1 **Mitonuclear sex determination? Empirical evidence from bivalves**

2 Chase H. Smith[1,*], Raquel Mejia-Trujillo[1], Sophie Breton[2], Brendan J. Pinto[3-5], Mark

3 Kirkpatrick[1,**], Justin C. Havird[1,**]

4 [1]Department of Integrative Biology, University of Texas at Austin, Austin, TX, USA

5 [2]Department of Biological Sciences, University of Montreal, Montreal, Canada

6 [3]School of Life Sciences, Arizona State University, Tempe, AZ USA

7 [4]Center for Evolution and Medicine, Arizona State University, Tempe, AZ USA

8 [5]Department of Zoology, Milwaukee Public Museum, Milwaukee, WI USA

9 [*]Corresponding Author: chase.smith@austin.utexas.edu

10 [**]Authors contributed equally

Abstract

Genetic elements encoded in nuclear DNA determine the sex of an individual in many animals. In bivalves, however, mitochondrial DNA (mtDNA) has been hypothesized to contribute to sex determination in lineages that possess doubly uniparental inheritance (DUI). In these cases, females transmit a female mtDNA (F mtDNA) to all offspring, while male mtDNA (M mtDNA) is transmitted only from fathers to sons. Because M mtDNA is inherited in the same way as Y chromosomes, it has been hypothesized that mtDNA may be responsible for sex determination. However, the role of mitochondrial and nuclear genes in sex determination has yet to be validated in DUI bivalves. In this study, we used DNA, RNA, and mitochondrial short non-coding RNA (sncRNA) sequencing to explore the role of mitochondrial and nuclear elements in the sexual development pathway of the freshwater mussel *Potamilus streckersoni* (Bivalvia: Unionida). We found that the M mtDNA shed a sncRNA partially within a male-specific mitochondrial gene that targeted pathways hypothesized to be involved in female development and mitophagy. RNA-seq confirmed the gene target was significantly upregulated in females, supporting a direct role of mitochondrial sncRNAs in gene silencing. These findings support the hypothesis that M mtDNA inhibits female development. Genome-wide patterns of genetic differentiation and heterozygosity did not support a nuclear sex determining region, although we cannot reject that nuclear factors are involved with sex determination. Our results provide further evidence that mitochondrial loci contribute to diverse, non-respiratory functions and provide a first glimpse into an unorthodox sex determining system.

Keywords: Sexual development, mitonuclear interactions, doubly uniparental inheritance, Unionida, mt-sncRNAs

33   **Introduction**

34   In many animals, sex is determined by genetic elements encoded in nuclear DNA, and

35   mitochondrial DNA (mtDNA) has not been demonstrated to play a role in sex determination

36   (Bachtrog et al. 2014). However, mechanisms underlying sex determination are highly variable

37   across the tree of life and have been shown to include both mitochondrial and nuclear gene

38   products in plants (Hanson and Bentolila 2004). mtDNA has also been hypothesized to play a

39   role in sex determination or sexual development in bivalves that possess a unique mitochondrial

40   biology known as doubly uniparental inheritance (DUI) (Breton et al. 2011; Breton et al. 2022).

41   Sex determining pathways in bivalves with DUI have been of interest to researchers since

42   its discovery in the early 1990's (Hoeh et al. 1991). Doubly uniparental inheritance involves the

43   biparental transmission of mtDNA, one passed by females to all offspring and a second

44   transmitted by males to only male offspring (Skibinski et al. 1994) (Fig. 1). Females typically

45   only possess female-transmitted mtDNA (F mtDNA), while males are often globally

46   heteroplasmic in somatic tissues but male-transmitted mtDNA (M mtDNA) is localized in

47   gonads and exclusively possessed in sperm (Breton et al. 2017; Ghiselli et al. 2019). Given M

48   mtDNA are inherited in the same way as Y chromosomes and associated with maleness, M

49   mtDNA has been hypothesized to trigger development of male phenotypes, which is

50   hypothesized to be suppressed when M mtDNA is degraded and only F mtDNA remains (Breton

51   et al., 2007).
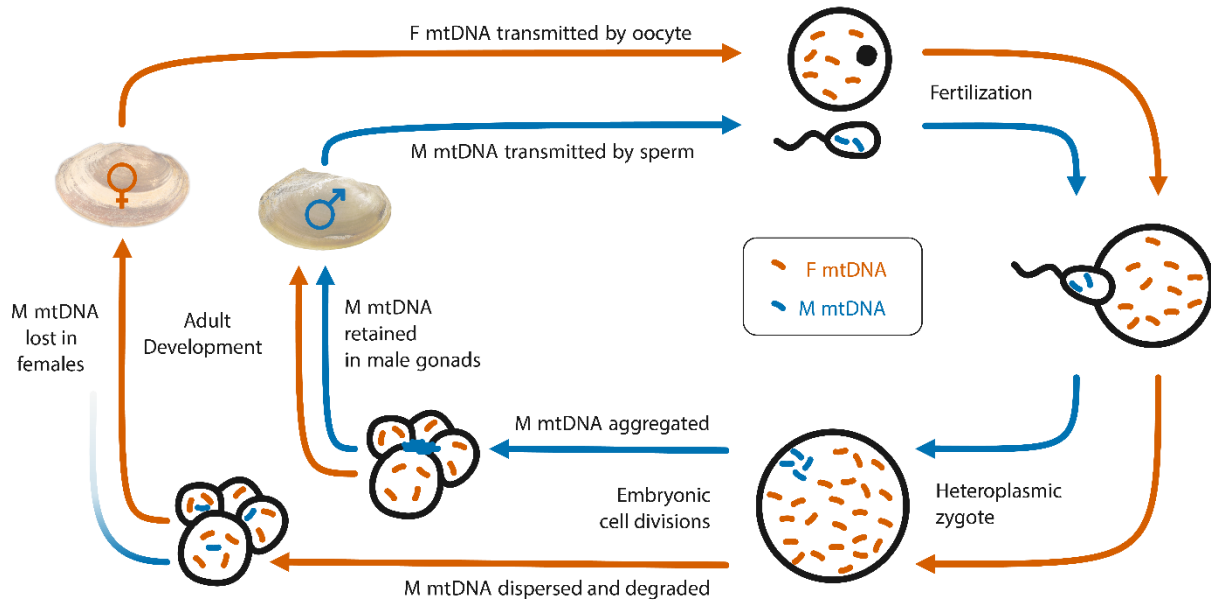
3

Figure 1. Overview of doubly uniparental inheritance of mitochondria in bivalves. Orange coloration represents the female-transmitted (F) mitochondrial DNA (mtDNA) and blue male-transmitted (M) mtDNA. Adapted from Breton et al. (2018).

In freshwater mussels (Bivalvia: Unionida), the evolutionary conservation of DUI and unique sex-specific mitochondrial genes has led to the hypothesis that mtDNAs could play a direct role in sex determination (Breton et al. 2011; Breton et al. 2022). Unlike most other DUI lineages where F and M mtDNA divergence is often ~15% or less, the F and M mtDNAs of freshwater mussels are more than 50% divergent in their amino acid sequences and have remained evolutionarily distinct for over 200 million years (Breton et al. 2007; Doucet-Beaupré et al. 2010; Smith et al. 2023). Further, both genomes contain open reading frames (ORFs; termed female ORF [F-ORF] and male ORF [M-ORF]) with no known homology that have been hypothesized to be either primary sex determining gene(s) or contribute to sexual development (Breton et al. 2011). This mtDNA sex determination hypothesis has been further supported by evolutionary transitions from dioecy to hermaphroditism leading to divergent evolution in F-ORF peptide sequence of hermaphrodites relative to other mitochondrially encoded proteins (Breton et al. 2011; Mitchell et al. 2016; Guerra et al. 2019). Because of these characteristics,

4

69   freshwater mussels have become an ideal model for investigating mitochondrial biology broadly

70   and sex determination hypotheses for DUI species and bivalves in general.

71   We highlight two specific hypotheses regarding the retention of M mtDNA and the role

72   of mtDNAs in the sexual development of DUI species: (1) In the egg factor hypothesis, sex is

73   determined by an egg factor contributed by the mother (Kenchington et al. 2009; Ghiselli et al.

74   2012; Milani et al. 2013; Zouros and Rodakis 2019; Zouros 2020). The egg factor is an inherited

75   genotype at a nuclear gene (or possibly multiple genes) that triggers sexual development,

76   including the retention of the M mtDNA and its localization in male gonadal cells. This

77   hypothesis generally follows expectations under a traditional XY or ZW sex determination

78   system and does not account for the potential role that mitochondrial-encoded elements may play

79   in sexual development. (2) In the cytoplasmic male sterility (CMS) hypothesis, which is

80   primarily based on evidence from freshwater mussels, mtDNAs play a direct role in sexual

81   development (Breton et al. 2022). In the model, the F-ORF acts as a feminizer (inhibiting

82   maleness) and the M-ORF antagonizes F-ORF in some way, acting as a 'restorer of maleness',

83   leading to downstream interactions with nuclear gene products to trigger development of male

84   phenotypes. These two hypotheses provide a foundation for future research to test if one or both

85   hypothetical pathways are involved in sexual development in DUI bivalves.

86   The CMS hypothesis provides an explanation for the evolutionary conservation of F-ORF

87   and M-ORF across freshwater mussels, but it remains uncertain whether these genes encode

88   proteins that trigger sexual development, if they interact with nuclear gene products, or if other

89   regulatory elements encoded in mtDNAs interact with the nuclear genome. In the nuclear

90   genome, short non-coding RNAs (sncRNAs) are critical factors that regulate nuclear gene

91   expression typically through RNA interference (RNAi), a process in which sncRNA binding

5

92   blocks translation of target messenger RNA (Ambros 2004). sncRNAs shed by mtDNA (mt-

93   sncRNAs) have been identified and confirmed to alter nuclear gene expression (Pozzi and

94   Dowling 2021; Pozzi and Dowling 2022), including in DUI species (Pozzi et al. 2017;

95   Passamonti et al. 2020). However, mt-sncRNA validation has yet to be performed in freshwater

96   mussels and their contribution to sexual development, therefore, remains uncertain.

