>
<tr>
<td></td>
<td>Habitat(s)*SL</td>
<td>2.007</td>
<td>11.078</td>
<td>0.190</td>
<td>0.856</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>Habitat(s)*Gono</td>
<td>0.655</td>
<td>0.503</td>
<td>1.346</td>
<td>0.193</td>
<td>0.275</td>
</tr>
<tr>
<td></td>
<td>Habitat(s)*Het</td>
<td>-11.115</td>
<td>14.159</td>
<td>-0.811</td>
<td>0.432</td>
<td>0.168</td>
</tr>
<tr>
<td></td>
<td>OSR(m)*Habitat(s)*SL</td>
<td>2.965</td>
<td>14.596</td>
<td>0.210</td>
<td>0.839</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>OSR(m)*Habitat(s)*Gono</td>
<td>0.305</td>
<td>0.752</td>
<td>0.333</td>
<td>0.685</td>
<td>0.087</td>
</tr>
<tr>
<td></td>
<td>OSR(m)*Habitat(s)*Het</td>
<td>-21.459</td>
<td>19.683</td>
<td>-1.092</td>
<td>0.276</td>
<td>0.232</td>
</tr>
</tbody>
</table>

Bold indicates significant effects
Table 2. The vector of experiment wide standardised linear selection gradients (β) for male traits in *Gambusia holbrooki*. Relative fitness was calculated within pools and male traits were standardised across the experiment. Selection gradients were estimated using linear multiple regression. The significance of selection gradients was determined using a linear mixed model with power-transformed relative fitness as the response variable to account for non-normal distribution of the data. Pool was included in this model as a random effect to account for potential non-independence of data from the same pool.

<table>
<thead>
<tr>
<th>Trait</th>
<th>B (SE)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent heterozygosity</td>
<td>0.355 (0.103)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Log standard length</td>
<td>-0.205 (0.101)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Residual gonopodium length</td>
<td>0.348 (0.101)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Bold indicates significant effect
Figure 1. The mean number of sires contributing to each brood (±SE) within each treatment. The number of females in each treatment: female-biased/complex habitat N=30; female-biased/simple habitat N=30; male-biased/complex habitat N=19; male-biased/simple habitat N=22.
We estimated non-directional selection on each trait (heterozygosity, log male length and relative gonopodium length) using the method of Henshaw & Zemel (2016). We first calculated the distributional selection differential (DSD) for each trait, which quantifies the total strength of selection, including both directional (i.e. linear) and non-directional selection. If trait values are ordered so that $z_1 \leq z_2 \leq \cdots \leq z_n$, then the DSD is given by:

$$d = \sum_{i=1}^{n-1} (z_{i+1} - z_i) \left( \sum_{j=1}^{i} \frac{1 - w_j}{n-1} \right)$$

Note that we use a divisor of $n - 1$, rather than $n$, for consistency with the linear selection differential, which is estimated as a sample covariance. We then calculated the non-directional component of the DSD as:

$$d_N = d - |s|,$$

where $s$ is the linear selection differential (Henshaw & Zemel 2016). This is an estimate of the total strength of non-directional selection of any kind (e.g. stabilising or disruptive: Brodie et al. 1995).

We used a permutation test to check the significance of selection estimates. We generated null distributions of $d$, $s$ and $d_N$ for each trait by randomly permuting relative fitness values among the sampled individuals 10,000 times and calculating these three statistics each time. We then estimated $p$-values by calculating the proportion of null-distributed values that were at least as large (in absolute value) as the observed values.

