Sperm competition shapes gene expression and sequence evolution in the ocellated wrasse

Running title: Sperm competition shapes gene evolution

Rebecca Dean[1,2†], Alison E. Wright[1], Susan E. Marsh-Rollo[3,4], Bridget M. Nugent[5,6], Suzanne H. Alonzo[3,5]*, Judith E. Mank[1]*

[1]Department of Genetics, Evolution and Environment, University College London

[2]School of Biological Sciences, Monash University, Clayton, Victoria, Australia

[3]Department of Ecology and Evolutionary Biology, University of California Santa Cruz

[4]Department of Psychology, McMaster University

[5]Department of Ecology and Evolutionary Biology, Yale University

[6]Department of Biomedical Sciences, University of Pennsylvania

*Co- senior authors

[†]Correspondence: r.dean@ucl.ac.uk

Keywords: masculinization, alternative reproductive tactics, morphs, positive selection, expression variance, sex-biased genes

ABSTRACT

Gene expression differences between males and females often underlie sexually dimorphic phenotypes and the expression levels of genes that are differentially expressed between the sexes is thought to respond to sexual selection. Most studies on the transcriptomic response to sexual selection treat sexual selection as a single force, but post-mating sexual selection in particular is expected to specifically target gonadal tissue. The three male morphs of the ocellated wrasse (*Symphodus ocellatus*) make it possible to test the role of post-mating sexual selection in shaping the gonadal transcriptome. Nesting males hold territories and have the highest reproductive success, yet we detected feminisation of their gonadal gene expression compared to satellite males. Satellite males are less brightly coloured and experience more intense sperm competition than nesting males. In line with post-mating sexual selection affecting gonadal gene expression, we detected a more masculinised expression profile in satellites. Sneakers are the lowest quality males and showed both de-masculinisation and de-feminisation of gene expression. We also detected higher rates of gene sequence evolution of male-biased genes compared to unbiased genes, which could at least in part be explained by positive selection. Together, these results reveal the potential for post-mating sexual selection to drive higher rates of gene sequence evolution and shape the gonadal transcriptome profile.

INTRODUCTION

Males and females within a species share the majority of the genome. Even in species with sex chromosomes, the sex-limited Y or W chromosome (in XY or ZW sex determination systems) is typically small and contains few genes (Skaletsky, et al. 2003; Koerich, et al. 2008). Therefore, many sexually dimorphic traits are the product of differences in regulation of loci present in both sexes (Dean and Mank 2016). Accordingly, just as many phenotypic traits differ substantially between the sexes, many genes show expression differences between females and males (Parsch and Ellegren 2013). These sex-biased genes are often viewed as the link between mating system, sexual selection and sexual dimorphism (Mank et al 2013).

Sexually dimorphic gene expression has recently been shown to respond to sexual selection across populations (Moghadam, et al. 2012; Hollis, et al. 2014; Immonen, et al. 2014) and among species (Harrison, et al. 2015). Additionally, intra-sexual comparisons within species have shown that transcriptional dimorphism scales with phenotypic dimorphism among individuals in both invertebrates (Snell-Rood, et al. 2011; Bailey, et al. 2013; Stuglik, et al. 2014) and vertebrates (Small, et al. 2009; Pointer, et al. 2013; Schunter, et al. 2014; Sharma, et al. 2014). These studies suggest that comparisons of transcriptomes both within and between the sexes can be useful for understanding the transcriptional architecture of sexual dimorphism, and the loci responding to sexual selection.

This previous work largely treats sexual selection as a single evolutionary force, focusing mainly on divergence between males and females. Pre-mating and post-mating sexual selection could, however, be expected to act in different ways and on different genomic targets. Pre-mating selection might be expected to largely target gene expression in somatic

63 tissues involved in competing for, attracting, or securing mates (Emlen, et al. 2012; Khila, et
64 al. 2012). In contrast, post-mating sexual selection (also referred to as post-copulatory
65 sexual selection in species with internal fertilisation), a major component of sexual selection
66 in polyandrous species, has the potential to shape the transcriptional profile of the testes
67 (Mank, et al. 2013; Harrison, et al. 2015). Focusing on these effects makes it possible to ask
68 how specific components of sexual selection have shaped variation within and between the
69 sexes, linking genomic changes to the phenotypic response.

70 Isolating the effects of post-mating sexual selection on gene expression requires analyzing
71 the gonad transcriptome separately from the soma. In addition to comparing the gonad
72 transcriptome between males and females, comparisons between multiple reproductive
73 males experiencing differing levels of post-mating sexual selection is also required. As well
74 as identifying the effects and targets of post-mating sexual selection, comparing gonadal
75 transcriptomes can also improve understanding of relative transcriptional investment and
76 potential costs associated with pre- and post-mating sexual selection. For example,
77 comparing the transcriptional profiles of the testes in males from the same species with
78 differing investment in somatic pre-mating sexually selected traits may reveal the
79 transcriptional tradeoffs males make in pre-mating versus post-mating sexually selected
80 traits. This shift in transcriptional investment in response to sperm competition could be
81 manifested in the testes, particularly for the genes that are involved in sperm production.

82 The ocellated wrasse, *Symphodus ocellatus*, (Fig. 1) allows for a full dissection of the
83 relationship between different aspects of sexual selection in shaping transcriptional
84 dimorphism. *S. ocellatus* males exhibit three morphs (Taborsky, et al. 1987). Nesting males
85 court females, build and defend nests and provide parental care. Sneaker males do not

court females, defend nests or provide care (Taborsky, et al. 1987; Alonzo, et al. 2000). Instead, they surreptitiously join a female and nesting male when they are spawning and release sperm. Satellite males do not build nests or provide care but they do associate with a nesting male and attempt to bring females to and chase sneaker males away from this male's nest. Females strongly prefer nesting males (Alonzo and Warner 2000b; Alonzo 2008). This preference may be adaptive as nesting males are older and/or exhibit faster growth than both sneakers and satellites (Alonzo, et al. 2000).

The high prevalence of multiple males at *S. ocellatus* nest sites, and multiple paternity within nests, indicates that sperm competition in this species is intense, and the intensity varies across male morphs (Alonzo and Warner 1999, 2000a, b; Alonzo and Heckman 2010). The risk and intensity of sperm competition is lowest for nesting males and highest for sneakers (Alonzo and Warner 2000b). Nesting males invest in traits favoured in pre-mating sexual selection (Alonzo 2008; Alonzo and Heckman 2010), as evidenced by the significant differences in somatic transcription among morphs (Alonzo and Warner 2000a; Stiver, et al. 2015). In contrast, sneaker males achieve fertilization success solely through post-mating sexual selection. Satellites are in many ways intermediate, investing in pre-mating behaviors as well as post-mating competitive traits (Stiver and Alonzo 2013; Stiver, et al. 2015). The three male types in this species therefore represent a continuum of pre- and post-mating sexual selection.

The particularly intense levels of sperm competition in *S. ocellatus* make it an ideal system to test whether sperm competition can drive elevated rates of evolution of male-biased genes (Ellegren and Parsch 2007). Higher rates of evolution of male-biased genes have been detected in many species (Ranz, et al. 2003; Zhang, et al. 2004; Cutter and Ward 2005;

Khaitovich, et al. 2005; Harrison, et al. 2015) and is generally thought to be the result of more intense sexual selection acting on males (Andersson 1994), which drives the evolution of male-biased genes. In a similar way, male traits subject to intense sexual selection are also rapidly evolving (Lande 1981). However, rapid rates of evolution could alternatively be non-adaptive, resulting from relaxed constraint or increased drift. Although positive selection has been shown to drive the elevated rate of male-biased gene evolution in *Drosophila* (Pröschel, et al. 2006), recent work in humans and birds (Gershoni and Pietrokovski 2014; Harrison, et al. 2015) suggests drift is the primary cause. However, these species may lack sufficient level of sperm competition to drive rapid rates of male-biased gene evolution. If sperm competition is important in explaining the rapid rates of male-biased gene evolution, it should be evident in *S. ocellatus*, given the intense sperm competition present in this species.

The three male morphs in *S. ocellatus* make it possible to test several aspects of how post-mating sexual selection affects expression and sequence evolution. First, because territorial males invest in costly somatic pre-mating sexually selected traits, it may be possible to identify the signature of this trade-off in the gonadal transcriptome related to post-mating sexual selection. Second, we can use this system to test whether sneaker males, which could be viewed as low quality, invest less in testes transcription compared to satellite males, which likely represent males of higher quality. Finally, the risk and intensity of sperm competition present across the three male morphs make it possible to test the power of sperm competition to shape coding sequence evolution of male-biased genes. Taken together, these analyses allow us to ask how post-mating sexual selection has shaped expression and sequence evolution in the ocellated wrasse with the potential to yield

general insights into how sexual dimorphism, intersexual variation and sexual selection are encoded in the genome.

MATERIALS & METHODS

Sample collection and preparation

The samples were netted using 25.4 x 20.3 cm hand nets in the Baie de Revellata of the Mediterranean Sea near the University of Liege Marine Station (STARESO) Calvi, Corsica using SCUBA. We collected gonad samples from wild individuals, totalling 7 females, 4 nesting males, 5 satellite males and 4 sneaker males during the breeding season of the ocellated wrasse (May and June 2014). All individuals were caught from a total of 6 different nests known to be in the spawning phase of the nest cycle (Lejeune 1985; Taborsky, et al. 1987; Alonzo 2004) meaning that males were courting females and spawning with these females in their nest. Behavioural observations were made for ten minutes prior to capture to verify individual phenotype (nesting male, satellite, sneaker or female) and that all individuals captured were actively involved in reproduction. All individuals were caught within minutes of being observed. Both males and females spawn repeatedly when at an actively spawning nest in this species and only individuals observed to be reproductively active were collected. We aimed to catch one of each type from each nest sampled, though in a few cases this was not possible. Individuals were brought to the surface, euthanized with an overdose of MS-222 and their gonads removed within 10-50 min (mean=27 min) of capture. Collection was authorized by a permit to the field research station STARESO by the French government (Arrêté no. 188 en date du 07 Avril 2014).