97   In this study, we set out to explore the role of mitochondrial and nuclear elements in the

98   sexual development of the freshwater mussel *Potamilus streckersoni*. Specifically, we use whole

99   genome resequencing, female and male gonad RNA sequencing, and female and male gonad mt-

100   sncRNA sequencing in attempts to identify a candidate SDR within the nuclear genome and

101   mitochondrial-encoded factors contributing to sexual development. We then synthesize our

102   results with previous sex determination hypotheses presented in the literature. Our findings

103   provide further support for an unusual sex determination system in DUI bivalves dependent on

104   mitonuclear interactions.

105

106   **Materials and Methods**

107   *DNA extraction and sequencing*

108   To investigate nuclear genes that may be involved with sex determination, we collected 22 adult

109   individuals of *P. streckersoni* from multiple localities in the Brazos River drainage in Texas.

110   This population has been shown to lack population structure based on genomic data (Smith et al.

111   2021). Sex was distinguished using shell characters (adult females typically have blunt posterior

112   shell margins) and external gill morphology (adult females have distinct serrated gill margins).

113   DNA was extracted from fresh mantle tissue using the Qiagen PureGene Kit (Hilden, Germany)

114   with standard protocols. High-molecular weight genomic DNA was confirmed from these

6

115   extractions by visualizing each isolation on a 1% agarose gel stained with GelGreen® nucleic

116   acid stain (Biotium, Hayward, CA, USA). Isolation quantity and quality was assessed using a

117   Qubit™ fluorometer and a NanoDrop™ One (ThermoFisher Scientific; Waltham, MA, USA),

118   respectively.

119   Whole genome resequencing (WGR) was performed on 11 female and 11 male

120   individuals of *P. streckersoni*. Libraries were prepared by the Texas A&M AgriLife Genomics

121   and Bioinformatics Service (College Station, TX, USA) from ~25 ng of genomic DNA in a

122   custom, automated, and miniaturized version of the PerkinElmer NEXTFLEX Rapid XP kit

123   protocol (Johnson et al. 2019). Briefly, genomic DNA was enzymatically fragmented for 6

124   minutes, ligated to unique dual-indexed barcodes, size selected between 520–720 bp using

125   SPRIselect beads (Beckman Coulter; Brea, CA, USA), and amplified with 10 PCR cycles. One

126   tenth of the manufacturer's prescribed volumes were used for enzymatic steps. Libraries were

127   diluted with elution buffer to a final concentration of 2 ng/μl and were pooled by equivolume.

128   The pool was sequenced on an Illumina NovaSeq S4 XP (San Diego, CA, USA) using 2 X 150

129   bp reads.

130

131   *DNA Analyses*

132   We used pool and individual based single nucleotide polymorphism (SNP) approaches to further

133   investigate a potential nuclear SDR in *P. streckersoni*. Raw reads were trimmed using TRIM

134   GALORE! V 0.6.7 (www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with default

135   settings and data quality was verified in FastQC v 0.11.9

136   (www.bioinformatics.babraham.ac.uk/projects/fastqc/). Trimmed reads were mapped to the *P.*

137   *streckersoni* nuclear genome assembly (Smith 2021) using bwa-mem v 2.2.1 (Md et al. 2019)

138   with default parameters. PCR duplicates were removed using picard v 2.27.4 (Broad Institute

139   2019). The resulting bam files were combined by sex and the Pooled Sequencing Analyses for

140   Sex Signal (PSASS) pipeline v 3.1.0 (Feron and Jaron 2021) was used to call SNPs and calculate

141   genotypic statistics. Pileup was used to call SNPs with a minimum mapping quality of 30. We

142   calculated sex-specific alleles within gene models using PSASS. We then calculated F_{ST} between

143   the sexes using 1 kb windows and considered windows with values greater than or equal to 0.1 as

144   elevated. Windows with at least an average of 3x coverage per individual per sex (33x) were

145   retained for the analysis.

146   Pool-based methods failed to identify large, contiguous areas of differentiation between

147   the sexes. To identify smaller regions of the genome with alleles consistent with XY or ZW sex

148   determination (XY- or ZW-like alleles), we investigated patterns of heterozygosity, which have

149   been used previously to identify SDRs in some bivalves (Han et al. 2022). First, we downloaded

150   high coverage Illumina reads for the individual used to generate the reference assembly from the

151   GenBank SRA (accession SRR13176629). We then recalled SNPs including all 22 WGR

152   samples and the reference individual using FreeBayes v 1.3.6 (Garrison and Marth 2012).

153   Variants were filtered to only include Q20 biallelic sites, and singletons were removed using

154   vcftools v 0.1.16 (Danecek et al. 2011). We then followed similar methods as Kirkpatrick et al.

155   (2022) to identify ZW or XY-like alleles. For ZW-like SNPs, the script required all males to be

156   homozygous for the same allele and females heterozygous with one copy of the same allele

157   found in males. For XY-like SNPs, we used the inverse. Although SNPs following this pattern

158   could occur across the genome by chance, they are expected to be enriched in SDRs. For our

159   analysis, we required called genotypes from at least 3 males and 3 females at each candidate site.

160   We used a modified version of the R script provided in Kirkpatrick et al. (2022) to perform the

161   analysis.

162

163   *RNA extraction and sequencing*

164   We extracted total RNA from fresh or preserved (RNAlater) gonadal tissue of 6 female and 3

165   male *P. streckersoni* to infer differential gene expression among nuclear genes. To improve the

166   existing genome annotation (Smith 2021), we also extracted RNA from six tissue types from an

167   adult male *P. streckersoni*: adductor, foot, gill, gonad, mantle, and stomach. RNA was extracted

168   from each tissue type independently and pooled into a single sample with equal representation

169   from RNA samples. All RNA was extracted using the RNeasy kit following the manufacturer's

170   protocol (Qiagen). RNA quality and integrity were determined using a NanoDrop™ and an

171   Agilent Bioanalyzer (Santa Clara, CA, USA), respectively. Messenger RNA was isolated from

172   150 ng of total RNA using a Nextflex Poly-A Selection kit (Perkin Elmer; Waltham, MA, USA).

173   cDNA libraries were prepared using a Nextflex Rapid Directional RNA 2.0 kit, miniaturized to

174   2/5 reaction volume and automated on a Sciclone NGSx liquid handler. All libraries were

175   sequenced by the Texas A&M AgriLife Genomics and Bioinformatics Service on an Illumina

176   NovaSeq S4 XP using 2 X 150 bp reads.

177

178   Potamilus streckersoni *genome annotation*

179   We used all novel RNA-Seq reads, as well as female and larval pools published previously (SRA

180   accession numbers SRR13176627 and SRR13176628), to improve the existing genome

181   annotation for *P. streckersoni* (Smith 2021). Structural and function annotation was performed

182   using the Funannotate pipeline v 1.8.10 (Palmer and Stajich 2017). Prior to running the pipeline,

9

183   repeats in the *P. streckersoni* genome assembly were identified and masked using

184   RepeatModeler v 2.0.1 (Flynn et al. 2020) and RepeatMasker v 4.0.9 (Smit et al. 2015),

185   respectively. Completeness of the previous and updated annotation was assessed using BUSCO v

186   5.4.3 (Manni et al. 2021) with the metazoan (v 10; 954 genes) and molluscan lineages (v 10;

187   5295 genes).

188

189   *RNA-seq analyses*

190   Gonadal RNA-Seq reads were trimmed using TRIM GALORE! with default parameters. Data

191   quality was verified using FastQC. Trimmed reads were then used to test for significant

192   differences in gene expression between sexes using the updated genome annotation. Reads were

193   mapped to the reference genome using Hisat2 v 2.2.1 (Kim et al. 2019) with default parameters.

194   Gene counts were summarized using the command "featureCounts" in the R package Rsubread v

195   2.10.5 (Liao et al. 2019). A differential gene expression analysis between males and females was

196   performed on the counts using the command "DESeq" in the R package DESeq2 v 1.36.0 (Love

197   et al. 2014) with default parameters. Genes with adjusted p-values < 0.05 were considered

198   differentially expressed genes (DEGs) among sexes. We then joined expression data for genes

199   that fell within windows of high F_{ST} based on DNA analyses and tested whether genes that fell

200   within windows of high F_{ST} were more likely to be differentially expressed using a chi-squared

201   test.

202   To infer co-expression among genes, we performed a weighted gene co-expression

203   network analysis (WGCNA) in the R package WGCNA v.1.72-1 (Langfelder and Horvath

204   2008). Before the analysis, we filtered out genes with a coefficient of variation > 200 and counts

205   per million value < 5 from the raw count matrix using the "filtered.data" command in the R

206 package NOISeq v 2.40.0 (Tarazona et al. 2015). The count data was then transformed using the

207 "vst" command in DESeq2. Once the co-expression network was constructed following

208 developers' documentation, we used eigengenes as representatives of each module to investigate

209 intermodular correlations. Additionally, we included binary traits (female and male) to the

210 eigengene network to reveal co-expression relationships across the sexes. We identified genes

211 with high intramodular connectivity (*i.e.*, hub genes) using the "intramodularConnectivity"

212 function from the WGCNA package.

213   We further investigated potential pathways that were differentially expressed among

214 female and male gonads using pre-ranked gene set enrichment analyses in the R package fgsea v

215 1.22.0 (Korotkevich et al. 2021). We used the hallmark (Liberzon et al. 2015) and GO Biological

216 Process gene sets from the Human Molecular Signatures Database v 2022.1.Hs (MSigDB)

217 (Subramanian et al. 2005) for the analyses. Gene names from functional annotations from

218 funannotate were edited to match human gene symbols from entries in MSigDB, while genes

219 without annotations were removed. Z-statistics from DESeq2 were used to rank genes.

220 Additionally, the Z-scores of duplicated genes were averaged such that each unique gene symbol

221 had one value contributing towards its rank. Pre-ranked gene set enrichment analyses were then

222 performed on each gene set using the command "fgseaMultilevel" in fgsea with a maximal gene

223 set size of 50. We considered pathways with an adjusted p-value < 0.05 as having significantly

224 different expression among female and male gonadal tissue.