Both directional and total selection on heterozygosity, male body length and gonopodium length were highly significant (Table S1), consistent with results in the main text (cf. Table 2). Estimates of non-directional selection were small and non-significant (Table S1).
Table S1. Total selection (measured as the distribution selection differential \( d \)), directional selection (linear selection differential \( s \)) and non-directional selection (\( d_N = d - |s| \)) for male traits in *Gambusia holbrooki*. Relative fitness was calculated within ponds and male traits were standardised across the experiment.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Statistic</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent heterozygosity</td>
<td>( d = 0.017 )</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>( s = 0.017 )</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>( d_N &lt; 0.001 )</td>
<td>1.00</td>
</tr>
<tr>
<td>Log standard length</td>
<td>( d = 0.013 )</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>( s = -0.013 )</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>( d_N &lt; 0.001 )</td>
<td>0.93</td>
</tr>
<tr>
<td>Relative gonopodium length</td>
<td>( d = 0.411 )</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>( s = 0.410 )</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>( d_N = 0.001 )</td>
<td>0.97</td>
</tr>
</tbody>
</table>

2. Methods for simulations to test for bias in selection gradients

We ran simulations to test for possible bias introduced by subsampling females in the female-biased pools (n = 5 of 20 females) in the experiment (see Mathematica code below). The simulations were designed and parameterised so that the data they generate is broadly consistent with those from the experiment (e.g. similar mean and variance in estimated male reproductive success). For each trial, we simulated mating and reproductive success in a pool of 10 males and 20 females (corresponding to the female-biased OSR in the experiment). We calculated selection gradients on two simulated male traits for both (i) the full sample of 20 females, and (ii) a random subsample of 5 females. We found negligible differences in average selection gradients between these two groups over a large number of trials (see below). We conclude that the subsampling procedure does not introduce appreciable bias.

Mating and post-copulatory success were assumed to be partly determined by male traits \( Z_M \) and \( Z_P \) respectively, which were drawn from standard normal distributions. Mating success \( M \) was generated as a random normal variable with a mean and standard deviation equal to 3, and a correlation of \( \rho \) with \( Z_M \). Higher values of \( \rho \) mean that \( Z_M \) has a stronger effect on mating success.\n
We rounded values of \( M \) to the nearest integer and then converted any negative values to zero. Sires were chosen from the mother’s mates at random, such that a male’s probability of siring any particular offspring was proportional to \( \exp(\alpha Z_P) \). Larger values of \( \alpha \) mean that \( Z_P \) has a stronger impact on male post-copulatory success.
We calculated relative reproductive success for each male over the full sample of 20 females, and over the random subsample of 5 females. We then calculated selection gradients on the two male traits, denoted as $\beta = [\beta_M, \beta_F]$ for the full samples and $\hat{\beta} = [\hat{\beta}_M, \hat{\beta}_F]$ for the subsamples. The bias due to subsampling is $\mathbb{E}(\hat{\beta} - \beta)$. We estimated this bias by running $10^6$ trials for each combination of the parameters $\rho = 0, 0.2, 0.5$ and $\alpha = 0, 0.5, 1$. This covers a range of selection intensities, including those observed in the experiment. Estimated bias was always less than $10^{-3}$, indicating that the subsampling procedure does not introduce appreciable bias.

3. Mathematica code for simulations

(*This code is written for Wolfram Mathematica, Version 11.0. The code appears in regular typeface, with comments in bold*)

```
NumRuns = 1000000;

(*Male and female sample sizes as in the experiment*)
FemaleFullSampleSize = 20;
FemaleSubsampleSize = 5;
MaleSampleSize = 10;

(*The parameters rho and alpha respectively*)
MatingTraitCorrelation = 0.5;
PaternitySlope = 0.5;

(*Start tallies that are used later*)
betaSubTally = 0;
betaDifTally = 0;
NumGoodRuns = NumRuns;

(*A matrix used to generate correlated normal random variables*)
MatingTraitCholesky = CholeskyDecomposition[{{1, MatingTraitCorrelation}, {MatingTraitCorrelation, 1}}];
```
(*This loop repeats 10^6 times*)

Do[

(*Generate the mating trait Z_M and a correlated normal variable that will be used to calculating mating success*)

{MatingTrait, MatingSuccessGenerator} = Transpose[Map[#.MatingTraitCholesky &, RandomReal[NormalDistribution[0, 1], {MaleSampleSize, 2}]]];

