De Novo Assembly of the Manila Clam *Ruditapes philippinarum* Transcriptome Provides New Insights into Expression Bias, Mitochondrial Doubly Uniparental Inheritance and Sex Determination

Fabrizio Ghiselli,[1,*] Liliana Milani,[1] Peter L. Chang,[2] Dennis Hedgecock,[3] Jonathan P. Davis,[4] Sergey V. Nuzhdin,[2] and Marco Passamonti[1]

[1]Dipartimento di Biologia Evoluzionistica Sperimentale, Università di Bologna, Bologna, Italy

[2]Program in Molecular and Computational Biology, Department of Biological Sciences, University of Southern California

[3]Marine and Environmental Biology Section, Department of Biological Sciences, University of Southern California

[4]Taylor Shellfish Farms, Quilcene, Washington

***Corresponding author:** E-mail: fabrizio.ghiselli@unibo.it.

Associate editor: Richard Thomas

Abstract

Males and females share the same genome, thus, phenotypic divergence requires differential gene expression and sex-specific regulation. Accordingly, the analysis of expression patterns is pivotal to the understanding of sex determination mechanisms. Many bivalves are stable gonochoric species, but the mechanism of gonad sexualization and the genes involved are still unknown. Moreover, during the period of sexual rest, a gonad is not present and sex cannot be determined. A mechanism associated with germ line differentiation in some bivalves, including the Manila clam *Ruditapes philippinarum*, is the doubly uniparental inheritance (DUI) of mitochondria, a variation of strict maternal inheritance. Two mitochondrial lineages are present, one transmitted through eggs and the other through sperm, as well as a mother-dependent sex bias of the progeny. We produced a de novo annotation of 17,186 transcripts from *R. philippinarum* and compared the transcriptomes of males and females and identified 1,575 genes with strong sex-specific expression and 166 sex-specific single nucleotide polymorphisms, obtaining preliminary information about genes that could be involved in sex determination. Then we compared the transcriptomes between a family producing predominantly females and a family producing predominantly males to identify candidate genes involved in regulation of sex-specific aspects of DUI system, finding a relationship between sex bias and differential expression of several ubiquitination genes. In mammalian embryos, sperm mitochondria are degraded by ubiquitination. A modification of this mechanism is hypothesized to be responsible for the retention of sperm mitochondria in male embryos of DUI species. Ubiquitination can additionally regulate gene expression, playing a role in sex determination of several animals. These data enable us to develop a model that incorporates both the DUI literature and our new findings.

Key words: *Ruditapes philippinarum*, de novo, transcriptome, doubly uniparental inheritance, sex bias, sex determination.

Introduction

Males and females undergo different selective pressures, some operating in opposite directions. Because both sexes share the same genome (except for sex chromosomes, where present), phenotypic divergence requires sex-specific regulation (Ellegren and Parsch 2007; Arnold et al. 2009), and males and females, even with the same set of genes, show differences in gene expression or use alternative splice forms (Long et al. 1995; Nuzhdin et al. 1997; Jin et al. 2001; McIntyre et al. 2006; Foley et al. 2007; Chang et al. 2011). Overall, sex-related differences in gene expression were observed across a wide range of taxa (Ellegren and Parsch 2007). For example, over 12% of the germ line transcripts of *Caenorhabditis elegans* showed a sex bias, and expression analyses on whole *Drosophila melanogaster* body showed that the proportion of genes presenting a sex bias is around 57% (Jin et al. 2001; Arbeitman et al. 2002; Meiklejohn et al. 2003; Parisi et al. 2003; Ranz et al. 2003; Reinke et al. 2004), and almost all are specific for reproductive tissues (Parisi et al. 2003). For these reasons, the analysis of their expression patterns is pivotal to the understanding of sex determination and differentiation mechanisms (Connallon and Knowles 2005). A common feature of sex-biased genes is that they evolve more rapidly than other genes (Zhang et al. 2004), and genes that are expressed exclusively in males show the greatest amino acid divergence (Richards et al. 2005). Whether or not these patterns would hold true across the animal kingdom is unknown. In this paper, we analyze expression pattern and polymorphism of sex- and family-biased genes in the Manila clam *Ruditapes philippinarum* in order to get insights into the mechanisms of sex determination and mitochondrial doubly uniparental inheritance (DUI) (Skibinski et al. 1994a, 1994b; Zouros et al. 1994a, 1994b).