225

226 *Mitochondrial sncRNA sequencing, validation, and target prediction*

227 High throughput sncRNA-seq was performed on gonadal tissue from 3 females and 3 males,

228 which included both new individuals and a subset of individuals used for RNA-Seq and DEG

11

229   analyses. Before sncRNA extraction, mitochondrial enrichment was performed on gonadal tissue

230   samples using a modified protocol described in Ballantyne & Moon (1985). Briefly, this process

231   involved an initial homogenization in ice-cold isolation buffer (480 mM sucrose, 100 mM KCl,

232   50 mM NaCl, 70 mM HEPES, 6 mM EGTA, 3 mM EDTA, and 1% BSA; pH 7.6) followed by

233   two centrifugation steps: 1) homogenates centrifuged at 4°C at 2,790 X g for 5 minutes, and 2)

234   supernatants centrifuged 4°C at 12,200 X g for 15 minutes. We used the Purelink miRNA

235   extraction kit (Invitrogen) with the standard protocol to extract sncRNA from pellets enriched

236   with mitochondria. Libraries were prepared for each sample at the University of Texas Genomic

237   Sequencing Facility using the NEBNext Small RNA library preparation kit (New England

238   Biolabs; Ipswich, MA, USA). Libraries were enriched with 14 cycles of PCR and the final

239   product was size selected for 105–155 bp using a 3% gel cassette on the Blue Pippin (Sage

240   Science; Beverly, MA, USA). The final size selected libraries were assessed for library quality

241   and quantity using a BioAnalyzer. Libraries were subsequently sequenced on an Illumina

242   NovaSeq 6000 using 1 X 75 or 1 X 100 bp reads.

243   Data processing generally followed previous studies identifying mt-sncRNAs (Pozzi et

244   al. 2017; Passamonti et al. 2020). Raw reads were trimmed using TRIM GALORE! while

245   enforcing a minimum read length of 15 bp and a maximum of 40 bp. Trimmed reads were then

246   mapped to reference female (GenBank: MW413895) and male (GenBank: ON881148)

247   mitochondrial genome assemblies for *P. streckersoni* (Smith 2021; Mejia-Trujillo and Smith

248   2022) using bowtie v 2.5.0 (Langmead and Salzberg 2012) with the parameters "-N 1 -i C,1 -L

249   18 --end-to-end". Mapped reads were then clustered using the UCLUST algorithm in USEARCH

250   V 11.0.667 (Edgar 2010) using an identity filter of 0.99. Clusters with centroids less than 17 bp

251   were removed. Centroids that followed the following criteria were considered mitochondrial

252   sncRNAs: (1) a cluster size greater than 200; (2) a perfect match of nucleotides 4–10 and a

253   minimum of 11 matches with a 3' UTR of a nuclear-encoded gene as determined by BLAST +

254   2.6.0105 (Camacho et al. 2009) using the options "-task blastn-short -strand minus"; and (3) a

255   $\Delta\Delta G$ score lower than −9 kJ for the centroid-target UTR interaction as determined by RNAup v

256   2.5.1 (Lorenz et al. 2011) using a temperature of 37 °C, (4) a Gibbs free energy score lower than

257   −20 kJ for the centroid-mRNA duplex as determined by RNAhybrid v 2.1.2 (Krüger and

258   Rehmsmeier 2006), and (5) at least a 1.5 fold difference in coverage at the 5' and 3' ends when

259   compared to the average value for 5 bp upstream and downstream as determined by SAMtools v

260   1.6 (Danecek et al. 2021). In cases where targets were annotated as hypothetical proteins, we

261   annotated genes (if able) by blasting the peptide sequence against the Kyoto Encyclopedia of

262   Genes and Genomes (KEGG) (Kanehisa et al. 2023). We then examined expression data for mt-

263   sncRNA targets and tested whether targeted genes were significantly more likely to be

264   differentially expressed than those that were not using a chi-squared test.

265

266   *Gene network analysis and protein-protein interaction prediction*

267   Our results identified a particularly interesting M mt-sncRNA at the 5' end of the M-ORF

268   targeting the nuclear-encoded gene *GCNT1*, prompting us to use STRING v 11.5 (Szklarczyk et

269   al. 2019) to investigate the gene interaction network of *GCNT1*. For the network, we only

270   considered validated interactions based on the following settings in STRING: 1) supported by

271   STRING databases, 2) supported by experimental evidence, 3) and predicted interactions with a

272   confidence score ≥ 0.15. We then visualized gene expression among genes in the network based

273   on expression data from DESeq2.

274   To infer if the F-ORF protein might interact with the *GCNT1* protein, we used

275   AlphaPulldown v 0.22.3 (Yu et al. 2023) and AlphaFold v 2.2.0 (Jumper et al. 2021). This tested

276   the hypothesis that the F-ORF protein is acting as a feminizer by interaction with the *GCNT1*

277   protein, while the M mt-sncRNA is acting as a restorer of maleness through RNAi-mediated

278   gene silencing. To control for random protein-protein interaction (PPI) support due to the small

279   size of the F-ORF protein (88 residues), we ran the same analysis on the F mtDNA copy of

280   ATP8 (72 residues) and the M-ORF (214 residues). Our logic behind this is that short peptide

281   sequences or mitochondrially based proteins would show similar support for PPI if a supported

282   F-ORF interaction was due to chance. We considered a protein docking score greater than or

283   equal to 0.23 as a predicted PPI (Basu and Wallner 2016).

284

285   **Results**

286   *No support for nuclear encoded SDR*

287   Whole genome resequencing performed on 11 male and 11 female individuals of *P. streckersoni*

288   generated approximately ~7x coverage per sample. Additional statistics regarding samples,

289   sequencing, voucher numbers, and accession numbers can be found in Table S1. Sliding window

290   statistics failed to identify large, contiguous areas of sequence or read depth differentiation

291   between the sexes (File S1), which would be expected if sex chromosomes or large, continuous

292   regions were responsible for sex determination. However, we identified 5629 1-kb windows with

293   elevated F_{ST} (0.1) across 880 scaffolds (roughly ⅓ of total scaffolds) (File S2). This included

294   2439 windows falling within 1346 genes (File S2). Genome-wide patterns of heterozygosity did

295   not identify any regions that were enriched with XY- or ZW-like SNPs, with only two ZW-like

296   SNPs (in two genes on two different scaffolds) and one XY-like SNPs (not within a gene)

14

297   identified by our analysis. All SNPs were only present in areas of low coverage (no more than

298   five individuals of either sex) and coinciding sliding windows showed no evidence of high

299   genetic differentiation, which would be expected if the region was a SDR. Therefore, we

