SUPPLEMENTARY MATERIALS

Table of Contents

1. Testing for non-directional selection

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.

Trait	Statistic	p
Percent heterozygosity	$d = 0.017$	**<0.001**
	$s = 0.017$	**<0.001**
	$d_N < 0.001$	1.00
Log standard length	$d = 0.013$	**0.01**
	$s = -0.013$	**0.01**
	$d_N < 0.001$	0.93
Relative gonopodium length	$d = 0.411$	**<0.001**
	$s = 0.410$	**<0.001**
	$d_N = 0.001$	0.97

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 ρ with Z_M. Higher values of ρ mean that Z_M has a stronger effect on mating success. We rounded values of M to the nearest integer and then converted any negative values to zero. Mates for each male were chosen from among the 20 females at random.

For every female that mated at least once, we generated a normally distributed number of offspring with a mean of 10 and a standard deviation of 5. We rounded offspring numbers 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 α 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 $\beta = [\beta_M, \beta_P]$ for the full samples and $\hat{\beta} = [\hat{\beta}_M, \hat{\beta}_P]$ 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 $a = 0, 0.5, 1$. 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 filed 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}]

(*End main loop*)

(*Print the average selection gradients in the subsample*)

betaSubTally/NumGoodRuns

(*Print the bias in the selection gradients due to subsampling*)

betaDifTally/NumGoodRuns

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-HpaII) 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.

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

Bold indicates significant effect

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

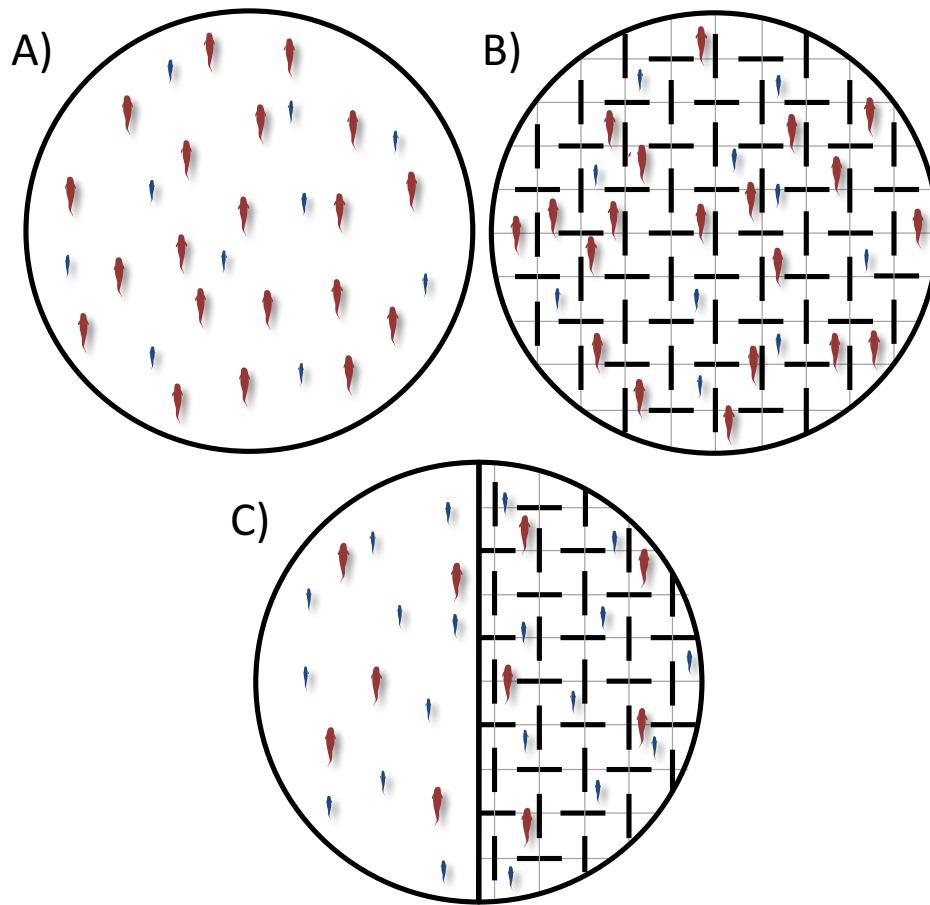|	Female-biased		Male-biased	
Trait	Complex	Simple	Complex	Simple
Number of broods	0.94 (0.065)	0.83 (0.060)	0.80 (0.130)	0.80 (0.101)
	N=120	N=120	N=30	N=30
Gestation time (days)	44.77 (1.786)	43.65 (1.810)	41.26 (3.334)	37.45 (1.389)
	N=86	N=83	N=19	N=22
Offspring brood 1	6.52 (0.493)	7.77 (0.574)	8.63 (1.501)	7.64 (1.411)
	N=86	N=83	N=19	N=22
Total offspring	8.26 (0.644)	8.64 (0.662)	10.79 (1.877)	8.09 (1.446)
	N=86	N=83	N=19	N=22