(*Generate male mating success*)

MaleMatingSuccess = Round[Max[0, #]] & /@ (3 + 3 MatingSuccessGenerator);

(*Generate the paternity trait Z_P*)

PaternityTrait = RandomReal[NormalDistribution[0, 1], MaleSampleSize];

(*Choose each male's mates at random from the 20 females*)

MaleMates = Table[RandomSample[Array[# &, FemaleFullSampleSize], MaleMatingSuccess[[i]]], {i, 1, MaleSampleSize}];

(*Assign each female the appropriate male mates*)

FemaleMates = ConstantArray[{}, FemaleFullSampleSize];

Do[AppendTo[FemaleMates[[MaleMates[[i, j]]]], i];
    , {i, 1, Length[MaleMates]}, {j, 1, Length[MaleMates[[i]]]}];

(*Generate female reproductive success. For females that don't mate this value is ignored*)

FemaleReproductiveSuccess = Round[Max[0, #]] & /@ RandomReal[NormalDistribution[10, 5], FemaleFullSampleSize];
(*Generate an empty parental table, i.e. the number offspring produced by each female-male pair, to be filled in below*)

ParentalTable = ConstantArray[0, {FemaleFullSampleSize, MaleSampleSize}];

(*This loop runs for each female*)
Do[
  If[FemaleMates[[fem]] != {},
    (*Determine the father for each of the female's offspring.
      Each mate's siring probability is proportional to his value of exp(alpha*Z_P)*)
    offspring = RandomChoice[Exp[PaternitySlope*PaternityTrait[[FemaleMates[[fem]]]]] ->
      FemaleMates[[fem]], FemaleReproductiveSuccess[[fem]]];

    (*This loops runs for each of the female's mates*)
    Do[
      (*Add that mate's offspring to the parental table*)
      ParentalTable[[fem, FemaleMates[[fem, malno]]]] = 
        Count[offspring, FemaleMates[[fem, malno]]];

      , {malno, 1, Length[FemaleMates[[fem]]]}];

    , {fem, 1, FemaleFullSampleSize}];

(*End both loops*)

(*Calculate male reproductive success for the full sample from the parental table*)
RFullSample = Total[ParentalTable];

(*Create a subsample of the parental table containing only 5 females*)
ParentalTableSubsample = RandomSample[ParentalTable, FemaleSubsampleSize];

(*Calculate male reproductive success for the subsample*)
RSubsample = Total[ParentalTableSubsample];

(*Standardise both males traits*)
{MatingTrait, PaternityTrait} = ((# - Mean[#])/StandardDeviation[#]) & /@ {MatingTrait, PaternityTrait};

(*Calculate selection gradients on the male traits for both the full sample (betaFull) and the subsample (betaSub) and then take their difference (betaDif)*)
Quiet[betaFull = {(#[1, 0] - #[0, 0]), (#[0, 1] - #[0, 0])} &@ LinearModelFit[Transpose[{MatingTrait, PaternityTrait, FullSample/Mean[RFullSample]}], {z1, z2}, {z1, z2}];

betaSub = {(#[1, 0] - #[0, 0]), (#[0, 1] - #[0, 0])} &@ LinearModelFit[Transpose[{MatingTrait, PaternityTrait, RSubsample/Mean[RSubsample]}], {z1, z2}, {z1, z2}];

betaDif = betaSub - betaFull;

(*Add betaSub and betaDif to their respective tallies, assuming nothing weird went wrong (e.g. zero variance in one variable). In case of weird things, remove the trial from the analysis. This affects only a very small proportion of trials*)
If[Element[betaDif[[1]], Reals] && Element[betaDif[[2]], Reals],
  betaSubTally = betaSubTally + betaSub;
  betaDifTally = betaDifTally + betaDif;,
  NumGoodRuns = NumGoodRuns - 1];