300   consider that these SNPs occurred by chance and are most likely not SDRs.
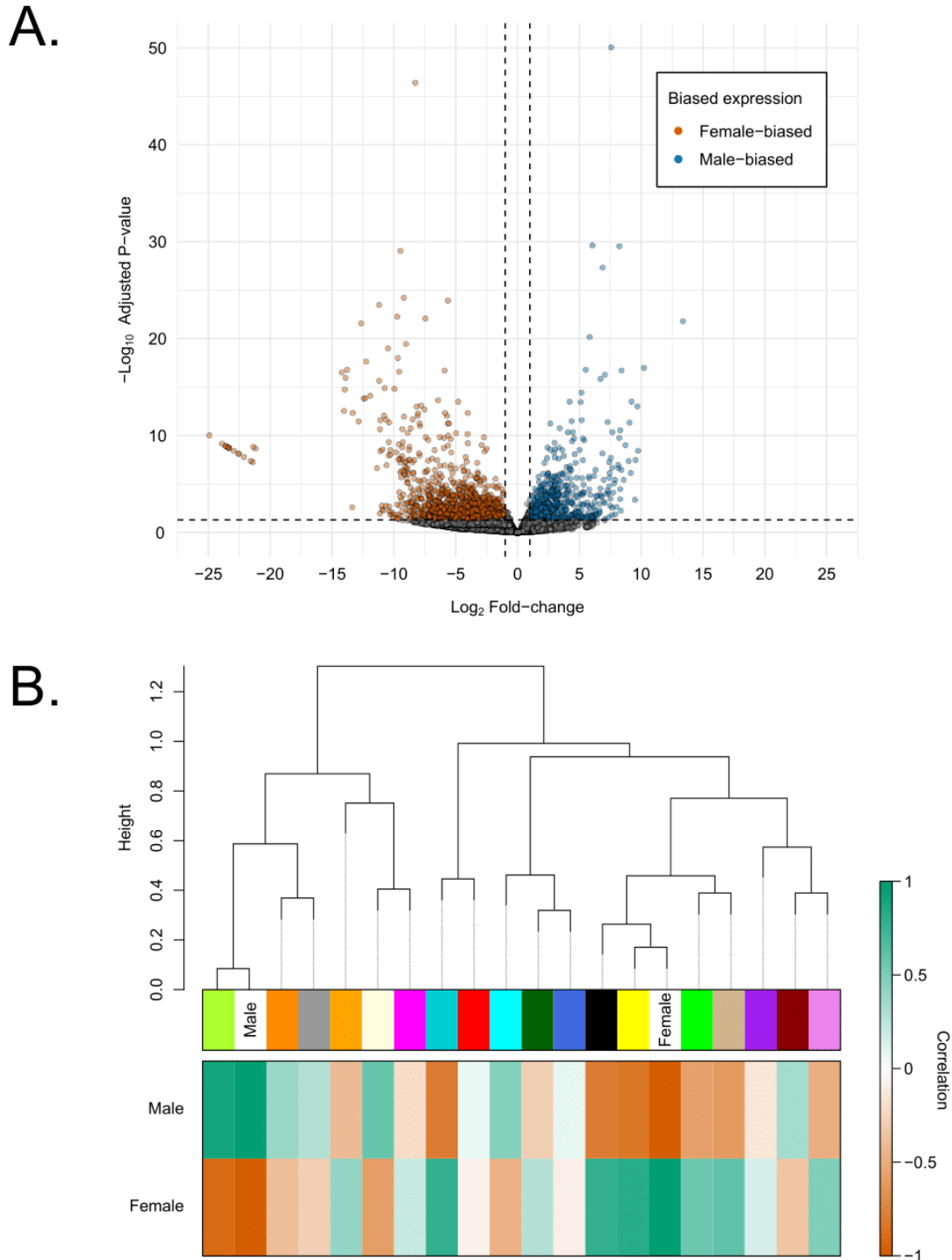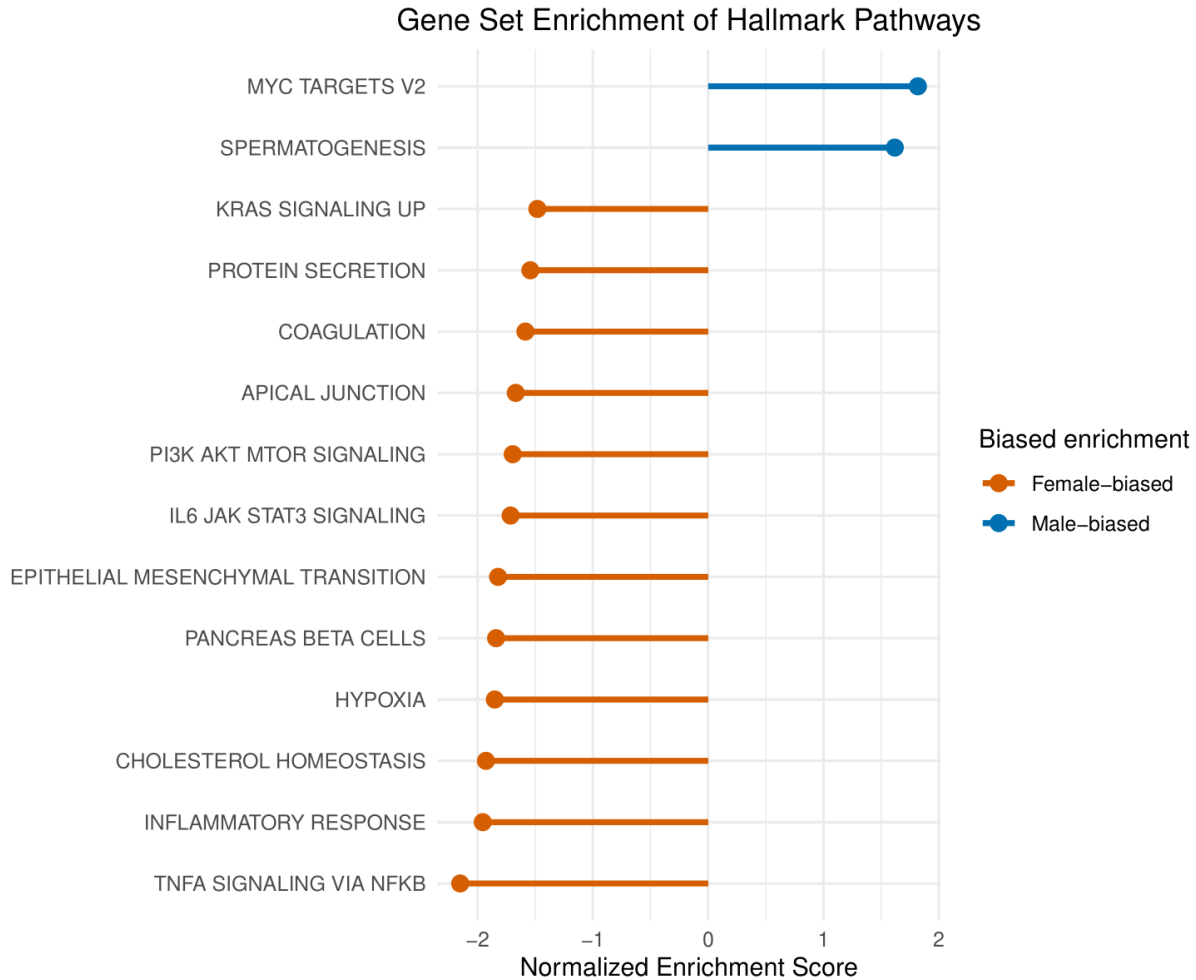
301

302   *Improved genome annotation and differential expression between the sexes*

303   Statistics regarding RNA-seq libraries can be found in Table S2. Functional annotation returned

304   45,268 gene models (34,937 protein-coding genes, 2,075 isoforms, 8,256 tRNAs), reducing the

305   number of protein coding genes from the previous annotation by ~15%. Despite the decreased

306   number of protein coding genes, the annotation had more than a 20% increase in complete

307   BUSCOs using the metazoan and molluscan lineages (up to 95% and 86% complete,

308   respectively, Table S3).

309   DESeq2 identified 2311 DEGs (adjusted p-value < 0.05) between female and male

310   gonadal tissue (Fig. 2A; File S3). Of these DEGs, 141 were also found to have at least 1 window

311   of elevated F_{ST} (0.1) (File S2). A chi-squared test supported the hypothesis that genes with

312   elevated F_{ST} were significantly more likely to be differentially expressed (6.1% vs 3.7%; $p <$

313   0.001), which is coincident with sexual antagonism in female and male gonadal tissues.

314   WGCNA selected 18 modules to best explain expression profiles (Fig. 2B; File S4). The

315   genes with the highest intramodular connectivity for each module (*i.e.*, the gene with the highest

316   level of co-expression with other genes in the module) are reported in File S5. Genes in the top

317   5% of intramodular connectivity were found in three modules: 1) 452 of the 1,620 genes in the

318   orange module, 2), 85 of the 1,140 genes in the violet module, and 3) 49 of the 1,909 genes in

319   the green-yellow module (most closely corresponded to maleness). The yellow module consisted

Figure 2. A) Volcano plot summarizing tests for differential gene expression between female and male gonadal tissue in *Potamilus streckersoni*. Orange dots represent genes found to have significant female-biased expression and blue those with significant male-biased expression. Black dots represent genes without significantly different expression. Dashed vertical lines represent a $\log_2$-fold change of -1 and 1, and dashed horizontal lines represent the -$\log_{10}$ value for adjusted p-value of 0.05. B) Dendrogram of WGCNA clusters based on expression profiles. In the heat plot, coloration represents positive (green) or negative (orange) correlation of each cluster with sex (female or male).

16

Figure 3. Gene set enrichment of Hallmark Pathways. Pathways with negative NES values contain genes with female-biased expression, while pathways with positive NES values contain genes with male-biased expression. All shown pathways showed significant enrichment (p-adj < 0.05) in female (orange) and male (blue) gonadal tissue.

of nearly ⅓ of all annotated genes (11,725 genes) and most closely corresponded to femaleness.

Additionally, 13 pathways were supported as enriched in the hallmark set and 246 pathways in the GOBP set (p < 0.05) (Fig. 3; File S6). Many of these pathways were involved with sex maintenance and not primary sex determination (*e.g.*, spermatogenesis gene enrichment in males).

Female and male mitochondrial sncRNAs inhibit nuclear gene expression

17

340   High throughput sncRNA-seq generated an average of 32 million reads per sample (female: 37

341   million, male: 29 million). Additional details regarding each library can be found in Table S4. In

342   total, 24 F and nine M mt-sncRNAs passed our inclusion criteria. All mt-sncRNAs, their

343   location, and validated targets are reported in Table S5. In total, 91 nuclear targets were

344   identified for the 34 mt-sncRNAs. The mt-sncRNAs were highly enriched, with read counts that

345   are orders of magnitude greater than in mt-sncRNAs previously described from DUI bivalves

346   (Pozzi et al., 2017). We noticed one male sample expressed F mt-sncRNAs at a higher level than

347   the other male samples (Fig. 4). However, this is reasonable and likely due to chance given our

348   sampling design, because male gonadal tissue could include cells containing F mtDNA (only

349   sperm show homoplasmy for M mtDNA) and M mtDNA concentration in gonadal tissue can

350   vary based on individual or seasonality in *P. streckersoni* (Mejia-Trujillo and Smith 2022).

351   We would generally expect nuclear targets of mt-sncRNAs to show sex-biased

352   expression, with M mt-sncRNA targets most likely showing increased expression in females and

353   vice versa for F mt-sncRNAs. However, we found limited expression differences in F and M mt-

354   sncRNA nuclear gene targets (Fig. 4). Of the 96 mt-sncRNA targets identified, only 13 were sex-

355   specific DEGs. Although we would expect all "true" mt-sncRNAs to lead to differential

356   expression, a chi-squared test did support the hypothesis that mt-sncRNA targets were

357   significantly enriched for DEGs (14% vs 6.6%, p = 0.006). Of the five F mt-sncRNA targets that

358   follow the expected pattern of upregulation in males, the most extreme was ~5-fold higher

359   expression in male gonads (*Transient receptor potential melastatin 2* [*TRPM2*]). We also had

360   seven F mt-sncRNA targets that showed female upregulation, with the most extreme having

361   ~141-fold higher expression in female gonads (*Endonuclease Domain Containing 1*). On the

362   other hand, both M mt-sncRNA targets with differential expression were in the predicted

363 direction and we choose to focus on the one with the most extreme sex-specific expression and

364 also the only one that was annotated: a M mt-sncRNA located at the 5' end of the M-ORF, which

365 was predicted to target *Glucosaminyl (N-Acetyl) Transferase 1* (*GCNT1*), which had ~28-fold

366 higher expression in female gonads when compared to male gonads (log$_2$-fold change = 4.8; p =

367 0.03) and in the top 5% of hub genes (Fig. 4; File S5). No M mt-sncRNA targets were

368 upregulated in male gonads.

369

370 *F-ORF may interact with GCNT1 to facilitate female development*

371 Because *GCNT1* showed the predicted patterns of a nuclear target gene under the CMS

372 hypothesis and could therefore potentially play a role in sex differentiation of bivalves, we chose

373 to further explore its potential interacting genes and proteins. The *GCNT1* gene network included

374 18 genes, most of which were *mucin* (*MUC*) and *N-acetylgalactosaminyltransferase* (*GALNT*)

375 genes (Fig. 5). Many genes in the network were upregulated in females, including *GCNT1*,

376 *MUC1* (log$_2$-fold change = 2.10), *MUC4* (log$_2$-fold change = 4.43), a duplicate of *MUC3A*

377 (*MUC3A_2*; log$_2$-fold change = 7), and a hypothetical gene showing homology to *MUC13*

378 (*MUC13*-like; log$_2$-fold change = 6.75) (Fig. 5). We found no acceptable support for PPIs

379 between the GCNT1 protein and any of the mitochondrial proteins (*i.e.,* ATP8, F-ORF, and M-

380 ORF), with the strongest interaction being between F-ORF and GCNT1 (Fig. S1). The potential

381 F-ORF and GCNT1 PPI had a larger protein docking score (0.19) when compared to ATP8

382 (0.03) and the M-ORF (0.02). This is consistent with expectations from the CMS hypothesis but

383 is limited evidence that these proteins interact.