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.
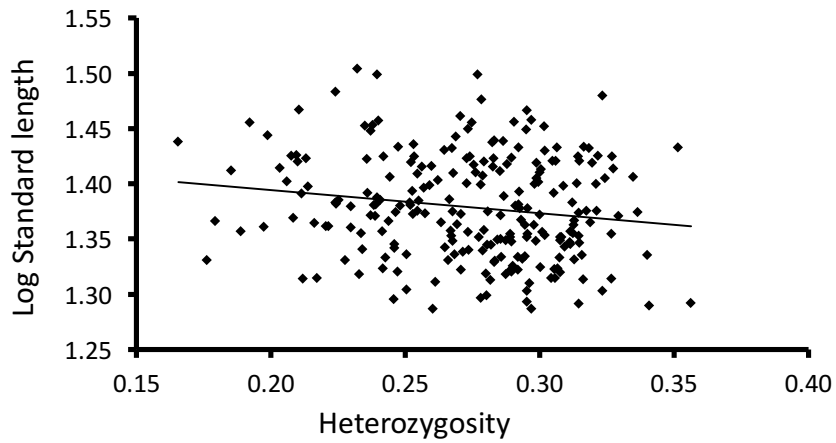
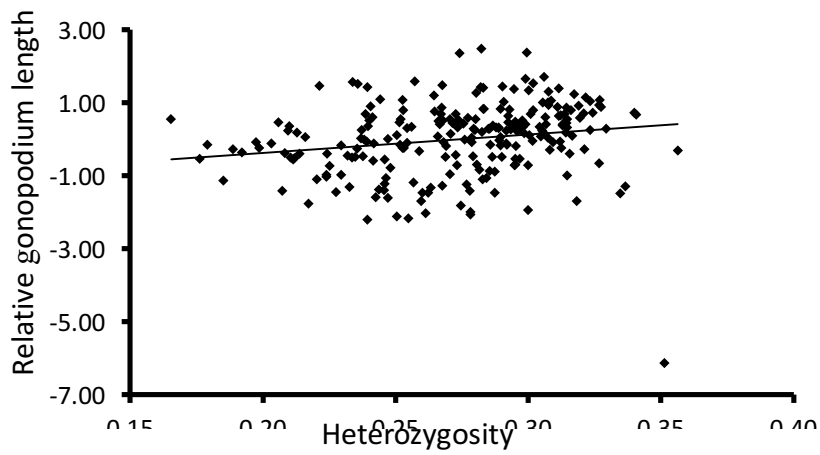**Figure S2.** The relationship between mean heterozygosity and log male standard length