Samples were cut into small pieces after dissection (to allow for better preservation) and immediately stored in RNAlater (Ambion). RNA extraction was done under standardised conditions using a Qiagen RNeasy kit (see Table S1 for RNA quality scores). The Wellcome Trust Centre for Human Genetics prepared mRNA libraries (TruSeq RNA Library Prep Kit v2), and each sample was individually barcoded and run across each of four lanes of Illumina HiSeq 2000 in order to eliminate technical variance. On average, we recovered 31 million 100 bp paired-end reads per sample before quality control (Table S1).

Read quality was assessed with FastQC v0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/download.html), and reads quality trimmed with Trimmomatic v0.32 (Lohse, et al. 2012). Specifically, reads were trimmed if the sliding window average Phred score over four bases was < 15 or if the leading/trailing bases had a Phred score < 3. Reads were removed post filtering if either read pair was < 36 bases in length. After trimming there were on average 29 million paired ends reads per sample totalling 199 million reads for females, 116 million reads for nesting males, 149 million reads for satellites and 119 million reads for sneaker males.

De novo *transcriptome assembly, mapping and normalisation*

We used the default parameters in Trinity v2.0.2 (Grabherr, et al. 2011) to construct a *de novo* transcriptome assembly on the combined pool of 583 million paired sequences. Each individual sample was mapped to the Trinity reference genome and RSEM v1.2.19 (Li and Dewey 2011) and Bowtie2 v2.2.4 used to obtain expression levels for the 567,384 contigs. *De novo* transcriptome assemblies generate many non-coding, chimeric or otherwise

176 spurious contigs, and we have previously developed and applied a series of filters to reduce
177 the number of erroneous contigs (Harrison, et al. 2012; Moghadam, et al. 2013; Chen, et al.
178 2015; Harrison, et al. 2015) when genome annotations are unavailable (see Fig S1 for
179 overview). Firstly, we selected the best isoform for each Trinity contig cluster, based upon
180 expression level and, in the case of ties, isoform length (Harrison, et al. 2015). These
181 sequences were subsequently used for analyses of coding sequence divergence. RSEM
182 v1.2.19 was then used to re-map expression to the set of best isoforms to facilitate accurate
183 comparisons between expression and coding sequence evolution. Secondly, non-coding
184 RNA was filtered using a BLASTn with an *E*-value cut-off of 1×10^{-10} between the set of best
185 isoform and *Gasterosteus aculeatus* (stickleback) non-coding RNA (Ensembl v81) (Flicek, et
186 al. 2013). Finally, we removed all contigs with < 2 FPKM in ¾ of the samples per morph. This
187 allowed morph-specific contigs to be retained with reasonable confidence (i.e. expressed in
188 at least 3 out of all the samples) and resulted in 39,453 contigs.

189 Orthology between *G. aculeatus* and *S. symphodus* sequences was assessed using BLAST
190 (Altschul etal. 1990). Specifically, the longest transcript for each gene was obtained for *G.*
191 *aculeatus* (Ensembl v81) (Flicek, et al. 2013) and a reciprocal BLASTn with an E-value cut-off
192 of 1×10^{-10} and minimum percentage identity of 30% was used to identify orthology.
193 Reciprocal orthologs between *G. aculeatus* and *S. symphodus* were identified using the
194 highest BLAST score. Open reading frames were obtained using BLASTx with E-value cut-off
195 of 1×10^{-10} and contigs with invalid open reading frames were removed. This resulted in
196 8,928 orthologous contigs with an average length of 2,951 bp (N50 = 3,575, N90 = 1,656).

197 Normalisation was performed using the TMM function in edgeR (Robinson, et al. 2010) and
198 RPKM values generated. We used hierarchical clustering, factor analysis and pairwise

199 correlations (Spearman's ρ) to identify any potential outliers. One satellite male sample
200 showed significant deviations from the male distribution (Wilcoxon rank sum test p-value =
201 0.0003, mean pairwise correlations among males $\rho = 0.966$, mean pairwise correlations
202 between excluded sample and the remaining satellite samples $\rho = 0.953$) and was removed,
203 and the normalisation was re-run. Pairwise correlations across samples within each morph
204 were high (median (min-max), nesting males = 0.982 (0.980-0.983); satellites = 0.983 (0.981-
205 0.984); sneakers = 0.980 (0.978-0.982); females = 0.977 (0.970-0.981). Post-normalisation
206 expression filtering resulted in the removal of genes that had expression < 2 RPKM in ¾ of
207 the samples per morph, leaving 8,906 contigs for the expression analysis. Average
208 expression for males, females and for each male morph was calculated as the logged mean
209 expression of the normalised data plus 1 (to avoid infinite values resulting from log 0). Males
210 and females had similar average $\log_2$ expression across all genes (female median = 4.05,
211 male median = 4.01, Wilcoxon rank sum test p-value = 0.555).

212

213 ***Sex-biased and morph-biased expression***

214 Differential expression between the sexes was quantified using edgeR (Robinson, et al.
215 2010), using both a fold-change threshold of 2 (Moghadam, et al. 2012; Harrison, et al.
216 2015; Grath and Parsch 2016) across all 3 male morphs *vs* females and $p_{adj} < 0.05$, with an
217 FDR correction for multiple testing. This resulted in 5,448 sex-biased contigs, classified as
218 those contigs with at least twice the expression in one sex compared to the other, as well as
219 $p_{adj} < 0.05$. We also identified sex-biased contigs by comparing each male morph to females
220 (i.e. nesting males *vs* females, sneakers *vs* females and satellites *vs* females). These
221 approaches identified consistent patterns of sex-bias, with an overlap of 94% between the

two methods when using nesting males as the reference, 96% when using sneakers and 99% when using satellite males as the reference. We also used DESeq (v1.18.0) (Anders and Huber 2010) to calculate differential expression between males and females, specifying the same parameters to identify sex-bias as used in the edgeR method. Of the sex-biased genes identified using the edgeR method, 97.4% were also identified as sex-biased using DESeq using the same fold-change thresholds.

Sex-biased contigs were then divided into male-biased genes, those with at least twice the expression in males compared to females and $p_{adj} < 0.05$, and female-biased genes, those with at least twice the expression in females compared to males and $p_{adj} < 0.05$. Because the strength of selection has been shown to correlate with expression level (Krylov, et al. 2003; Nuzhdin, et al. 2004), male-biased (n = 2,590) and female-biased (n = 2,858) contigs were divided into quartiles based on average expression level in females for female-biased contigs and average expression level in males (across the three morphs) for male-biased contigs.

Morph-biased genes are expressed more highly in one morph compared to the other two morphs. They can be expressed in multiple morphs, but at significantly different levels in one morph compared to the other two. Morph-biased contigs were identified using edgeR, comparing differences between the focal morph and the other two morphs (i.e. nesting males *vs* sneaker and satellite males; satellite males *vs* sneaker and nesting males; sneaker males *vs* nesting and satellite males), using an FDR adjustment for false discovery rate at $p_{adj} < 0.05$. Because we expect few contigs to exhibit large expression changes across morphs (Pointer, et al. 2013; Hollis, et al. 2014), we report morph-biased contigs both with and without 2-fold expression thresholds.

245 Hierarchical clustering for the average for each morph was performed using Euclidean
246 clustering in the R package pvclust v1.3-2 (Suzuki and Shimodaira 2006) with bootstrap
247 resampling (1000 replicates) for female-biased and male-biased genes. Heatmaps were
248 generated using $\log_2$ average expression for each gene in each morph using the R package
249 pheatmap v1.0.2. Factor analysis was conducted in R (using Factanal) with varimax rotation.

250 Gene Ontology functional enrichment was assessed for sex-biased genes in each expression
251 quartile using Gene Ontology Enrichment Analysis and Visualization tool (Eden, et al. 2009)
252 using two unranked lists of genes and function ontology specified with a p-value threshold <
253 0.001. Biomart (Ensembl Genes 86) was used to assign orthology between *G. aculeatus* and
254 *Danio rerio*. Sex-biased genes for each quartile were specified as the target gene set with all
255 other genes as the background. Significant GO terms ($p_{fdr} < 0.05$) are listed in Table S2.

256

257 ***Sequence divergence***

258 Coding sequences for *Xiphophorus maculatus* (platyfish) and *G. aculeatus* (stickleback) were
259 obtained from Ensembl v81 and the longest transcript for each gene identified (Flicek, et al.
260 2013). A reciprocal BLASTn with an E-value cut-off of 1×10^{-10} and minimum percentage
261 identity of 30% was used to identify reciprocal 1:1:1 orthologs, resulting in 5,366
262 orthogroups (Fig. S1). Open reading frames were obtained using BLASTx and *G. aculeatus*
263 protein sequences as the BLAST database.

264 Orthologs were aligned with PRANK v140603 in codon mode (Löytynoja and Goldman 2008)
265 specifying the tree ((*X. maculatus*, *G. aculeatus*), *S. ocellatus*). Alignments were quality
266 filtered using SWAMP v.09 (Harrison et al 2014) to remove poorly aligned regions that might

give false signals of position selection. Specifically, codons were masked if there were more than seven non-synonymous mutations in a sliding window scan of 15 codons. Gaps and masked codons were removed from the alignment and orthogroups discarded if the length < 300bp.

We obtained divergence estimates for each orthogroup using the branch model (model=2, nssites=0) in the CODEML package in PAML v4.8 (Yang 2007) using the tree (*X. maculatus*, *G. aculeatus*, *S. ocellatus*#1). The branch model was used to calculate d_N/d_S for the wrasse-specific branch. Contigs were excluded if tree length $d_S > 2$ in order to remove sequences which have reached mutational saturation (Axelsson, et al. 2008). This resulted in 4,912 orthogroups remaining out of a total of 5366.

For each expression class, we calculated mean d_N and mean d_S from the PAML outputs as the sum of the number of substitutions across all contigs in a given category divided by the number of sites ($d_N = D_N/N$; $d_S = D_S/S$; where D_N = number of non-synonymous substitutions, N = number of non-synonymous sites, D_S = number of synonymous substitutions, S = number of synonymous sites). This approach avoids the problems of infinitely high d_N/d_S estimates arising from contigs with extremely low d_S (Harrison, et al. 2015; Wright, et al. 2015) and prevents disproportionate weighting and skew from shorter contigs. 1000 bootstrap replicates were generated to estimate 95% confidence intervals and permutations tests were used to test for significant differences between pairwise comparisons.