19

384
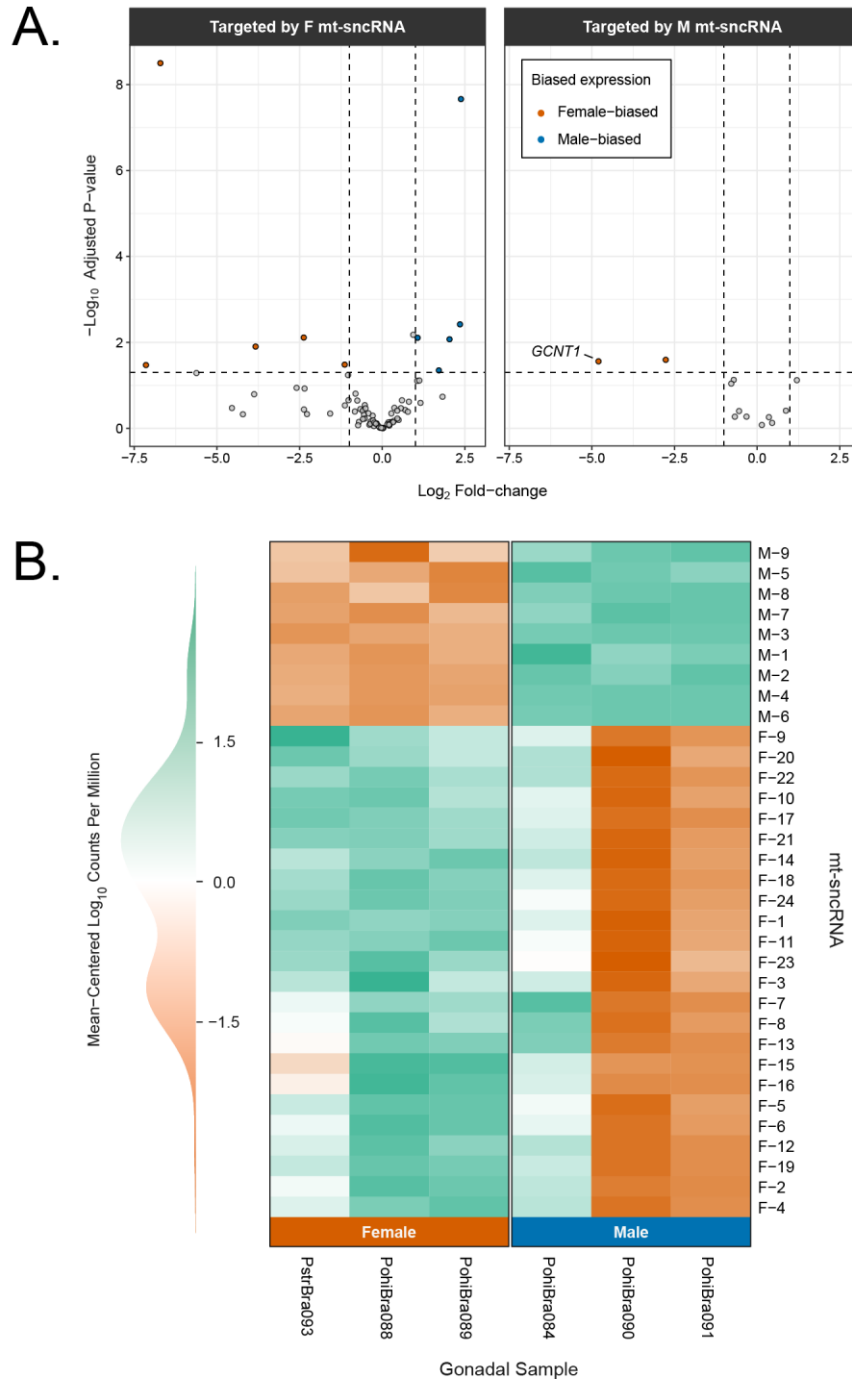
Figure 4. A) Volcano plot summarizing gene expression of genes targeted by female (left) and male mt-sncRNAs
(right). Orange dots represent genes found to have significant female-biased expression and blue those with
significant male-biased expression. All targets with significant differences in expression were hypothetical other
than *GCNT1*. B) Heat map visualizing mt-sncRNAs expression in female (left) and male (right) gonadal tissue.
Green coloration represents higher expression and orange represents lower or no expression.
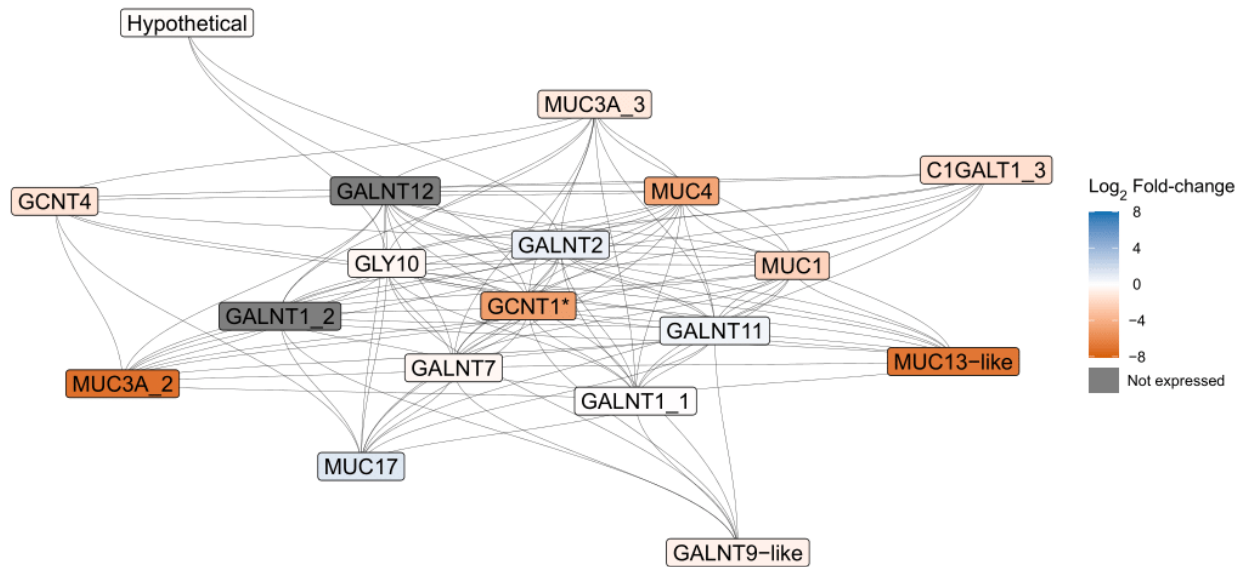
Figure 5. *GCNT1* gene network as predicted by STRING based on the *Potamilus streckersoni* proteome. Gray lines represent known interactions among genes based on database and experimental evidence. Genes are colored based on expression profiles between female (orange) and male (blue) gonadal tissue with color intensity representing differential gene expression between female and male gonadal tissue. Dark gray coloration represents genes with no expression. Asterisks indicated genes with significant differences in expression (p-adj < 0.05). Protein names with the suffix '-like' are hypothetical and annotated as their best KEGG-hit when applicable. Protein names with an underscore represent protein products from genes with multiple annotated copies.

Discussion

Male mt-sncRNAs may inhibit mitophagy and contribute to sexual development in DUI species

In the CMS hypothesis presented by Breton et al. (2022), the F-ORF gene originated to act as a

"feminizer" in an evolutionary transition from hermaphroditism to gynodioecy (*i.e.*, consisting of

hermaphrodites and females). Subsequently, the M-ORF is hypothesized to have originated

during a transition from gynodioecy to dioecy (*i.e.*, consisting of females and males) by

counteracting feminization and leading to the origin of M mtDNA. While this hypothesis

provides an explanation for the origin of DUI, empirical data has not supported that the M-ORF

is responsible for counteracting the F-ORF or if other elements encoded within M mtDNA

contribute to sex determination or sexual development. Our investigation found highly enriched

mt-sncRNAs transcribed from M mtDNA that may play such a role (Fig. 4). Most interestingly,

21

410   we validated a M mt-sncRNA at the 5' end of the M-ORF that targeted *GCNT1*. This gene is in

411   the mucin-type *O*-glycosylation pathway, which has a conserved role in eukaryotic development

412   (reviewed by Tran and Ten Hagen 2013). Our RNA-seq data found *GCNT1* is differentially

413   expressed and highly upregulated in females. This supports the hypothesis that M mtDNA is

414   interacting with the nuclear genome through RNAi-mediated gene silencing and provides a

415   plausible explanation for the evolutionary conservation of M-ORF across freshwater mussels.

416   The exact role that the knockdown of *GCNT1* plays in inhibiting degradation of M

417   mtDNA and female developmental pathways remains uncertain, but our findings suggest two

418   hypothetical functions. First, *GALNT* genes in the *GCNT1* gene network have been hypothesized

419   to trigger female development in scallops based on comparative transcriptomics (Zhou et al.

420   2019), which supports the hypothesis that the M mt-sncRNA could inhibit female developmental

421   pathways in freshwater mussels if it is evolutionarily conserved across bivalves. This hypothesis

422   is further supported from findings in mice, where *GCNT1* has been demonstrated to be

423   upregulated following the knockout of *SOX8* (Singh et al. 2009). These results suggest a non-

424   altered male development pathway keeps *GCNT1* at low levels and provide further support for a

425   role of M mt-sncRNAs in sexual development of bivalves through the inhibition of female

426   developmental pathways.