, {gen, 1, NumRuns}]
4. Paternity Analysis

DNA samples were sent to the commercial genotyping service Diversity Arrays. This company has developed a widely used technique called DArTseq™. DArTseq™ represents a combination of DArT complexity reduction methods and next generation sequencing platforms (Kilian et al, 2012; Courtois et al, 2013; Cruz et al, 2013; Raman et al, 2014). It is a new implementation of sequencing complexity reduced representations (Altshuler et al, 2000) and more recent applications of this concept on next generation sequencing platforms (Baird et al, 2008; Elshire et al, 2011). The technology is optimized by selecting the most appropriate complexity reduction method based on both the size of the representation and the genome fraction selected for assays. Four methods of complexity reduction were tested in Gambusia and the PstI-HpaII method was selected. DNA samples were processed in digestion/ligation reactions principally as per Kilian et al (2012), but replacing a single PstI-compatible adaptor with two different adaptors corresponding to two different Restriction Enzyme (RE) overhangs. The PstI-compatible adapter was designed to include Illumina flowcell attachment sequence, sequencing primer sequence and “staggered”, varying length barcode region, similar to the sequence reported by Elshire et al (2011). The reverse adapter contained flowcell attachment region and HpaII-compatible overhang sequence. Only “mixed fragments” (PstI-HpalII) were effectively amplified in 30 rounds of PCR using the following reaction conditions: 1. 94°C for 1 min; 2. 30 cycles of 94°C for 20 sec, 58°C for 30 sec, 72°C for 45 sec; 3. 72°C for 7 min. After PCR equimolar amounts of amplification products from each sample of the 96-well microtiter plate were bulked and applied to c-Bot (Illumina) bridge PCR followed by sequencing on Illumina Hiseq2500. The sequencing (single read) was run for 77 cycles.

Sequences generated from each lane were processed using proprietary DArT analytical pipelines. In the primary pipeline the fastq files were first processed to filter out poor quality sequences, applying more stringent selection criteria to the barcode region than the rest of the sequence. In that way the assignments of the sequences to specific samples carried in the “barcode split” step are very reliable. Approximately 2 500 000 (±7%) sequences per barcode/sample are used in marker calling in routine DArTseq assay, but we applied a more cost effective version using 1 300 000 per sample). Finally, identical sequences were collapsed into “fastqcall files” used in the secondary pipeline for DArT PL’s proprietary SNP and SilicoDArT (presence/absence of restriction fragments in representation) calling algorithms (DArTsoft14). This clean-up process resulted in a comprehensive data set of approximately 3171 SNPs with an average call rate of 97.7% and a reproducibility rate of 99.3%.
5. Reproductive success of females

To test whether our treatments influenced female reproductive output we ran GLMMs. The number of broods (0, 1 or 2) was analysed using an ordinal logistic regression in the package “ordinal” using the command clmm to allow for random effects. For those females that bred we also analysed the number of offspring in her first brood, gestation time (days between leaving the pool and giving birth), and the total number of offspring a female produced using models with Poisson error that included an individual level random effect when data were over-dispersed (Harrison 2014). In all models OSR, habitat and female standard length (centred to a mean of 0; (Gelman 2008)) and their interactions were specified as fixed effects. Pool identity was treated as a random effect.

Neither gestation time, the number of offspring in the first brood nor the total number of offspring a female produced were influenced by the adult sex ratio, habitat complexity or the interaction between them (Table S2). The number of broods per female was, however, influenced by an interaction between female length and the OSR. With a female-biased adult sex ratio smaller females produced more broods than larger females, whereas with a male-biased sex ratio there was a weak relationship in the opposite direction (Fig. S6). There was no effect of habitat complexity.