Sex Determination in Bivalves and DUI

Many bivalves are stable gonochoric species, but the mechanism of gonad sexualization and the genes involved are still unknown (Paz et al. 2005; Breton et al. 2007). During the period of sexual rest, a gonad is not present and sex cannot be determined. Every year in the reproductive season, testis and ovary develop from a group of germ cells (Devauchelle 1990; Milani L, Ghiselli F, Maurizii MG, and Passamonti M, in preparation), and sex can be determined by detecting sperm or oocytes microscopically. In addition, heteromorphic sex chromosomes appear to be absent from bivalves (Sastry 1979; Borsa and Thiriot-Quiévreux 1990).

DUI is a mechanism associated with germ line differentiation in some bivalves, including *R. philippinarum*, a noteworthy variation of the strict maternal inheritance. In DUI species, two mitochondrial lineages are present, one transmitted through eggs (called F, for female-transmitted) and the other transmitted through sperm (called M, for male-transmitted). In DUI, both sexes inherit F mitochondria from the mother, whereas M mitochondria are transmitted from father to sons only (Breton et al. 2007; Passamonti and Ghiselli 2009). Thus, two different mitochondrial genomes, with an unexpectedly high level of sequence divergence, up to 52% (Doucet-Beaupré et al. 2010), are detectable. In addition, a high variability of progeny sex ratio was observed in Mytilus DUI species: Some females produce female-biased offspring, others male-biased and still others a 1:1 ratio (Saavedra et al. 1997; Kenchington et al. 2002, 2009; Cogswell et al. 2006). The existence of lineages presenting skewed sex ratios in DUI animals has been proposed to be a peculiarity of their sex determination mechanism. Experimental evidence suggests that control over sex ratio is exercised by the mother's nuclear genome (Saavedra et al. 1997; Kenchington et al. 2002, 2009; Cogswell et al. 2006). Indeed, matings of the same female with different males always give the same sex ratio, but matings of the same male with different females result in different sex ratios. The genetic factors involved are so far unknown, but probably sex determination in bivalves is oligogenic, with multiple coexisting genes (Kenchington et al. 2002).

In embryos of analyzed DUI species, sperm mitochondria follow two different distribution patterns. In males, they aggregate near the cleavage furrow at the first cell division, eventually segregating into the primordial germ cells, whereas in females they are dispersed among blastomeres and degraded (Cao et al. 2004; Obata and Komaru 2005; Cogswell et al. 2006; Kenchington et al. 2009; Milani L, Ghiselli F, and Passamonti M, in preparation). Sperm mitochondria degradation in mammal embryos is mediated by ubiquitination (Sutovsky et al. 2000; Sutovsky 2003), and a modification of this mechanism has been hypothesized as responsible for the retention of sperm mitochondria in male embryos of DUI species (Kenchington et al. 2002). Other than its role in protein degradation, ubiquitination can regulate gene expression: Ubiquitin proteolysis can control transcription through degradation of specific transcription factors (Salghetti et al. 2001) and can be involved in mRNA processing (Muratani et al. 2005). A role of ubiquitination was observed in sex determination (*Drosophila*: Bayrer et al. 2005; *C. elegans*: Hodgkin 1987; Hansen and Pilgrim 1999; Starostina et al. 2007; Kulkarni and Smith 2008), in sex transition (i.e., gonadal transformation from ovary to testis in proterogynic species) and testis maturation (teleost fishes: Fujiwara et al. 1994; Sun et al. 2008; *C. elegans*: Shimada et al. 2006), and in human male germ cell development (Ginalski et al. 2004).

Sex ratio bias, together with sperm mitochondrial maintenance in male embryos, led to the hypothesis of a relationship between DUI and germ line specification. In more detail, a role of sperm mitochondria in inducing the development of the undifferentiated gonad into a testis was proposed (Saavedra et al. 1997; Kenchington et al. 2002). However, whether the relationship between DUI and sex determination is causative (DUI having an active role in sex determination) or associative (DUI being a byproduct of sex determination) is still an object of debate (see, e.g., Kenchington et al. 2009; Breton, Stewart et al. 2011).