427   Genes in the *GCNT1* gene network have also been demonstrated to play a role in

428   mitophagy. The gene *MUC1* is localized to the mitochondrial membrane to increase mitophagy

429   of mitochondria with decreased membrane potential (Li et al. 2022). Our results indicated that

430   *MUC1* is somewhat upregulated in female gonads of *P. streckersoni* ($\log_2$-fold change of 2.1),

431   which supports the hypothesis that upregulation of the *GCNT1* gene network in female gametes

432   is involved in mitophagy of male-transmitted mitochondria and provides a plausible explanation

22

433   for the M mt-sncRNA target. This hypothesis does require that male-transmitted mitochondria

434   have a lower membrane potential than female-transmitted mitochondria. Decreased membrane

435   potential has been demonstrated to trigger the degradation of paternal mitochondria post

436   fertilization in multiple model species (Rojansky et al. 2016; Zhou et al. 2016). While sperm

437   mitochondria in some marine DUI species have been demonstrated to exhibit high membrane

438   potential (Milani and Ghiselli 2015), mitochondrial performance and function is certainly altered

439   in DUI sperm vs. eggs (Bettinazzi et al. 2019; Bettinazzi et al. 2020) and M mtDNA OXPHOS

440   genes have been suggested to be under relaxed selection (Maeda et al. 2021), suggesting that

441   altered mitochondrial function in sperm could target M mtDNA for degradation. Studies of F vs.

442   M mtDNA function have not been performed in freshwater mussels and are necessary to

443   determine if male-transmitted mitochondria have decreased membrane potential compared to

444   female-transmitted mitochondria, which would further support the role of the *GCNT1* gene

445   network in degradation of M mtDNA during early development.

446

447   *No evidence of nuclear sex determination in freshwater mussels*

448   Our findings did not show evidence of a nuclear SDR. This result is not surprising given bivalves

449   lack heteromorphic sex chromosomes, and that SDRs on homomorphic sex chromosomes can be

450   small. However, genome-wide heterozygosity was also unable to identify regions that were

451   enriched for SDR-like alleles using similar methodologies as previous studies in bivalves (Han et

452   al. 2022). Although we did not find evidence of nuclear genes involved in sex determination, we

453   cannot reject that sex is determined by nuclear genes (Kenchington et al. 2009; Ghiselli et al.

454   2012; Milani et al. 2013; Zouros and Rodakis 2019; Zouros 2020), or directly by the male-

455   transmitted mitochondria (Breton et al. 2011; Breton et al. 2022).

23

456   Under the hypothesis of nuclear sex determination, one or more loci could cause the

457   retention or degradation of male-transmitted mitochondria in early development, which may or

458   may not contribute to sexual development. This hypothesis does not require fixed differences

459   between the sexes. We agree with previous hypotheses that nuclear encoded loci must contribute

460   to the degradation or retention of male-transmitted mitochondria in early development, because

461   while mitochondrial loci most likely contribute to more than just respiration, they must interact

462   with nuclear gene products to perform any function (Rand 2001). We did observe high levels of

463   genetic differentiation in numerous genes across the genome of *P. streckersoni* (1346 genes

464   across 1/3 of all scaffolds) concomitant to differential expression of many of these genes, which

465   is consistent with polygenic sex determination. However, this type of sex determination is

466   hypothesized to be evolutionarily unstable (Rice 1986), and there are a only few hypothesized

467   instances of this type of sex determining system across animals (Alexander et al. 2015; Roberts

468   et al. 2016; Schartl et al. 2023). Future studies are necessary to test its potential presence in

469   freshwater mussels.

470

471   *Hypothetical sexual development in freshwater mussels and future directions*

472   We hypothesize that freshwater mussels either have a relatively small, cryptic ZW sex

473   determining region that has remained undetected (as in other bivalves; Han et al. 2022), or have a

474   polygenic sex determining system (Fig. 6). However, it is worth noting that nuclear factors may

475   not be involved in sex determination. Crosses in several DUI lineages (*i.e.*, Mytilida, Unionida,

476   Venerida) suggest that offspring sex is solely impacted by the mother and offspring sex ratios

477   can vary from all female to all male progeny dependent on female brood (Saavedra et al. 1997;

478   Kenchington et al. 2002; Ghiselli et al. 2011; Machordom et al. 2015). Although offspring sex

479   being determined by the mother is consistent with ZW nuclear sex determination, extreme

480   variation in sex ratios under natural conditions deviates from the expected sex ratios (50% of

481   each sex). In *Mytilus*, mothers that give rise to all-male progeny are relatively rare (Saavedra et

482   al. 1997; Kenchington et al. 2002), which is coincident in other groups with CMS and may

483   suggest mtDNAs act as the primary sex determination signal. Evaluating each of these

484   hypotheses in *P. streckersoni*, however, will require a more contiguous genome assembly, more

485   thorough genome resequencing, and/or sequencing of parents and progeny from controlled

486   crosses. Controlled crosses are feasible in captivity, but difficult for two reasons: 1) the

487   freshwater mussel life cycle requires temporary larval encystment on vertebrate hosts to

488   complete metamorphosis, and 2) most species reach sexual maturity at a relatively old age (1-2

489   years) compared to model systems (Barnhart et al. 2008; Haag 2012).

490   Despite not identifying a nuclear genetic sex determining region, our results do provide

491   support for a direct role of M mt-sncRNAs acting as a "restorer" of maleness by directly

492   inhibiting female developmental pathways through RNAi, with the mucin-type *O*-glycosylation

493   pathway bring a prime target that may be important across bivalves. The knockdown of the

494   pathway may inhibit female development or halt apoptosis of male-transmitted mitochondria in

495   early development, but this remains speculative. Our hypothetical pathway is supported, at least

496   in part, by empirical data and is consistent with previous hypotheses of sex determination in

497   freshwater mussels (Breton et al. 2011; Breton et al. 2022). Although we did observe female

498   upregulation of genes in the mucin-type *O*-glycosylation pathway (Fig. 5), we did not observe

499   high genetic differentiation ($F_{ST} \geq 0.1$) between the sexes in *GCNT1* or any other genes in the

500   pathway. We also did not identify any ZW-like SNPs in the genes, which together with patterns

501   of genetic differentiation, suggest that the pathway is epigenetically regulated. Future studies
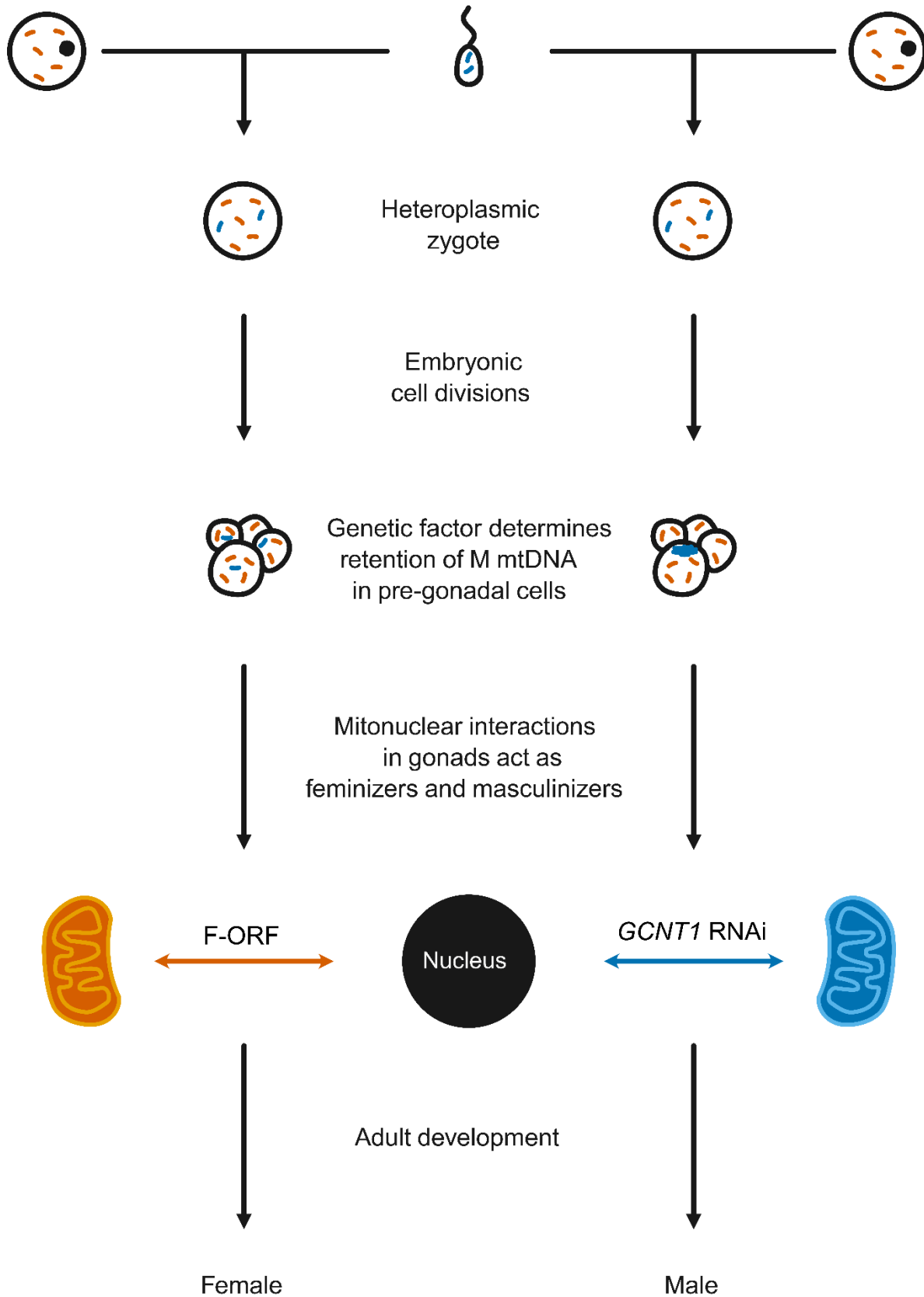
Figure 6. Hypothetical sexual developmental pathway in *Potamilus streckersoni*.