Polymorphism analysis

289 Polymorphism data was obtained by first mapping RNA-seq reads to the best isoform Trinity
290 assembly using the two-pass alignment method of the STAR aligner v2.4.2a with default
291 parameters (Dobin, et al. 2013). Only uniquely mapping reads were retained. SAMTOOLS
292 mpileup v0.1.19 (Li, et al. 2009) and VARSCAN2 v2.3.9 mpileup2snp (Koboldt, et al. 2009;
293 Koboldt, et al. 2012) were used to call SNPs. SAMTOOLS mpileup was run with the
294 probabilistic alignment disabled, a max read depth of 10,000,000 and default minimum base
295 quality of 13. VARSCAN2 mpileup2snp was run with a minimum frequency for homozygote
296 of 0.85, minimum coverage of 2, minimum average quality of 20, strand-filter on and p-
297 value = 1. Valid SNPs were required to have a minimum coverage of 20 in at least four
298 individuals and a minor allele frequency > 0.15, resulting in 218,913 SNPs. SNPs were
299 matched to the reading frame to determine whether they were synonymous or
300 nonsynonymous.

301 In order to ensure the divergence and polymorphism data was comparable for subsequent
302 analyses, similar criteria were used to filter both analyses. Specifically, codons that (i) were
303 masked by SWAMP (ii) failed the minimum coverage threshold of 20 in at least four
304 individuals or (iii) were excluded from PAML due to alignment gaps and the clean filter
305 function, were filtered from both the polymorphism and divergence analyses.

306 The McDonald–Kreitman test (McDonald and Kreitman 1991) was used to estimate the
307 number of contigs evolving under adaptive and neutral evolution by contrasting the number
308 of nonsynonymous and synonymous substitutions (D_N and D_S) with polymorphisms (P_N and
309 P_S). Fisher's Exact tests were run for each contig using D_N, D_S, P_N and P_S. Contigs were
310 removed if the total observations across rows and columns in the 2x2 contingency table was
311 < 6 (Begun, et al. 2007; Andolfatto 2008). For those contigs with significant deviations in D_N,

312 D_S, P_N and P_S, positive selection was indicated by $D_N/D_S > P_N/P_S$ (McDonald and Kreitman
313 1991).

314 To examine expression levels of genes under positive selection we expanded our group of
315 genes under putative positive selection by using the Direction of Selection (DoS) test
316 (Stoletzki and Eyre-Walker 2011). DoS (DoS = $D_n/(D_n + D_s) - P_n/(P_n + P_s)$) was calculated as
317 the difference in the proportion of fixed non-synonymous sites and the proportion of
318 polymorphic non-synonymous sites (Stoletzki and Eyre-Walker 2011). An excess of non-
319 synonymous substitutions compared to polymorphisms (i.e. DoS > 0) indicates putative
320 positive selection.

321 Lastly, we tested morph-biased genes (identified within the expression analysis) for standing
322 variation using polymorphism data to test for an excess or under-representation of
323 nonsynonymous polymorphisms across morph-biased genes. Excess or underrepresentation
324 is indicative of relaxed purifying selection or positive selection, respectively. For this
325 analysis, we separately concatenated P_N and P_S for each gene class and used Fisher's Exact
326 tests (in R v3.1.3) to test for significant differences in P_N/P_S between pairwise comparisons
327 between morph-biased genes and male-biased genes.

328 Commands are included in the supplementary material.

329

330 **RESULTS**

331 We had a total of 583 million paired-end reads across all samples after trimming, which we
332 used for *de novo* transcriptome assembly. After filtering our assembly, we recovered 8,928

reciprocal orthologs with *Gasterosteus aculeatus* (stickleback), representing coding sequence used for all downstream analysis.

Sex-biased expression across male morphs

We estimated expression for each sample, based on an average of 29 million paired-end reads after trimming. We identified greater inter-sexual expression variation, with 5,448 sex biased genes ($\log_2$ M:F > 1 or < -1, $p_{adj} < 0.05$) compared to intra-sexual expression variation, with 34 morph-biased genes (24 nesting male-biased, 2 satellite male-biased, 8 sneaker male-biased, $\log_2$ fold change between morph comparisons > 1 or < -1, $p_{adj} < 0.05$). Of the nesting male-biased contigs, 1 contig was also male-biased and 2 were female-biased, the satellite male-biased contigs 1 was also male-biased and 1 female-biased, and the sneaker male-biased contigs 3 were male-biased and 3 female-biased. Previous work has indicated few contigs have large expression changes among morphs (Pointer, et al. 2013; Hollis, et al. 2014), therefore we also assessed morph-bias without expression thresholds, and only based on statistical thresholds ($p_{adj} < 0.05$). Using this more relaxed threshold, we recovered 41 nesting male-biased contigs, 11 satellite male-biased contigs, and 9 sneaker male-biased contigs.

We first used these expression estimates for hierarchical clustering, which can be used to assess overall transcriptional similarity across morphs and sexes. For male- and female-biased genes, male morphs cluster more closely to each other than to females (Fig 2). Our clustering also indicates some intra-sexual variation among male morphs, as sneaker and satellite males show greater transcriptional similarity to each other than to nesting males

355 for female-biased genes expressed in the gonad. All three male morphs were statistically
356 indistinguishable via bootstrapping in the hierarchical clustering across male-biased genes.
357 Intra-sexual variation in expression was also evident with factor analysis based on all
358 expressed genes, which indicated greater transcriptional difference between nesting males
359 and satellite males (Fig. 3). However, sneaker males showed substantial variation across
360 both factors, and overlap with both nesting males and satellites (Fig. 3).

361 In order to test for differences in transcriptional investment among the three morphs we
362 next tested for masculinization and feminization of gonadal expression (Jaquiery, et al.
363 2013; Pointer, et al. 2013; Hollis, et al. 2014) in each of the male morphs, and combined that
364 with estimates of expression variance. Masculinization and feminization of gene expression
365 is the increase in expression of male-biased or female-biased genes, respectively. Similarly
366 demasculinization and defeminisation of gene expression is the reduction in expression of
367 male-biased and female-biased genes, respectively. Gene expression variance estimates are
368 increasingly used to infer selection acting on expression level under assumptions that
369 selection on expression will decrease expression variance across replicates (Moghadam, et
370 al. 2012; Romero, et al. 2012; Dean, et al. 2015). Nesting males express female-biased genes
371 at significantly higher levels compared to the other male morphs at the lower expression
372 levels (Fig. 4A). Although initially surprising given recent studies in birds and *Drosophila*
373 (Pointer, et al. 2013; Hollis, et al. 2014), our results indicate that nesting males also exhibit
374 higher variance in expression for female-biased genes at the lower expression levels (Fig.
375 5A) compared to the other morphs, suggesting that although they show some feminization,
376 it is unlikely to be due to selection acting to increase expression. In contrast, satellite males
377 had higher expression (Fig. 4B) and lower variance (Fig. 5B) for male-biased genes,

particularly for highly expressed male-biased genes, suggesting that the gonadal masculinization is the response to selection for higher expression in this morph. These results do not qualitatively change if we increase the sex-biased threshold to four-fold expression differences between the sexes (Figure S2, S3), suggesting these results are robust to comparisons between samples that potentially vary in tissue composition (Harrison, et al. 2015; Montgomery and Mank 2016). Gene Ontology terms for sex-biased genes in the different expression level quartiles are presented in Table S2.

Rates of evolution

In order to test the power of sperm competition to shape gene sequence evolution we compared rates of evolution and population-level polymorphism across sex-biased and unbiased genes. As observed in many other animals (Ellegren and Parsch 2007), male-biased gonadal genes showed higher rates of gene sequence evolution than unbiased genes, driven by both an increase in d_N and a decrease in d_S (Table 1). Interestingly, female-biased genes in our wrasse data also showed significantly higher rates of sequence evolution, again driven by both an increase in d_N and a decrease in d_S (Table 1). McDonald-Kreitman (MK) tests (McDonald and Kreitman 1991) revealed a higher proportion of male-biased genes with signatures of positive selection than unbiased genes (Table 2), indicating adaptive evolution explains at least some of the elevated rate of evolution for male-biased genes. However, only five female-biased genes showed significant evidence of positive selection with McDonald-Kreitman tests, which was not significantly different from the level observed for unbiased genes. We also tested morph-biased genes for differences in rates of evolution and standing variation. Although the small number of morph-biased genes results in low

401 statistical power (Tables S3 and S4), nesting-male biased genes do show a marginally
402 significant elevated rate of evolution compared to male-biased genes (Table S3), consistent
403 with previous work in the bulb mite and pea aphid (Stuglik et al 2014, Purandare et al 2014).
404 Our analysis of standing polymorphism suggests that this fast rate of evolution may be due
405 in part to drift (Table S4).

406 To test whether male-biased genes under positive selection have highest expression in
407 morphs subject to strong post-mating sexual selection, we expanded our group of genes
408 under putative positive selection by using the Direction of Selection (DoS) test (Stoletzki and
409 Eyre-Walker 2011), which is more permissive than the MK test. Satellite males express male-
410 biased genes with DoS > 0 (indicative of putative positive selection) at higher levels than
411 nesting males (Fig. 6). Male-biased genes under putative positive selection also tend to be
412 more highly expressed (Fig. S4A) but tend to be less male-biased in their expression (Fig.
413 S4B) than genes under relaxed constraint.

414

415 **DISCUSSION**

416 Here we report patterns of regulatory and sequence evolution in the gonad transcriptome
417 of wild caught ocellated wrasse, a species with three male morphs which experiences
418 particularly high levels of sperm competition and a continuum of pre- and post-mating
419 sexual selection (Alonzo and Warner 2000b; Alonzo and Heckman 2010). Previous work in
420 this system has revealed somatic variation in gene expression among the male morphs that
421 may be important in pre-mating sexual selection (Stiver, et al. 2015). Here we focus on the
422 consequences of post-mating sexual selection on transcriptome evolution in the gonads.