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.
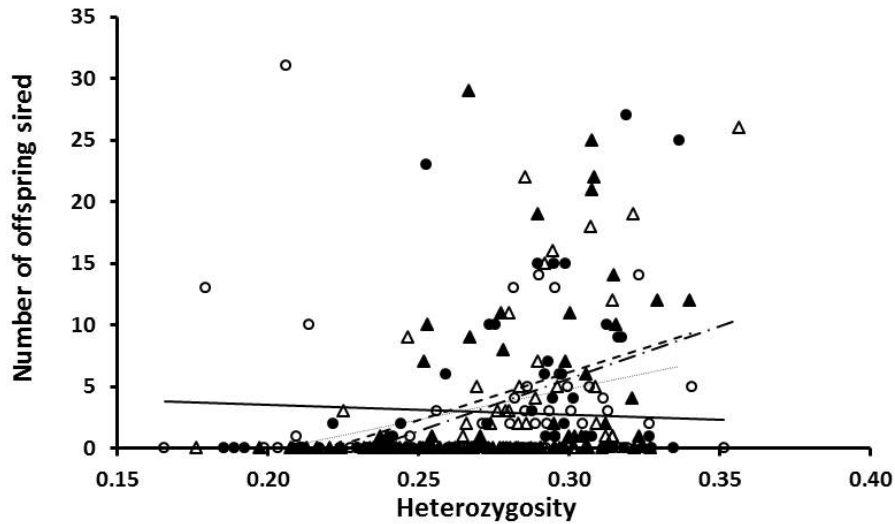
Figure S4. The relationship between male mean heterozygosity and the number of offspring he sired. Male-biased simple habitat (open circles, solid line), Male-biased complex habitat (open triangles dash-dot line), Female-biased simple habitat (closed circles, dot line), female-biased complex habitat (closed triangles, dash line).
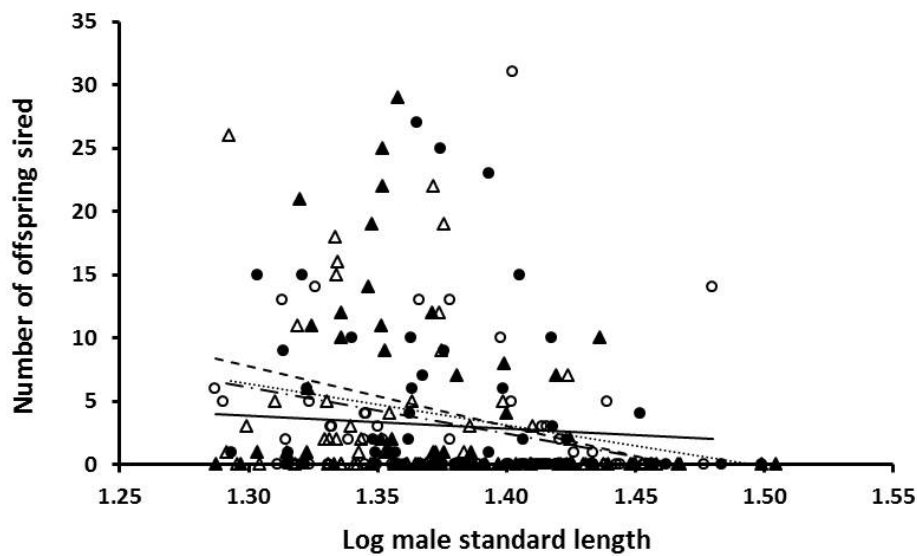

Figure S5. The relationship between male log standard length and the number of offspring he sired. Male-biased simple habitat (open circles, solid line), Male-biased complex habitat (open triangles dash-dot line), Female-biased simple habitat (closed circles, dot line), female-biased complex habitat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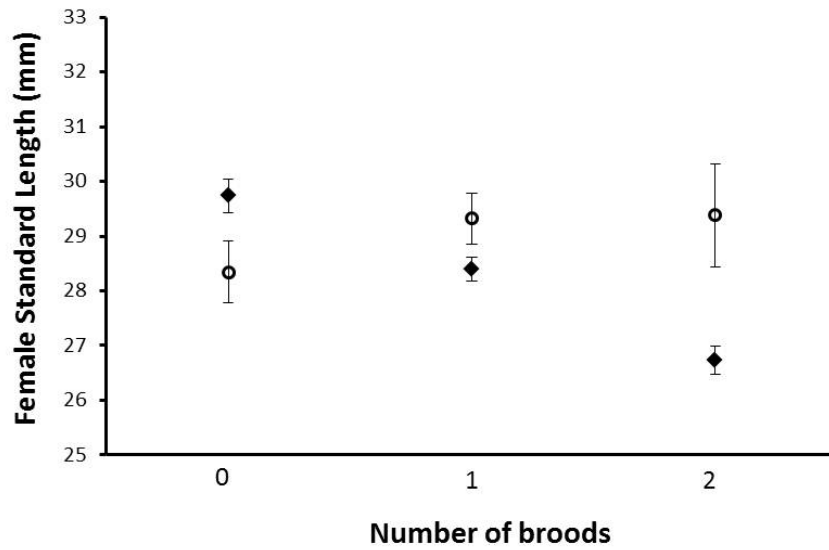
Figure S6. The relationship between number of broods and female size for male biased (open circles) and female biased (black diamonds) treatments. Number of female that had no broods - female biased = 71, male biased = 19; number of females that had one brood – female biased = 125, male biased = 34;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