504 testing the regulatory roles of mt-sncRNAs, profiling gene and mt-sncRNAs expression during

505 early development, and the use of mediated RNAi will be helpful in determining the role of

506 mucin-type *O*-glycosylation pathway in female development. Further support could come from

507 data on the *in vivo* activity of M mt-scnRNAs (as in Passamonti et al. 2020), their functional

508 binding (Pozzi and Dowling 2022), their expression, and the expression of their targets during

509 early development. Silencing of *GCNT1* through RNAi in female embryos will be helpful to

510 determine if the mucin-type *O*-glycosylation pathway plays a role in female development. If the

511 pathway is necessary for female development, silencing of that pathway should be strongly

512 selected against in female embryos or lead to exclusively male progeny.

513   The role of the F-ORF or F mtDNA acting as a feminizer remains unclear. We were able

514 to find limited support for the F-ORF protein interacting with the *GCNT1* (Fig. S1), but we

515 cannot confirm that the two proteins interact. It is worth noting that the accuracy PPI prediction

516 in non-model species is relatively poor (*e.g.*, Sledzieski et al. 2021). Future analytic advances or

517 experimental datasets may provide empirical support for the F-ORF interacting with *GCNT1*. We

518 were also able to validate 24 F mt-sncRNAs, but at this point we were unable to identify any

519 targets that were supported by RNA-seq data to have a role in sexual development. It is worth

520 noting that we were able to identify a putative F mt-sncRNA at the 5' end of the F-ORF but

521 could not validate a nuclear target at this time. Of the four genes targeted by F mt-sncRNAs that

522 could be annotated (*i.e.*, *Fanconi Anemia Complementation Group E* [*FANCE*], *Krüppel*

523 *associated box* [*KRAB*], *Nucleoporin 155*, *TRPM2*) and followed patterns expected under RNAi

524 (*i.e.*, significant upregulation in males), we were unable to determine any biological functions

525 potentially relevant to feminization. This is because the gene targets are broadly involved with

526 complex gene pathways that provide multiple biological functions, albeit *FANCE*, *KRAB*, and

27

527 *TRPM2* appear to have some level of contribution to mitophagy (Barde et al. 2013; Rodríguez

528 and D'Andrea 2017; Kang et al. 2018). Future research will be necessary to infer the role of F

529 mt-sncRNAs in female development, but we hypothesize most are involved in mitochondrial

530 maintenance in gonadal and somatic cells.

531

532 **Data Availability Statement**

533 All novel sequencing reads used in this study are available under BioProject PRJNA926666 on

534 the NCBI SRA. Additional previously published reads used in the study are available under

535 BioProject PRJNA681676. The updated genome annotation is available on NCBI under the

536 WGS project accession JAEAOA01. Associated files and scripts used in this study are available

537 in supplemental information and on GitHub

538 (https://github.com/raquelmejiatrujillo/Mitonuclear_sex_determination_in_freshwater_mussels).

539

540 **Acknowledgements**

28

References

547

548 Alexander HJ, Richardson JML, Edmands S, Anholt BR. 2015. Sex without sex chromosomes:
549   genetic architecture of multiple loci independently segregating to determine sex ratios in
550   the copepod *Tigriopus californicus*. *J Evol Biol* 28:2196–2207.

551 Ambros V. 2004. The functions of animal microRNAs. *Nature* 431:350–355.

552 Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman T-L, Hahn MW, Kitano J,
553   Mayrose I, Ming R, et al. 2014. Sex determination: Why so many ways of doing it? *PLoS*
554   *Biol* 12:e1001899.

555 Ballantyne JS, Moon TW. 1985. Hepatopancreas mitochondria from *Mytilus edulis*: substrate
556   preferences and effects of pH and osmolarity. *Mar Biol* 87:239–244.

557 Barde I, Rauwel B, Marin-Florez RM, Corsinotti A, Laurenti E, Verp S, Offner S, Marquis J,
558   Kapopoulou A, Vanicek J, et al. 2013. A KRAB/KAP1-miRNA cascade regulates
559   erythropoiesis through stage-specific control of mitophagy. *Science* 340:350–353.

560 Barnhart MC, Haag WR, Roston WN. 2008. Adaptations to host infection and larval parasitism
561   in Unionoida. *J N Am Benthol Soc* 27:370–394.

562 Basu S, Wallner B. 2016. DockQ: A quality measure for protein-protein docking models. *PLoS*
563   *ONE* 11:e0161879.

564 Bettinazzi S, Nadarajah S, Dalpé A, Milani L, Blier PU, Breton S. 2020. Linking paternally
565   inherited mtDNA variants and sperm performance. *Phil Trans R Soc B* 375:20190177.

566 Bettinazzi S, Rodríguez E, Milani L, Blier PU, Breton S. 2019. Metabolic remodelling associated
567   with mtDNA: insights into the adaptive value of doubly uniparental inheritance of
568   mitochondria. *Proc R Soc B* 286:20182708.

569 Breton S, Beaupré HD, Stewart DT, Hoeh WR, Blier PU. 2007. The unusual system of doubly
570   uniparental inheritance of mtDNA: isn't one enough? *Trends Genet* 23:465–474.

571 Breton S, Bouvet K, Auclair G, Ghazal S, Sietman BE, Johnson N, Bettinazzi S, Stewart DT,
572   Guerra D. 2017. The extremely divergent maternally- and paternally-transmitted
573   mitochondrial genomes are co-expressed in somatic tissues of two freshwater mussel
574   species with doubly uniparental inheritance of mtDNA. *PLoS ONE* 12:e0183529.

575 Breton S, Stewart DT, Brémaud J, Havird JC, Smith CH, Hoeh WR. 2022. Did doubly
576   uniparental inheritance (DUI) of mtDNA originate as a cytoplasmic male sterility (CMS)
577   system? *BioEssays* 44:2100283.

578 Breton S, Stewart DT, Shepardson S, Trdan RJ, Bogan AE, Chapman EG, Ruminas AJ,
579   Piontkivska H, Hoeh WR. 2011. Novel protein genes in animal mtDNA: A new sex
580   determination system in freshwater mussels (Bivalvia: Unionoida)? *Mol Biol Evol*
581   28:1645–1659.

582 Broad Institute. 2019. Picard toolkit. *Broad Institute, GitHub repository* [Internet]. Available
583   from: https://broadinstitute.github.io/picard/

584 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009.
585   BLAST+: architecture and applications. *BMC Bioinform* 10:421.

586 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G,
587   Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics*
588   27:2156–2158.

589 Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T,
590   McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools.
591   *GigaScience* 10:giab008.

592 Doucet-Beaupré H, Breton S, Chapman EG, Blier PU, Bogan AE, Stewart DT, Hoeh WR. 2010.
593   Mitochondrial phylogenomics of the Bivalvia (Mollusca): searching for the origin and
594   mitogenomic correlates of doubly uniparental inheritance of mtDNA. *BMC Evol Biol*
595   10:50.

596 Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*
597   26:2460–2461.

598 Feron R, Jaron KS. 2021. PSASS.

599 Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020.
600   RepeatModeler2 for automated genomic discovery of transposable element families. *Proc*
601   *Natl Acad Sci USA* 117:9451–9457.

602 Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing.
603   *arXiv*:1207.3907.

604 Ghiselli F, Maurizii MG, Reunov A, Ariño-Bassols H, Cifaldi C, Pecci A, Alexandrova Y, Bettini
605   S, Passamonti M, Franceschini V, et al. 2019. Natural heteroplasmy and mitochondrial
606   inheritance in bivalve molluscs. *Integr Comp Biol* 59:1016–1032.

607 Ghiselli F, Milani L, Chang PL, Hedgecock D, Davis JP, Nuzhdin SV, Passamonti M. 2012. De
608   novo assembly of the Manila Clam *Ruditapes philippinarum* transcriptome provides new
609   insights into expression bias, mitochondrial doubly uniparental inheritance and sex
610   determination. *Mol Biol Evol* 29:771–786.

611 Ghiselli F, Milani L, Passamonti M. 2011. Strict sex-specific mtDNA Segregation in the germ
612   line of the DUI species *Venerupis philippinarum* (Bivalvia: Veneridae). *Mol Biol Evol*
613   28:949–961.

614 Guerra D, Lopes-Lima M, Froufe E, Gan HM, Ondina P, Amaro R, Klunzinger MW, Callil C,
615   Prié V, Bogan AE, et al. 2019. Variability of mitochondrial ORFans hints at possible
616   differences in the system of doubly uniparental inheritance of mitochondria among
617   families of freshwater mussels (Bivalvia: Unionida). *BMC Evol Biol* 19:229.

618 Haag WR. 2012. North American freshwater mussels: Natural history, ecology, and conservation.
619   Cambridge, NY, USA: Cambridge University Press

620 Han W, Liu L, Wang J, Wei H, Li Yuli, Zhang Lijing, Guo Z, Li Yajuan, Liu T, Zeng Q, et al.
621   2022. Ancient homomorphy of molluscan sex chromosomes sustained by reversible sex-
622   biased genes and sex determiner translocation. *Nat Ecol Evol* 6:1891–1906.

623 Hanson MR, Bentolila S. 2004. Interactions of mitochondrial and nuclear genes that affect male
624   gametophyte development. *Plant Cell* 16:S154–S169.

625 Hoeh WR, Blakley KH, Brown WM. 1991. Heteroplasmy suggests limited biparental inheritance
626   of *Mytilus* mitochondrial DNA. *Science* 251:1488–1490.

627 Johnson CD, Wang S, Stull M, Schulze E, Brun M, Bennett G, McKown M, Metz RP. 2019.
628   Reducing the cost on plant genotyping by low coverage sequencing.

629 Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates
630   R, Žídek A, Potapenko A, et al. 2021. Highly accurate protein structure prediction with
631   AlphaFold. *Nature* 596:583–589.

632 Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. 2023. KEGG for
633   taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res* 51:D587–D592.

634 Kang P, Zhang W, Chen X, Yi X, Song P, Chang Y, Zhang S, Gao T, Li C, Li S. 2018. TRPM2
635   mediates mitochondria-dependent apoptosis of melanocytes under oxidative stress. *Free
636   Radic Biol Med* 126:259–268.