Gene expression and sexual selection

The design of our study makes it possible to compare both inter and intra-sexual transcriptional variation to explore how post-mating sexual selection shapes the gonadal transcriptome and drives gene sequence evolution. Previous work on somatic tissues showed equal or greater intra-sexual than inter-sexual differences in transcription (Snell-Rood, et al. 2011; Schunter, et al. 2014). This is in clear contrast to our results here (Fig. 1) and our previous work on male morphs in wild turkeys (Pointer, et al. 2013), which show the greatest difference is first by sex, then within sex by morph. This may reflect fundamental differences between somatic and gonad transcriptional variation, as somatic tissues in general tend to show far less inter-sexual variation than the gonad (Pointer, et al. 2013; Harrison, et al. 2015), in the latter case resulting from the profound functional and physiological differences in producing and delivering male versus female gametes.

Relatively few genes showed significant expression differences among male morphs, however average expression across male- and female-biased gene categories was similar to previous studies in birds (Pointer, et al. 2013), and mites (Stuglik, et al. 2014). Previous work has suggested that sex-biased genes shift expression in response to sexual selection (Hollis, et al. 2014; Immonen, et al. 2014) and are correlated with the magnitude of male sexually selected traits (Pointer, et al. 2013). Based on this, we might expect territorial males in *S. ocellatus* to show greater male-biased expression. However, we observed the opposite, and territorial males instead exhibit significant feminization of expression for female-biased genes. Although models of gene expression evolution have yet to be validated, the high variance exhibited across replicates suggests that expression of female-biased genes is

unlikely to be the result of positive selection in territorial males. High expression variance in female-biased genes, coupled with the high investment nesting males make in somatic traits (Alonzo 2008; Alonzo and Heckman 2010; Alonzo and Warner 2000a; Stiver et al 2015), is consistent with a trade-off between maintaining costly pre-mating sexually selected traits and the constraints of post-mating sexual selection shaping gene expression in the gonad.

In contrast to territorial males, satellite males showed the highest expression level for male-biased genes, consistent with the greater transcriptional investment in genes likely to be important in post-mating sexual selection compared to nesting males. Satellites also showed significantly less variation in expression for both male- and female-biased genes, consistent with positive selection under recent models of gene expression evolution (Moghadam, et al. 2012; Romero, et al. 2012; Dean, et al. 2015). Satellite males also have the largest absolute gonad mass out of the three morphs (Alonzo & Stiver unpublished data), illustrating concordance between male-biased gene expression and absolute gonadal mass. These results are consistent with the possibility that, freed from the costs of pre-mating somatic sexually selected traits, satellite males are able to invest more in post-mating sexually selected gene expression patterns that may aid them in sperm competition. This is somewhat at odds with results from the wild turkey, where subordinate males, which are analogous in many ways to satellite males in the wrasse, show reduced expression of male-biased genes (Pointer, et al. 2013). However, it is worth noting that subordinate male turkeys are effectively non-reproductive (Krakauer 2005, 2008), and therefore do not experience sperm competition.

Sneaker males show defeminization of female-biased genes demasculinization of male-biased expression and high variance, indicating no directional selection and suggesting that

469 they are simply lower quality on average than the other morphs. This is consistent with the
470 life history trajectory of sneaker males, which tend to be the males with slowest early
471 growth rate (Alonzo, et al. 2000). Small males typically breed as 1 year old sneaker males,
472 then go on to become 2 year old satellite males, never becoming nesting males, while larger
473 males become satellites as 1 year olds then nesting males as 2 year olds (Alonzo et al. 2000).
474 Though, sneaker males release the most sperm per spawn out of all three morphs (Alonzo
475 and Warner 2000a), they also generate the least total sperm (across successive mating
476 attempts) and have the lowest average individual mating success and paternity of all three
477 morphs (Alonzo and Warner 2000b; Alonzo et al. in prep). Therefore, demasculinization and
478 high variance of male-biased gene expression may reflect the low overall male quality of
479 sneaker males.

480 Alternative mating tactics are common in fish, and appear to have evolved many times
481 independently (Mank and Avise 2006). Given the repeated origin of these phenotypes, it will
482 be interesting for future studies to determine whether our observed patterns of gonadal
483 gene expression differentiation among morphs are conserved among systems with similar
484 mating systems.

485 ***Sperm competition and sequence evolution***

486 Our results are somewhat unusual compared to previous studies in animals in that both
487 male- and female-biased genes show elevated rates of sequence evolution. This was due to
488 elevated d_N and a reduced d_S in both male- and female-biased genes, and suggests that
489 sexual selection alone, which in this species would act more forcefully on males, cannot
490 explain elevated rates of evolution observed in both sexes. Although female-biased genes
491 have been observed to evolve more rapidly in a yeast (Whittle and Johannesson 2013), and

both male- and female-biased genes exhibit faster rates of evolution in an alga (Lipinska, et al. 2015), expression studies in adult animals have tended to show faster rates of evolution primarily in male-biased genes (Ellegren and Parsch 2007). It is not clear at this point whether our results represent a species-specific pattern, or are exhibited by other species with similar mating systems. We also note that the contigs we removed from the d_N/d_S analysis due to mutational saturation could also have been the most rapidly evolving genes, causing an overall underestimation of divergence.

Interestingly, recent work in birds (Harrison, et al. 2015) and humans (Gershoni and Pietrokovski 2014) have suggested that fast rates of evolution for male-biased genes might be due to relaxed constraint rather than positive selection, and this is consistent with studies in insects showing relaxed constraint characterizes caste-biased genes (Hunt, et al. 2011). In contrast, our results show that male-biased genes have a higher proportion of loci showing evidence of positive selection, suggesting that at least some of the acceleration in rates of evolution for male-biased genes is due to adaptive evolution. This may be due to the intense level of sperm competition experienced by males of these species (Alonzo and Heckman 2010) and the resulting strength of post-mating sexual selection among alternative male types. Extended haploid selection due to external fertilization may also explain our results.

In summary, the alternative mating strategies of *S. ocellatus* make it possible to isolate the complex effects of post-mating sexual selection on gonadal genome evolution. Our results reveal the potential for post-mating sexual selection to masculinize the transcriptome and drive adaptive evolution of male-biased genes.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation (under Grant Number IOS-0950472 to SHA), the European Research Council (grant agreements 260233 and 680951 to JEM), Marie Curie Actions (grant agreement 655392 to RD) and Australian Research Council (DE150101853 to RD). We thank Pierre Lejeune, Alexandre Volpon and the staff of Station de Recharches Sous-marins at Océangraphique for their support in the field and Holly Kindsvater for sample collection. The authors acknowledge the use of the UCL Legion High Performance Computing Facility (Legion@UCL), and associated support services, in the completion of this work. We also thank S. Montgomery, V. Oostra, I. Darolti, N. Bloch and 3 anonymous reviewers for helpful comments and suggestions.

REFERENCES

Alonzo SH. 2008. Female mate choice copying affects sexual selection in wild populations of the ocellated wrasse. Animal Behaviour 75:1715-1723.

Alonzo SH. 2004. Uncertainty in territory quality affects the benefits of usurpation in a Mediterranean wrasse. Behavioral Ecology 15:278-285.

Alonzo SH, Heckman KL. 2010. The unexpected but understandable dynamics of mating, paternity and paternal care in the ocellated wrasse. Proceedings of the Royal Society B-Biological Sciences 277:115-122.

Alonzo SH, Taborsky M, Wirtz P. 2000. Male alternative reproductive behaviours in a Mediterranean wrasse, Symphodus ocellatus: Evidence from otoliths for multiple life-history pathways. Evolutionary Ecology Research 2:997-1007.

Alonzo SH, Warner RR. 2000a. Allocation to mate guarding or increased sperm production in a Mediterranean wrasse. American Naturalist 156:266-275.

Alonzo SH, Warner RR. 2000b. Dynamic games and field experiments examining intra- and intersexual conflict: explaining counterintuitive mating behavior in a Mediterranean wrasse, Symphodus ocellatus. Behavioral Ecology 11:56-70.

Alonzo SH, Warner RR. 1999. A trade-off generated by sexual conflict: Mediterranean wrasse males refuse present mates to increase future success. Behavioral Ecology 10:105-111.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. Genome Biology 11:R106.

Andersson M. 1994. Sexual Selection. Princeton, NJ: Princeton University Press.

Andolfatto P. 2008. Controlling Type-I Error of the McDonald-Kreitman Test in Genomewide Scans for Selection on Noncoding DNA. Genetics 180:1767-1771.

Axelsson E, Hultin-Rosenberg L, Brandstrom M, Zwahlen M, Clayton DF, Ellegren H. 2008. Natural selection for avian protein-coding genes expressed in brain. Molecular Ecology 17:3008-2017.

Bailey NW, Veltsos P, Tan YF, Millar AH, Ritchie MG, Simmons LW. 2013. Tissue-Specific Transcriptomics in the Field Cricket Teleogryllus oceanicus. G3-Genes Genomes Genetics 3:225-230.

Begun DL, Holloway Ak, Stevens K, Hillier L, Poh Y-P, Hahn MW, Nista PM, Jones CD, Kern AD, Dewey CN, et al. 2007. Population genomics: Whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. PLoS Biology 5:e310.

Chen Y-C, Harrison PW, Kotrschal A, Kolm N, Mank JE, Panula P. 2015. Expression change in Angiopoietin-1 underlies change in relative brain size in fish. Proceedings of the Royal Society B-Biological Sciences 282.

Cutter AD, Ward S. 2005. Sexual and temporal dynamics of molecular evolution in *C. elegans* development. Molecular Biology and Evolution 22:178-188.

Dean R, Harrison PW, Wright AE, Zimmer F, Mank JE. 2015. Positive Selection Underlies Faster-Z Evolution of Gene Expression in Birds. Molecular Biology and Evolution 32:2646-2656.

Dean R, Mank JE. 2016. Tissue-specific and sex-specific regulatory variation permit the evolution of sex-biased gene expression. American Naturalist 188:E74-E84.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batus P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15-21.

Eden E, Navon R, Steinfold I, Lipson D, Yakhini. 2009. GOrilla: A tool for discovery and visualization of enriched FO terms in ranked gene lists. BMC Bioinformatics 10:48.

Ellegren H, Parsch J. 2007. The evolution of sex-biased genes and sex-biased gene expression. Nature Reviews Genetics 8:689-698.

Emlen DJ, Warren IA, Johns A, Dworkin I, Lavine LC. 2012. A mechanism of extreme growth and reliable signaling in sexually selected ornaments and weapons. Science 337:860-864.

Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fairley S, et al. 2013. Ensembl 2013. Nucleic Acids Research 41:D48-D55.