Trait means for each treatment are shown in Table S3.
**Table S2.** The effects of OSR and habitat complexity on female reproductive output.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Term</th>
<th>Estimate</th>
<th>SE</th>
<th>z</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of broods</td>
<td>OSR (m)</td>
<td>-0.614</td>
<td>0.821</td>
<td>-0.748</td>
<td>0.454</td>
</tr>
<tr>
<td></td>
<td>Habitat (s)</td>
<td>-0.444</td>
<td>0.748</td>
<td>-0.594</td>
<td>0.553</td>
</tr>
<tr>
<td></td>
<td>Standard length (SL)</td>
<td>-1.226</td>
<td>0.532</td>
<td>-2.302</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>OSR(m)*Habitat (s)</td>
<td>0.469</td>
<td>1.157</td>
<td>0.405</td>
<td>0.685</td>
</tr>
<tr>
<td></td>
<td>OSR(m)*SL</td>
<td>2.323</td>
<td>1.004</td>
<td>2.314</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>Habitat(s)*SL</td>
<td>0.374</td>
<td>0.733</td>
<td>0.511</td>
<td>0.670</td>
</tr>
<tr>
<td></td>
<td>OSR(m)*Habitat(s)*SL</td>
<td>-0.953</td>
<td>1.352</td>
<td>-0.705</td>
<td>0.481</td>
</tr>
<tr>
<td>Gestation time</td>
<td>Intercept</td>
<td>3.766</td>
<td>0.037</td>
<td>102.11</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>OSR (m)</td>
<td>-0.096</td>
<td>0.082</td>
<td>-1.17</td>
<td>0.243</td>
</tr>
<tr>
<td></td>
<td>Habitat (s)</td>
<td>-0.029</td>
<td>0.052</td>
<td>-0.57</td>
<td>0.569</td>
</tr>
<tr>
<td></td>
<td>Standard length (SL)</td>
<td>0.034</td>
<td>0.077</td>
<td>0.44</td>
<td>0.661</td>
</tr>
<tr>
<td></td>
<td>OSR(m)*Habitat (s)</td>
<td>-0.028</td>
<td>0.114</td>
<td>-0.25</td>
<td>0.806</td>
</tr>
<tr>
<td></td>
<td>OSR(m)*SL</td>
<td>0.107</td>
<td>0.168</td>
<td>0.64</td>
<td>0.525</td>
</tr>
<tr>
<td></td>
<td>Habitat(s)*SL</td>
<td>-0.029</td>
<td>0.106</td>
<td>-0.28</td>
<td>0.782</td>
</tr>
<tr>
<td></td>
<td>OSR(m)*Habitat(s)*SL</td>
<td>-0.149</td>
<td>0.222</td>
<td>-0.67</td>
<td>0.501</td>
</tr>
<tr>
<td>Offspring brood 1</td>
<td>Intercept</td>
<td>1.692</td>
<td>0.089</td>
<td>19.108</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>OSR (m)</td>
<td>0.282</td>
<td>0.197</td>
<td>1.437</td>
<td>0.151</td>
</tr>
<tr>
<td></td>
<td>Habitat (s)</td>
<td>0.181</td>
<td>0.122</td>
<td>1.479</td>
<td>0.139</td>
</tr>
<tr>
<td></td>
<td>Standard length (SL)</td>
<td>0.050</td>
<td>0.181</td>
<td>0.278</td>
<td>0.781</td>
</tr>
<tr>
<td></td>
<td>OSR(m)*Habitat (s)</td>
<td>-0.393</td>
<td>0.275</td>
<td>-1.428</td>
<td>0.153</td>
</tr>
<tr>
<td></td>
<td>OSR(m)*SL</td>
<td>-0.198</td>
<td>0.410</td>
<td>-0.483</td>
<td>0.629</td>
</tr>
<tr>
<td></td>
<td>Habitat(s)*SL</td>
<td>0.061</td>
<td>0.277</td>
<td>0.219</td>
<td>0.827</td>
</tr>
<tr>
<td></td>
<td>OSR(m)*Habitat(s)*SL</td>
<td>0.263</td>
<td>0.548</td>
<td>0.480</td>
<td>0.631</td>
</tr>
<tr>
<td>Total offspring</td>
<td>Intercept</td>
<td>1.887</td>
<td>0.128</td>
<td>14.769</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>OSR (m)</td>
<td>0.255</td>
<td>0.236</td>
<td>1.079</td>
<td>0.281</td>
</tr>
<tr>
<td></td>
<td>Habitat (s)</td>
<td>0.034</td>
<td>0.180</td>
<td>0.192</td>
<td>0.848</td>
</tr>
<tr>
<td></td>
<td>Standard length (SL)</td>
<td>0.095</td>
<td>0.203</td>
<td>0.469</td>
<td>0.639</td>
</tr>
<tr>
<td></td>
<td>OSR(m)*Habitat (s)</td>
<td>-0.369</td>
<td>0.332</td>
<td>-1.111</td>
<td>0.266</td>
</tr>
<tr>
<td></td>
<td>OSR(m)*SL</td>
<td>-0.277</td>
<td>0.422</td>
<td>-0.656</td>
<td>0.512</td>
</tr>
<tr>
<td></td>
<td>Habitat(s)*SL</td>
<td>0.096</td>
<td>0.295</td>
<td>0.325</td>
<td>0.746</td>
</tr>
<tr>
<td></td>
<td>OSR(m)*Habitat(s)*SL</td>
<td>0.239</td>
<td>0.562</td>
<td>0.425</td>
<td>0.671</td>
</tr>
</tbody>
</table>