637 Kenchington E, MacDonald B, Cao L, Tsagkarakis D, Zouros E. 2002. Genetics of mother-
638   dependent sex ratio in Blue Mussels (*Mytilus* spp.) and implications for doubly
639   uniparental inheritance of mitochondrial DNA. *Genetics* 161:1579–1588.

640 Kenchington EL, Hamilton L, Cogswell A, Zouros E. 2009. Paternal mtDNA and maleness are
641   co-inherited but not causally linked in mytilid mussels. *PLoS ONE* 4:e6976.

642 Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and
643   genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 37:907–915.

644 Kirkpatrick M, Sardell JM, Pinto BJ, Dixon G, Peichel CL, Schartl M. 2022. Evolution of the
645   canonical sex chromosomes of the guppy and its relatives. *G3-Genes Genom Genet*
646   12:jkab435.

647 Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. 2021. Fast gene
648   set enrichment analysis. *bioRxiv*:060012.

649 Krüger J, Rehmsmeier M. 2006. RNAhybrid: microRNA target prediction easy, fast and flexible.
650   *Nucleic Acids Res* 34:W451–W454.

651 Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network
652   analysis. *BMC Bioinform* 9:559.

653 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*
654   9:357–359.

655 Li Q, Chu Y, Li S, Yu L, Deng H, Liao C, Liao X, Yang C, Qi M, Cheng J, et al. 2022. The
656   oncoprotein MUC1 facilitates breast cancer progression by promoting Pink1-dependent
657   mitophagy via ATAD3A destabilization. *Cell Death Dis* 13:899.

658 Liao Y, Smyth GK, Shi W. 2019. The R package Rsubread is easier, faster, cheaper and better for
659   alignment and quantification of RNA sequencing reads. *Nucleic Acids Res* 47:e47–e47.

660 Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. 2015. The molecular
661   signatures database hallmark gene set collection. *Cell Syst* 1:417–425.

662 Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL.
663   2011. ViennaRNA Package 2.0. *Algorithms Mol Biol* 6:26.

664 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for
665   RNA-seq data with DESeq2. *Genome Biol* 15:550.

666 Machordom A, Araujo R, Toledo C, Zouros E, Ladoukakis ED. 2015. Female-dependent
667   transmission of paternal mtDNA is a shared feature of bivalve species with doubly
668   uniparental inheritance (DUI) of mitochondrial DNA. *J Zool Syst Evol Res* 53:200–204.

669 Maeda G, Iannello M, McConie HJ, Ghiselli F, Havird JC. 2021. Relaxed selection on male
670   mitochondrial genes in DUI bivalves eases the need for mitonuclear coevolution. *J Evol*
671   *Biol* 34:1722–1736.

672 Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update: Novel and
673   streamlined workflows along with broader and deeper phylogenetic coverage for scoring
674   of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* 38:4647–4654.

675 Md V, Misra S, Li H, Aluru S. 2019. Efficient architecture-aware acceleration of bwa-mem for
676   multicore systems.

677 Mejia-Trujillo R, Smith CH. 2022. The male-type mitochondrial genome of the freshwater
678   mussel *Potamilus streckersoni* Smith, Johnson, Inoue, Doyle, & Randklev, 2019
679   (Bivalvia: Unionidae). *Mitochondrial DNA B Resour* 7:1854–1858.

680 Milani L, Ghiselli F. 2015. Mitochondrial activity in gametes and transmission of viable mtDNA.
681   *Biol Direct* 10:22.

682 Milani L, Ghiselli F, Nuzhdin SV, Passamonti M. 2013. Nuclear genes with sex bias in *Ruditapes*
683   *philippinarum* (Bivalvia, veneridae): Mitochondrial inheritance and sex determination in
684   DUI species. *J Exp Zool (Mol Dev Evol)* 320:442–454.

685 Mitchell A, Guerra D, Stewart D, Breton S. 2016. In silico analyses of mitochondrial ORFans in
686   freshwater mussels (Bivalvia: Unionoida) provide a framework for future studies of their
687   origin and function. *BMC Genom* 17:597.

688 Palmer J, Stajich J. 2017. Funannotate: eukaryotic genome annotation pipeline. Available from:
689   https://funannotate.readthedocs.io/en/latest/

690 Passamonti M, Calderone M, Delpero M, Plazzi F. 2020. Clues of in vivo nuclear gene
691   regulation by mitochondrial short non-coding RNAs. *Sci Rep* 10:8219.

692 Pozzi A, Dowling DK. 2021. Small mitochondrial RNAs as mediators of nuclear gene
693   regulation, and potential implications for human health. *BioEssays* 43:2000265.

694 Pozzi A, Dowling DK. 2022. New Insights into mitochondrial–nuclear interactions revealed
695   through analysis of small RNAs.Yi S, editor. *Genome Biology and Evolution* 14:evac023.

696 Pozzi A, Plazzi F, Milani L, Ghiselli F, Passamonti M. 2017. SmithRNAs: Could mitochondria
697   "bend" nuclear regulation? *Mol Biol Evol* 34:1960–1973.

698 Rand DM. 2001. The units of selection on mitochondrial DNA. *Annu Rev Ecol Syst* 32:415–448.

699 Rice W. 1986. On the instability of polygenic sex determination: The effect of sex-specific
700   selection. *Evolution* 40:633–639.

701 Roberts NB, Juntti SA, Coyle KP, Dumont BL, Stanley MK, Ryan AQ, Fernald RD, Roberts RB.
702   2016. Polygenic sex determination in the cichlid fish Astatotilapia burtoni. *BMC Genom*
703   17:835.

704 Rodríguez A, D'Andrea A. 2017. Fanconi anemia pathway. *Curr Biol* 27:R986–R988.

705 Rojansky R, Cha M-Y, Chan DC. 2016. Elimination of paternal mitochondria in mouse embryos
706   occurs through autophagic degradation dependent on PARKIN and MUL1. *eLife*
707   5:e17896.

708 Saavedra C, Reyero M-I, Zouros E. 1997. Male-dependent doubly uniparental inheritance of
709   mitochondrial DNA and female-dependent sex-ratio in the mussel *Mytilus*
710   *galloprovincialis*. *Genetics* 145:1073–1082.

711 Schartl M, Georges A, Marshall Graves JA. 2023. Polygenic sex determination in vertebrates – is
712   there any such thing? *Trends Genet* 39:242–250.

713 Singh AP, Harada S, Mishina Y. 2009. Downstream genes of *Sox8* that would affect adult male
714   fertility. *Sex Dev* 3:16–25.

715 Skibinski DO, Gallagher C, Beynon CM. 1994. Sex-limited mitochondrial DNA transmission in
716   the marine mussel *Mytilus edulis*. *Genetics* 138:801–809.

717   Sledzieski S, Singh R, Cowen L, Berger B. 2021. D-SCRIPT translates genome to phenome with
718   sequence-based, structure-aware, genome-scale predictions of protein-protein
719   interactions. *Cell Syst* 12:969-982.e6.

720   Smit AFA, Hubley R, Green P. 2015. RepeatMasker Open-4.0. Available from:
721   http://www.repeatmasker.org

722   Smith CH. 2021. A high-quality reference genome for a parasitic bivalve with doubly
723   uniparental. *Genome Biol Evol* 13:evab029.

724   Smith CH, Johnson NA, Robertson CR, Doyle RD, Randklev CR. 2021. Establishing
725   conservation units to promote recovery of two threatened freshwater mussel species
726   (Bivalvia: Unionida: *Potamilus*). *Ecol Evol* 11:11102–11122.

727   Smith CH, Pinto BJ, Kirkpatrick M, Hillis DM, Pfeiffer JM, Havird JC. 2023. A tale of two
728   paths: The evolution of mitochondrial recombination in bivalves with doubly uniparental
729   inheritance. *J Hered*:esad004.

730   Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A,
731   Pomeroy SL, Golub TR, Lander ES, et al. 2005. Gene set enrichment analysis: A
732   knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl.*
733   *Acad. Sci. U.S.A.* 102:15545–15550.

734   Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva
735   NT, Morris JH, Bork P, et al. 2019. STRING v11: protein–protein association networks
736   with increased coverage, supporting functional discovery in genome-wide experimental
737   datasets. *Nucleic Acids Research* 47:D607–D613.

738   Tarazona S, Furió-Tarí P, Turrà D, Pietro AD, Nueda MJ, Ferrer A, Conesa A. 2015. Data quality
739   aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package.
740   *Nucleic Acids Res*:gkv711.

741   Tran DT, Ten Hagen KG. 2013. Mucin-type *O*-Glycosylation during Development. *J Biol Chem*
742   288:6921–6929.

743   Yu D, Chojnowski G, Rosenthal M, Kosinski J. 2023. AlphaPulldown—a python package for
744   protein–protein interaction screens using AlphaFold-Multimer. *Bioinformatics*
745   39:btac749.

746   Zhou L, Liu Z, Dong Y, Sun X, Wu B, Yu T, Zheng Y, Yang A, Zhao Q, Zhao D. 2019.
747   Transcriptomics analysis revealing candidate genes and networks for sex differentiation
748   of yesso scallop (*Patinopecten yessoensis*). *BMC Genom* 20:671.

749   Zhou Q, Li Haimin, Li Hanzeng, Nakagawa A, Lin JLJ, Lee E-S, Harry BL, Skeen-Gaar RR,
750   Suehiro Y, William D, et al. 2016. Mitochondrial endonuclease G mediates breakdown of
751   paternal mitochondria upon fertilization. *Science* 353:394–399.

752   Zouros E. 2020. Doubly uniparental inheritance of mitochondrial DNA: Might it be simpler than
753   we thought? *J Zool Syst Evol Res* 58:624–631.

754   Zouros E, Rodakis GC. 2019. Doubly uniparental inheritance of mtDNA: An unappreciated
755   defiance of a general rule. In: Sutovsky P, editor. Cellular and Molecular Basis of
756   Mitochondrial Inheritance. Vol. 231. Advances in Anatomy, Embryology and Cell
757   Biology. Cham: Springer International Publishing. p. 25–49. Available from:
758   https://link.springer.com/10.1007/978-3-030-04570-8