Gershoni M, Pietrokovski S. 2014. Reduced selection and accumulation of deleterious mutations in genes exclusively expressed in men. Nature Communications 5:4438.
Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29:644-U130.
Grath S, Parsch J. 2016. Sex-Biased Gene Expression. Annual Review of Genetics 50.
Harrison PW, Mank JE, Wedell N. 2012. Incomplete Sex Chromosome Dosage Compensation in the Indian Meal Moth, Plodia interpunctella, Based on De Novo Transcriptome Assembly. Genome Biology and Evolution 4:1118-1126.
Harrison PW, Wright AE, Zimmer F, Dean R, Montgomery SH, Pointer MA, Mank JE. 2015. Sexual selection drives evolution and rapid turnover of male gene expression. Proceedings of the National Academy of Sciences, USA 112:4393-4398.
Hollis B, Houle D, Yan Z, Kawecki TJ, Keller L. 2014. Evolution under monogamy feminizes gene expression in *Drosophila melanogaster*. Nature Communications 5:3482.
Hunt BG, Ometto L, Wurm Y, Shoemaker D, Yi SV, Keller L, Goodisman MAD. 2011. Relaxed selection is a precursor to the evolution of phenotypic plasticity. Proceedings of the National Academy of Sciences, USA 108:15936-15941.
Immonen E, Snook RR, Ritchie MG. 2014. Mating system variation drives rapid evolution of the female transcriptome in *Drosophila pseudoobscura*. Ecology and Evolution 4:2186-2201.
Jaquiery J, Rispe C, Roze D, Legeai F, Le Trionnaire G, Stoeckel S, Mieuzet L, Da Silva C, Poulain J, Prunier-Leterme N, et al. 2013. Masculinization of the X Chromosome in the Pea Aphid. Plos Genetics 9:e1003690.
Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Paabo S. 2005. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. Science 309:1850-1854.
Khila A, Abouheif E, Rowe L. 2012. Function, developmental genetics, and fitness consequences of a sexually antagonistic trait. Science 336:585-589.
Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstock GM, Wilson RK, Ding L. 2009. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics 25:2283-2285.
Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Research 22:568-576.
Koerich LB, Wang XY, Clark AG, Carvalho AB. 2008. Low conservation of gene content in the *Drosophila* Y chromosome. Nature 456:949-951.
Krakauer AH. 2005. Kin selection and cooperative courtship in wild turkeys. Nature 434:69-72.
Krakauer AH. 2008. Sexual selection and the genetic mating system of wild turkeys. Condor 110:1-12.
Krylov DM, Wiolf YI, Rogozin IB, Koonin EV. 2003. Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. Genome Research 13:2229-2235.
Lande R. 1981. Models of speciation by sexual selection on polygenic traits. Proceedings of the National Academy of Sciences of the United States of America 78:3721-3725.
Lejeune P. 1985. Etude écoéthologique des comportements reproducteur et sociaux des Labridae méditerranéens des genres Symphodus (Rafinesque 1810) et Coris (Lacepede 1802). Cahier d'Ethologie Appliquee 5:1-208.
Li B, Dewey CN. 2011. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPDP. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-2079.

Lipinska A, Cormier A, Luthringer R, Peters AF, Corre E, Gachon CM, Cock JM, Coelho SM. 2015. Sexual Dimorphism and the Evolution of Sex-Biased Gene Expression in the Brown Alga Ectocarpus. Molecular Biology and Evolution 32:1581-1597.
Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, Usadel B. 2012. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Research 40:W622-W627.
Löytynoja A, Goldman N. 2008. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. Science 320:1632-1635.
Mank JE, Avise JC. 2006. Comparative phylogenetic analysis of male alternative reproductive tactics in ray-finned fishes. Evolution 60:1311-1316.
Mank JE, Wedell N, Hosken DJ. 2013. Polyandry and sex-specific gene expression. Philosophical Transactions of the Royal Society, B 368:20120047.
McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the ADH locus in Drosophila. Nature 351:652-654.
Moghadam HK, Harrison PW, Zachar G, Szekely T, Mank JE. 2013. The plover neurotranscriptome assembly: transcriptomic analysis in an ecological model species without a reference genome. Molecular Ecology Resources 13:696-705.
Moghadam HK, Pointer MA, Wright AE, Berlin S, Mank JE. 2012. W chromosome expression responds to female-specific selection. Proceedings of the National Academy of Sciences, USA 109:8207-8211.
Montgomery SH, Mank JE. 2016. Inferring regulatory change from gene expression: the confounding effects of tissue scaling. Molecular Ecology.
Nuzhdin SV, Wayne ML, Harmon KL, McIntyre LM. 2004. Common pattern of evolution of gene expression level and protein sequence in Drosophila. Molecular Biology and Evolution 21:1308-1317.
Parsch J, Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene expression. Nature Reviews Genetics 14:83-87.
Pointer MA, Harrison PW, Wright AE, Mank JE. 2013. Masculinization of gene expression is associated with exaggeration of male sexual dimorphism. Plos Genetics 9:e1003649.
Pröschel M, Zhang Z, Parsch J. 2006. Widespread adaptive evolution of drosophila genes with sex-biased expression. Genetics 174:893-900.
Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL. 2003. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. Science 300:1742-1745.
Robinson M, CMcCarthy D, Smyth G. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression dats. Bioinfomratics 26:139-140.
Romero IG, Ruvinsky I, Gilad Y. 2012. Comparative studies of gene expression and the evolution of gene regulation. Nature Reviews Genetics 13:505-516.
Schunter C, Vollmer SV, Macpherson E, Pascual M. 2014. Transcriptome analyses and differential gene expression in a non-model fish species with alternative mating tactics. Bmc Genomics 15.
Sharma E, Kunstner A, Fraser BA, Zipprich G, Kottler VA, Henz SR, Weigel D, Dreyer C. 2014. Transcriptome assemblies for studying sex-biased gene expression in the guppy, Poecilia reticulata. Bmc Genomics 15.
Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, et al. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825-U822.
Small CM, Carney GE, Mo QX, Vannucci M, Jones AG. 2009. A microarray analysis of sex- and gonad-biased gene expression in the zebrafish: Evidence for masculinization of the transcriptome. Bmc Genomics 10:579.
Snell-Rood EC, Cash A, Han MV, Kijimoto T, Andrews J, Moczek AP. 2011. Developmental decoupling of alternative phenotypes: Insights from the transcriptomes of horn-polyphenic beetles. Evolution 65:231-245.

675 Stiver KA, Alonzo SH. 2013. Does the Risk of Sperm Competition Help Explain Cooperation between
676 Reproductive Competitors? A Study in the Ocellated Wrasse (Symphodus ocellatus). American
677 Naturalist 181:357-368.
678 Stiver KA, Harris RM, Townsend JP, Hofmann HA, Alonzo SH. 2015. Neural Gene Expression Profiles
679 and Androgen Levels Underlie Alternative Reproductive Tactics in the Ocellated Wrasse, Symphodus
680 ocellatus. Ethology 121:152-167.
681 Stoletzki N, Eyre-Walker A. 2011. Estimation of the Neutrality Index. Molecular Biology and Evolution
682 28:63-70.
683 Stuglik MT, Babik W, Prokop Z, Radwan J. 2014. Alternative reproductive tactics and sex-biased gene
684 expression: the study of the bulb mite transcriptome. Ecology and Evolution 4:623-632.
685 Suzuki R, Shimodaira H. 2006. Pvclust: an R package for assessing the uncertainty in hierarchical
686 clustering. Bioinformatics 22:1540-1542.
687 Taborsky M, Hudde B, Wirtz P. 1987. Reproductive behavior and ecology of Symphodus (Crenilabrus)
688 Ocellatus, a European wrasse with 4 types of male behavior. Behaviour 102:82-118.
689 Whittle CA, Johannesson H. 2013. Evolutionary Dynamics of Sex-Biased Genes in a Hermaphrodite
690 Fungus. Molecular Biology and Evolution 30:2435-2446.
691 Wright AE, Harrison PW, Zimmer F, Montgomery SH, Pointer MA, Mank JE. 2015. Variation in
692 promiscuity and sexual selection drives avian rate of Faster-Z evolution. Molecular Ecology 24:1218-
693 1235.
694 Yang ZH. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Molecular Biology and
695 Evolution 24:1586-1591.
696 Zhang Z, Hambuch TM, Parsch J. 2004. Molecular evolution of sex-biased genes in *Drosophila*.
697 Molecular Biology and Evolution 21:2130-2139.

698

699 **DATA ACCESSIBILITY**

700 Data are available in the NCBI short read archive (www.ncbi.nlm.nih.gov/sra) BioProject ID
701 PRJNA344726.

702

703 **AUTHOR CONTRIBUTIONS**

704 JEM, SHA, RD designed research, SEM, BMN, SHA performed research, RD, AEW analysed
705 data, RD, JEM, SHA, AEW wrote paper, all authors revised paper.

706

TABLES

Table 1. Rates of evolution (d_N/d_S) for sex biased genes. Values for sex-biased gene categories that are significantly different from unbiased genes are in bold.

Expression class	Total contigs	Orthologs[a]	Filter[b]	d_N (95% CI) significance[c]	d_S (95% CI) significance[c]	d_N/d_S (95% CI) significance[c]
Male-biased	2590	1567	1435	**0.022 (0.021-0.023) *P < 0.0001***	**0.344 (0.337-0.351) *P < 0.0001***	**0.064 (0.061-0.067) *P < 0.0001***
Female-biased	2858	1603	1481	**0.023 (0.022-0.024) *P < 0.0001***	**0.347 (0.342-0.353) *P = 0.024***	**0.066 (0.063-0.068) *P < 0.0001***
Unbiased	3458	2175	1996	0.021 (0.020-0.021)	0.351 (0.345-0.356)	0.059 (0.057-0.061)

[a]Number of contigs that are 1:1:1 orthologs with *X. maculatus* and *G. aculeatus.*
[b]Number of 1:1:1 orthologs after filtering.
[c]Significance based on 2-tailed permutations tests (1000 replicates), compared to unbiased genes

Table 2. McDonald-Kreitman tests for selection

Expression class	Contigs[a]	Positive selection[b] (significance)
Male-biased	425	**19 (*P = 0.0056*)**
Female-biased	231	5 *n.s*
Unbiased	623	9

[a]Number of 1:1:1 orthologs after filtering.
[b]Number of contigs with significant positive selection (significant deviations in D_N, D_S, P_N and P_S, and $D_N/D_S > P_N/P_S$). P values from Fisher's Exact test compare sex-biased to unbiased genes.