**Table S3.** Female reproductive output in each treatment, mean (SE)

<table>
<thead>
<tr>
<th>Trait</th>
<th>Female-biased</th>
<th>Male-biased</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Complex</td>
<td>Simple</td>
</tr>
<tr>
<td>Number of broods</td>
<td>0.94 (0.065)</td>
<td>0.83 (0.060)</td>
</tr>
<tr>
<td></td>
<td>N=120</td>
<td>N=120</td>
</tr>
<tr>
<td>Gestation time (days)</td>
<td>44.77 (1.786)</td>
<td>43.65 (1.810)</td>
</tr>
<tr>
<td></td>
<td>N=86</td>
<td>N=83</td>
</tr>
<tr>
<td>Offspring brood 1</td>
<td>6.52 (0.493)</td>
<td>7.77 (0.574)</td>
</tr>
<tr>
<td></td>
<td>N=86</td>
<td>N=83</td>
</tr>
<tr>
<td>Total offspring</td>
<td>8.26 (0.644)</td>
<td>8.64 (0.662)</td>
</tr>
<tr>
<td></td>
<td>N=86</td>
<td>N=83</td>
</tr>
</tbody>
</table>

Bold indicates significant effect
3. Supplementary figures

Figure S1. Schematic of experimental ponds A) female-biased simple habitat, B) female-biased complex habitat, C) male-biased simple (left) and complex (right). Ponds were 1m in diameter and water depth was 15cm. Females are red and males are blue.
Figure S2. The relationship between mean heterozygosity and log male standard length.

Figure S3. The relationship between mean heterozygosity and male relative gonopodium length. The outlier in this figure is not due to measurement error. Its exclusion does not alter the conclusions of our analyses.
Figure S4. The relationship between male mean heterozygosity and the number of offspring he sired. Male-biased simple habitat (open circles, solid line), Male-biased complex habitat (open triangles dash-dot line), Female-biased simple habitat (closed circles, dot line), female-biased complex habitat (closed triangles, dash line).

Figure S5. The relationship between male log standard length and the number of offspring he sired. Male-biased simple habitat (open circles, solid line), Male-biased complex habitat (open triangles dash-dot line), Female-biased simple habitat (closed circles, dot line), female-biased complex habitat (closed triangles, dash line).
Figure S6. The relationship between number of broods and female size for male biased (open circles) and female biased (black diamonds) treatments. Number of female that had no broods - female biased = 71, male biased = 19; number of females that had one brood – female biased = 125, male biased = 34; number of females that had 2 broods – female biased = 44, male biased = 7.