FIGURES

Figure 1. The ocellated wrasse (*Symphodus ocellatus*) has three male morphs. Brightly coloured nesting males (top left) hold territories and females (bottom left) prefer to mate with nesting males. Satellites associate (bottom right) with nesting males and help defend the nest by chasing away sneakers (top right). Illustration credit: Clara Lacy.

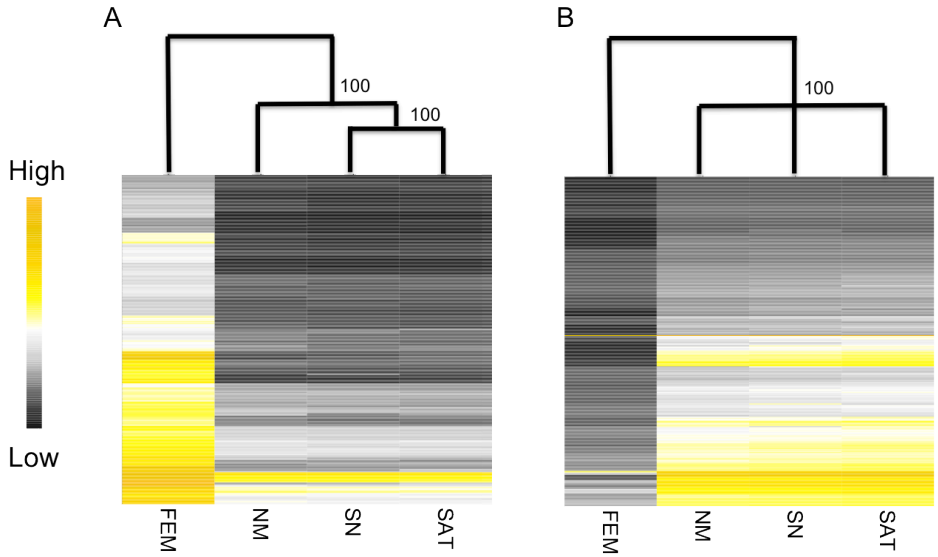


Figure 2. Hierarchical clustering and heat maps based on average within-morph expression for females (FEM), nesting males (NM), satellite males (SAT) and sneaker males (SN) for (A) female-biased genes (n = 2,858) and (B) male-biased genes (n = 2,590). Confidence intervals for hierarchical clustering are based on 1000 bootstrap replicates, and branches with <80% support have been collapsed.

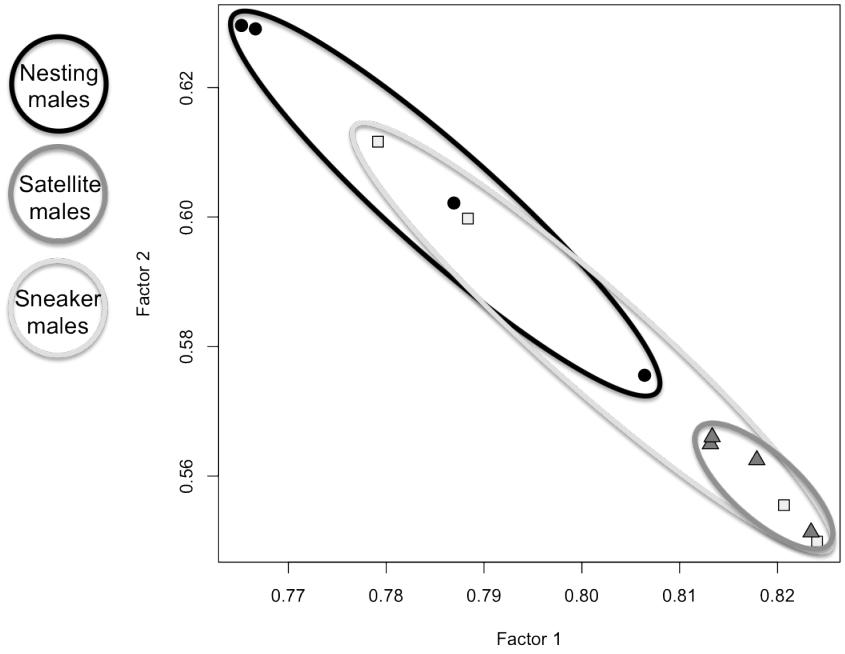
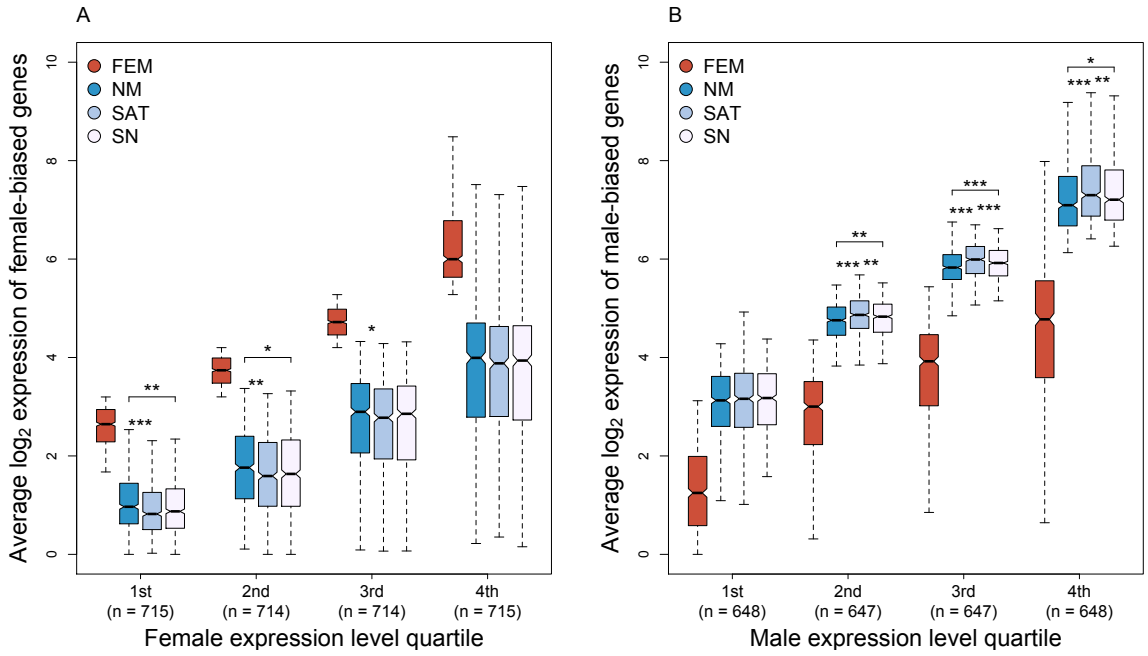


Figure 3. Factor analysis of three male morphs based on normalized RPKM values for all expressed contigs. Nesting males are identified with circles and a black ellipse, satellite males with triangles and a dark grey ellipse, and sneaker males with squares and a light grey ellipse.



749

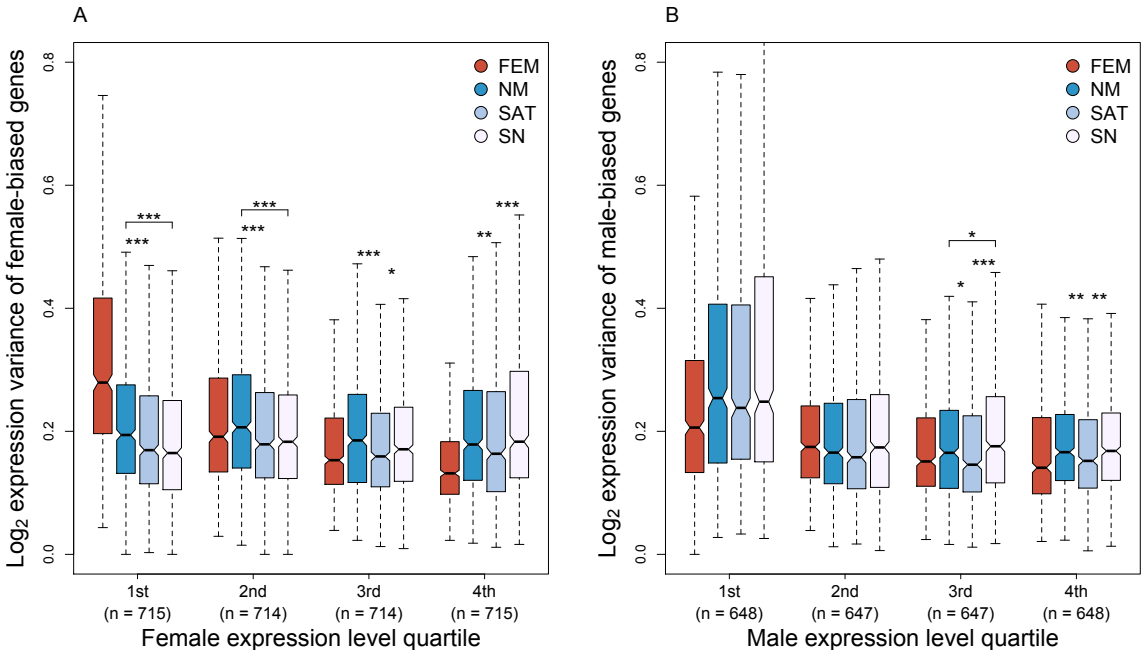
750 **Figure 4**. Gene expression for each of the morphs for (A) female-biased genes and (B) male-
751 biased genes. Data is divided into quartiles based upon expression level in females for panel
752 (A) and males for panel (B). Red = females (FEM), dark blue = nesting males (NM), light blue
753 = satellite males (SAT) and white = sneaker males (SN). Significance is indicated based on
754 Wilcoxon rank sum tests (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$).

755

756

757

758



759

760 **Figure 5**. Gene expression variance for each of the morphs for (A) female-biased genes and
761 (B) male-biased genes. Data is divided into quartiles based upon expression level in females
762 for panel (A) and males for panel (B). Red = females, dark blue = nesting males, light blue =
763 satellite males and white = sneaker males. Significance is indicated based on Wilcoxon rank
764 sum tests (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$).

765

766

767

768

769

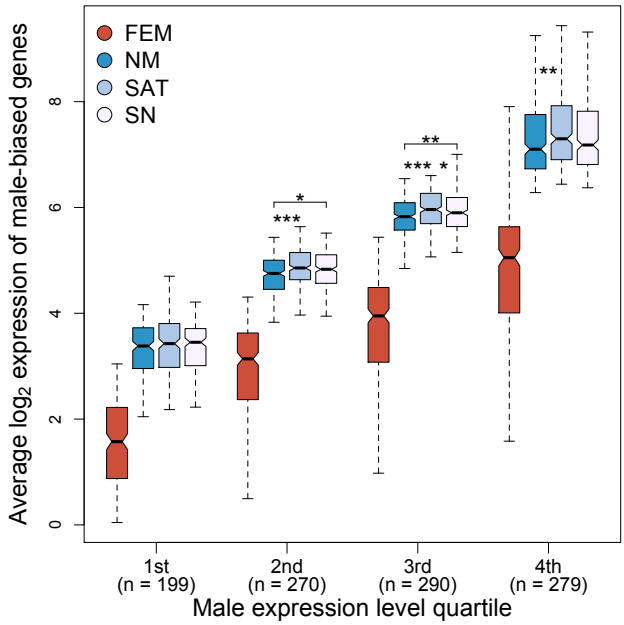


Figure 6. Male-biased gene expression for genes under putative positive selection for each of the morphs. Red = females, dark blue = nesting males, light blue = satellite males and white = sneaker males. Significance is indicated based on Wilcoxon rank sum tests (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$).

Table S1. RNA quality scores and raw and post-trimming read counts per sample

Morph	Sample	RNA quality		Read number		% retained
		260/280	260/230	Raw paired reads	Paired reads after trimming	
Female	14003	2.09	2.32	32,912,641	30,062,449	91.4
	14007	2.11	1.70	26,425,448	24,203,325	91.6
	14012	2.09	2.28	28,318,371	26,370,307	93.1
	14013	2.10	2.30	26,683,758	25,049,998	93.9
	14016	2.10	2.30	36,207,483	33,893,881	93.6
	14018	2.12	2.33	29,186,747	27,638,121	94.7
	14019	2.09	2.32	33,958,247	31,701,901	93.4
Nesting male	14004	2.08	2.24	32,343,023	30,122,757	93.1
	14008	2.13	2.26	32,060,846	29,487,537	92.0
	14011	2.10	2.21	27,493,908	25,817,118	93.9
	14021	2.11	1.69	32,753,374	30,686,654	93.7
Satellite	14002	2.10	2.26	31,681,769	29,615,187	93.5
	14006	2.09	2.28	31,704,381	29,925,131	94.4
	14009	2.09	1.96	30,237,815	28,125,718	93.0
	14015	2.10	2.26	33,690,970	31,635,100	93.9
	14022	2.07	1.67	31,521,712	29,784,095	94.5
Sneaker	14001	2.10	2.10	31,262,426	29,109,185	93.1
	14005	2.14	2.06	30,852,163	28,750,108	93.2
	14010	2.10	2.28	31,366,093	29,394,148	93.7
	14014	2.11	1.95	33,214,909	31,279,877	94.2

Table S2. Functionally enriched Gene ontology terms for sex-biased genes in each expression quartile compared to background all genes (FDR < 0.05)

GO term	Description	Sex bias	Expression quartile	p-value (FDR)
0004872	Receptor activity	Female-biased	1st	0.0002
0060089	Molecular transducer activity	Female-biased	1st	0.0001
0005102	Receptor binding	Female-biased	1st	0.002
0038023	Signalling receptor activity	Female-biased	1st	0.002
0004888	Transmembrane signalling receptor activity	Female-biased	1st	0.001
0099600	Transmembrane receptor activity	Female-biased	1st	0.001
0004871	Signal transducer activity	Female-biased	1st	0.006
0043565	Sequence-specific DNA binding	Female-biased	1st	0.018
0005506	Iron ion binding	Female-biased	4th	0.008
0003777	Microtubule motor activity	Male-biased	4th	0.08

Table S3. Rates of evolution (d_N/d_S) for morph-biased contigs.

Expression class	Total contigs	Orthologs [a]	Filter[b]	d_N (95% CI) significance[c]	d_S (95% CI) significance[c]	d_N/d_S (95% CI) significance[c]
Nesting male-biased	41	18	17	0.031 (0.025-0.038) *n.s.*	0.332 (0.296-0.374) *n.s.*	**0.094 (0.074-0.117)** ***P = 0.044***
Satellite male-biased	11	3	2	0.007 (0.003-0.008) *n.s*	0.287 (0.278-0.292) *n.s*.	0.023 (0.012-0.029) *n.s*
Sneaker male-biased	9	6	6	0.036 (0.023-0.041) *n.s*	0.330 (0.244-0.413) *n.s*	0.110 (0.082-0.137) *n.s*

[a]Number of contigs that are 1:1:1 orthologs with *X. maculatus* and *G. aculeatus.*
[b]Number of 1:1:1 orthologs after filtering
[c]P-values from permutations tests, 1,000 replicates, comparing morph-biased genes to male-biased genes

Table S4. Polymorphism (p_N/p_S) for morph-biased contigs.

Expression class	Contigs SNPs[a]	p_N	p_S	p_N/p_S
Nesting male-biased	15	12	24	**0.5** **P = 0.0008**
Satellite male-biased	2	0	17	0 P = 0.2525
Sneaker male-biased	5	7	14	**0.5** **P = 0.0097**
Male-biased	1326	901	6530	0.14

P values from Fisher's exact tests between morph-biased genes and male-biased genes.

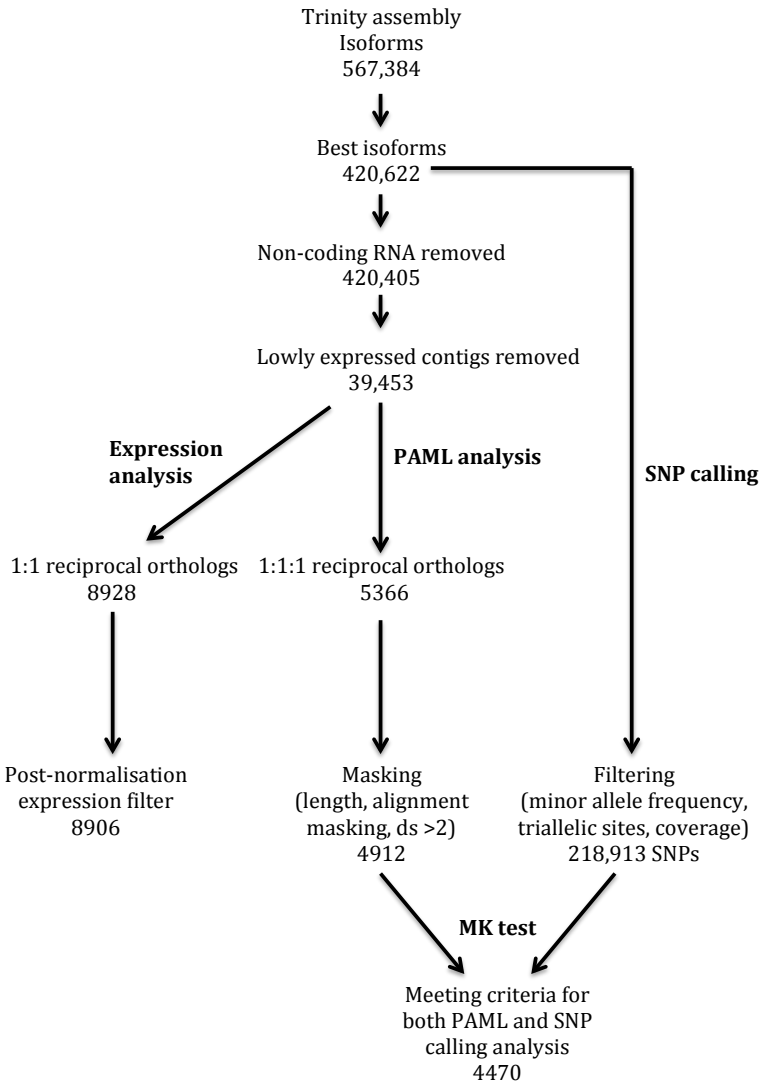


Figure S1. Flowchart of the number of contigs present at each stage of the filtering to produce a dataset for the expression analysis, a dataset for the PAML analysis and a dataset for the McDonald-Kreitman (MK) test.

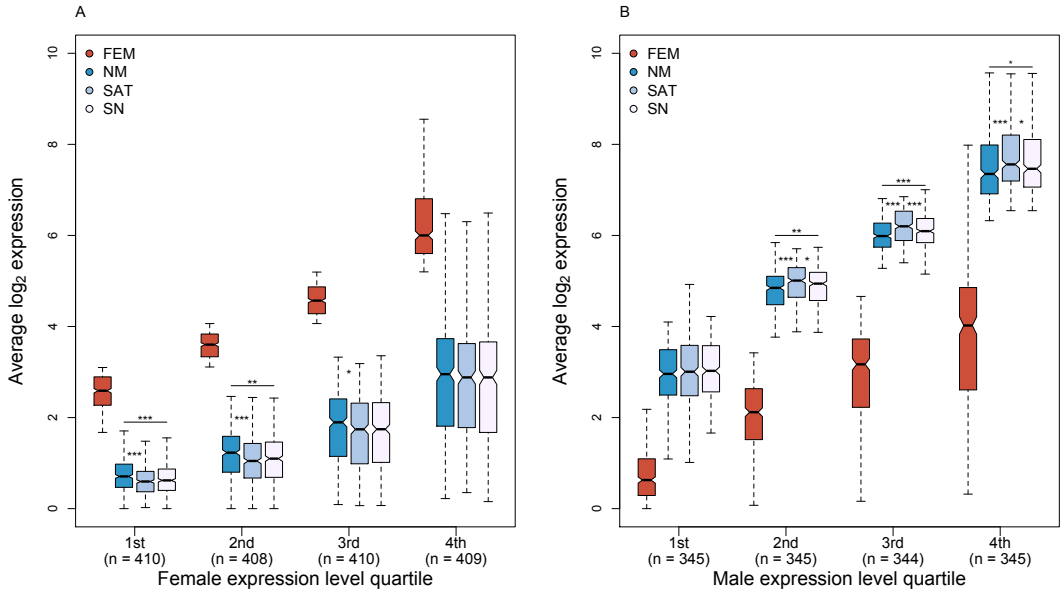


Figure S2. Gene expression for each of the morphs for (A) female-biased genes and (B) male-biased genes using a 4-fold threshold difference in expression between males and females. Data is divided into quartiles based upon expression level in females for panel (A) and males for panel (B). Red = females, dark blue = nesting males, light blue = satellite males and white = sneaker males. Significance is indicated based on Wilcoxon tests (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$).

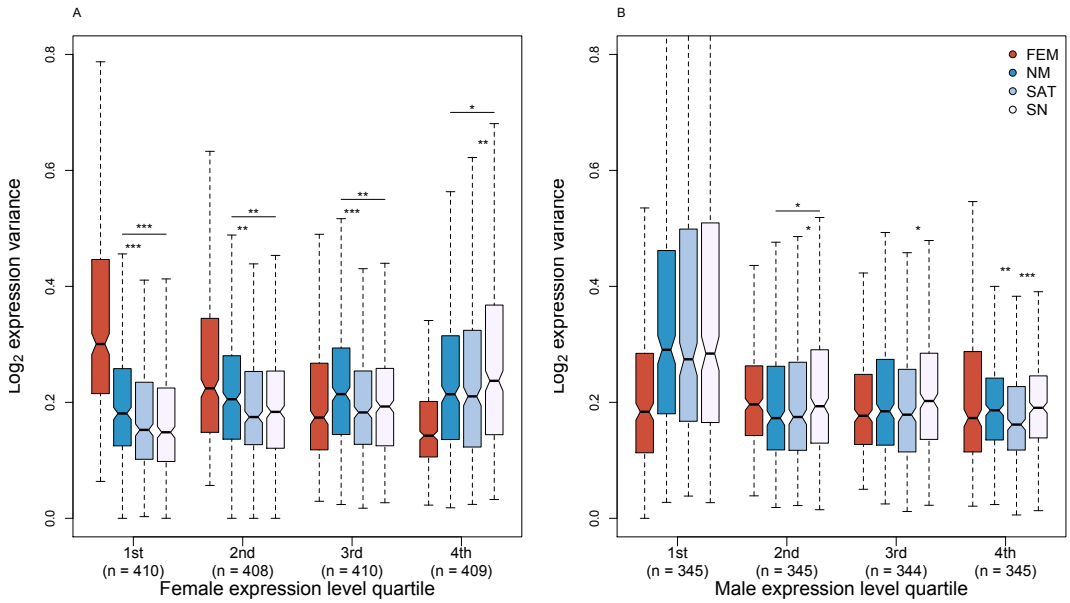


Figure S3. Gene expression variance for each of the morphs for (A) female-biased genes and (B) male-biased genes using a 4-fold threshold difference in expression between males and females. Data is divided into quartiles based upon expression level in females for panel (A) and males for panel (B). Red = females, dark blue = nesting males, light blue = satellite males and white = sneaker males. Significance is indicated based on Wilcoxon tests (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$).

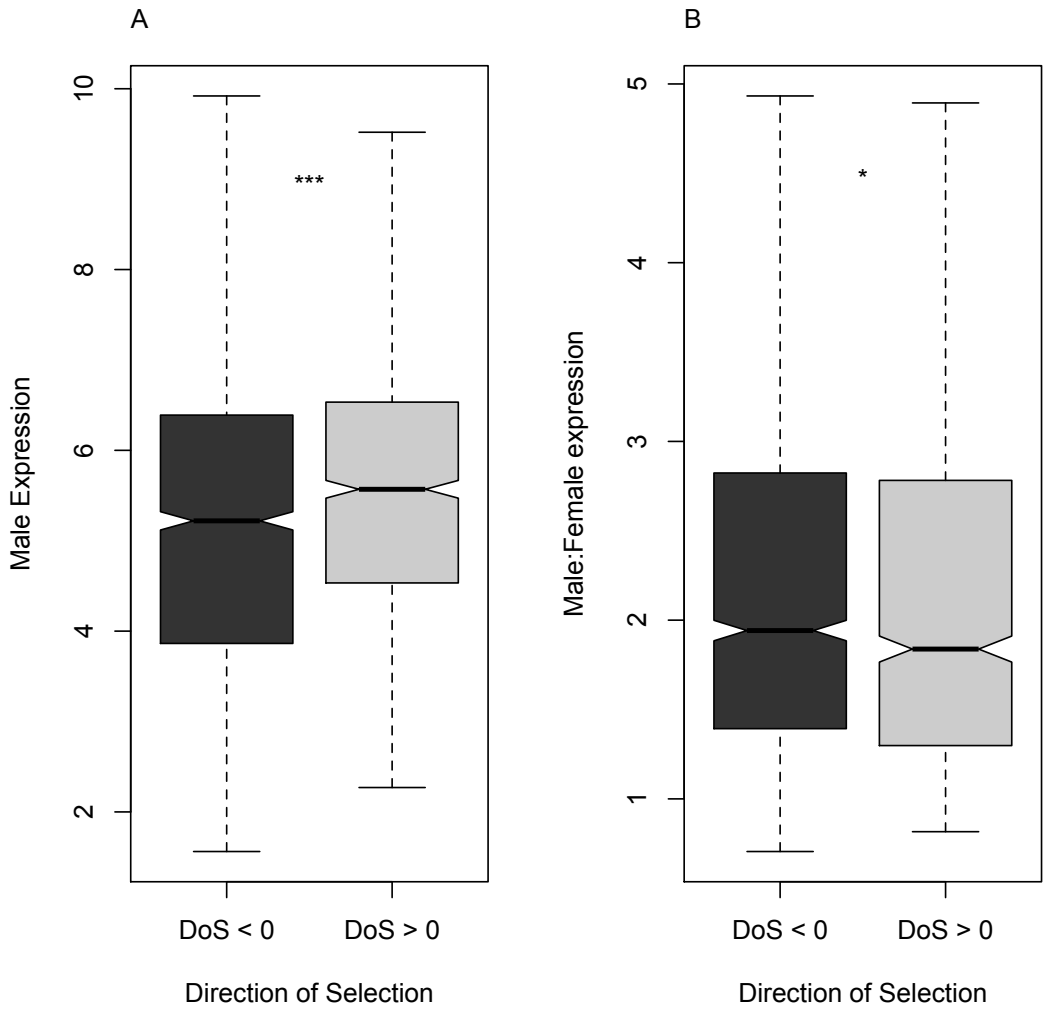


Figure S4 (A) Gene expression level and (B) sex-biased expression of male-biased genes under putative positive selection (DoS > 0, light grey) and under relaxed constraint (DoS < 0, dark grey). Boxes represent medians and first and third quartiles and notches represent approximate 95% confidence intervals. Wilcoxon tests comparing the two gene classes * $P < 0.05$, ***$P < 0.001$.

Supplementary methods

Trim RNA-seq data
Trimmomatic v0.32

java -jar trimmomatic-0.32.jar PE -phred 33 forward.fastq reverse.fastq forward_paired.fastq.gz forward_unpaired.fastq.gz reverse_paired.fastq.gz reverse_unpaired.fastq.gz ILLUMINACLIP:adaptors.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36

De novo transcriptome assembly
Trinity v2.0.2

Trinity --seqType fq --max_memory 10G --left forward.fastq --right reverse.fastq --CPU 12 --normalize_reads --output trinity_output_directory --bflyHeapSpaceMax 16G

Map RNA-seq data and quantify expression for expression analysis
RSEM v1.2.19 and Bowtie2 v2.2.4

extract-transcript-to-gene-map-from-trinity reference.fa Trinity_transcript_gene_map

rsem-prepare-reference --bowtie2 --transcript-to-gene-map Trinity_transcript_gene_map reference.fa Trinity_reference

For each sample:

rsem-calculate-expression -p 12 --bowtie2 --paired-end Sample1_forward.fastq Sample1_reverse.fastq Trinity_reference Sample1

These exact steps were repeated on the final set of best isoforms chosen on the basis of contig expression and length.

Map RNA-seq data for polymorphism analysis
STAR v2.4.2a

STAR --runMode genomeGenerate --genomeDir indexdirectory1 --genomeFastaFiles reference.fa --runThreadN --limitGenomeGenerateRAM

For each sample:

STAR --genomeDir indexdirectory1 --readFilesIn forward.fastq reverse.fastq --runThreadN --outFilterMultimapNmax 1

Run second index, include SJ.out.tab.Pass1.sjdb for samples 1 to n:

STAR --runMode genomeGenerate --genomeDir indexdirectory2/ --genomeFastaFiles reference.fa --sjdbFileChrStartEnd Sample1/SJ.out.tab.Pass1.sjdb Sample2/SJ.out.tab.Pass1.sjdb /Sample-n/SJ.out.tab.Pass1.sjdb --sjdbOverhang 99 --runThreadN 4 --limitGenomeGenerateRAM 800000000000

STAR --genomeDir indexdirectory2 --readFilesIn forward.fastq reverse.fastq --runThreadN --outFilterMultimapNmax 1

Identify SNPs
SAMTOOLS v0.1.19 and VARSCAN2 v2.3.9

samtools faidx reference.fa

For each sample:

samtools view -bS file.sam | samtools sort - file_sorted

samtools mpileup -Bd 10000000 -f reference.fa sample1_sorted.bam sample2_sorted.bam ... > allsamples.mpileup

varscan mpileup2snp allsamples.mpileup.vcf --min-coverage 2 --min-reads2 2 --min-avg-qual 20 --min-var-freq 0.01 --min-freq-for-hom 0.85 --p-value 1 --strand-filter 1 --output-vcf 1 --vcf-sample-list vcf-sample-list.txt > allsamples.mpileup2snp.vcf

Sequence divergence analysis
PRANK v140603 and SWAMP v.09 and PAML v4.8

python SWAMP.py -i input_folder - branchcodes_all.txt -m 100 -t 7 -w 15 > swampoutput