Genomic Resources in Molluscs

Among metazoans, the phylum Mollusca is second only to arthropods in the number of living species and is by far the largest group of the Lophotrocozoa. The class Bivalvia includes both marine and freshwater species; its largest recent family, the Veneridae, originated 350 Ma and contains about 800 species (Mikkelsen et al. 2006). Bivalve molluscs make up an important source of food all over the world, with a production of over 11.7 million metric tons in 2008, corresponding to 22% of the global aquaculture production. Among them, the family Ostreidae has the highest production, closely followed by the Veneridae (Food and Agriculture Organization Statistical Division data). Among Veneridae, *R. philippinarum* alone represents 23.5% of all bivalve production, being one of the most important species in global aquaculture. The importance of bivalves in marine ecosystems and aquaculture argues for the development of bivalve genomics and genomic resources (Hedgecock et al. 2005; Saavedra and Bachère 2006). Some libraries have been reported for commercial bivalves (see, e.g., Boutet et al. 2008; Craft et al. 2010; Milan et al. 2011). However, the structure and gene content of bivalve genomes have been poorly understood and even the most important aquacultured organisms on a global scale are minimally represented in GenBank. Of bivalves entrees in GenBank, *R. philippinarum* represents 1.1% of nucleotide sequences (405 of 36,445), 1.6% of the expressed sequence tags (5,656 of 358,773), and 1.5% of protein sequences (303 of 20,225), all about an order of magnitude lower than for oysters and mussels.

In this paper, we produced a de novo annotation of 17,186 transcripts from *R. philippinarum*, improving significantly the amount of data available to the scientific community. Moreover, our data provide the basis for the development of sex-specific genetics markers that would make the manipulation of sex determination possible, providing a useful tool for selective breeding programs of economically important species.

Table 1. Family Sex Ratios.

Family ID	Total Numbers	Males	Females	Number of Sexed	Percentage of Sexed	Percentage of Males
032	19	1	11	12	63	8
023[a]	60	8	38	46	77	17
014	18	4	9	13	72	31
001	17	5	10	15	88	33
003	43	6	12	18	42	33
007	41	4	8	12	29	33
012	44	11	19	30	68	37
022	33	12	17	29	88	41
030	19	5	7	12	63	42
017	47	8	11	19	40	42
021	36	10	13	23	64	43
027	41	8	10	18	44	44
026	15	5	6	11	73	45
019	36	11	13	24	67	46
010	40	15	16	31	78	48
009	29	4	4	8	28	50
011	23	6	6	12	52	50
029	54	22	19	41	76	54
006	28	12	10	22	79	55
016	35	13	9	22	63	59
008	25	7	4	11	44	64
028	21	6	3	9	43	67
031	21	6	3	9	43	67
020	35	18	7	25	71	72
005	28	10	3	13	46	77
002	31	13	3	16	52	81
024	16	13	3	16	100	81
025[a]	47	31	7	38	88	82
004	49	19	4	23	47	83
Average	32.8	10.1	9.8	19.9	61.5	51.2
Total	951	293	285	578		

NOTE.—Sex ratios in 29 clam families from Taylor Shellfish Farms, Inc. (Quilcene, WA). The overall sex ratio is balanced (percentage of males=51.2), but the heterogeneity of sex ratios across all families is highly significant (chi-square test $P<0.001$).

[a] Includes additional samples taken at the time of selecting clams from transcriptomic analyses.

Here, we report the first whole transcriptome analysis by RNA-Seq performed to identify genes involved in bivalve sex determination and DUI. The characterization of genes associated with reproduction and the analysis of their expression pattern and polymorphism can provide insight into molecular mechanisms regulating sex determination. We compared the transcriptomes of males and females and identified 1,575 genes with strong sex-specific expression and 166 sex-specific single nucleotide polymorphisms (SNPs), obtaining preliminary information about genes that could be involved in sex determination. Furthermore, for the first time in a DUI species outside the Mytilus complex, we confirmed the presence of sex-biased families. Then, we compared the transcriptomes between a family producing predominantly females and a family producing predominantly males to identify candidate genes involved in the regulation of sex-specific aspects of DUI system. Finally, we produced a model that is consistent with the DUI literature and with our new transcriptomic data.

Materials and Methods

Clam Families

In 2006, Taylor Shellfish Farm (Quilcene, WA) generated 29 families of clams by pairwise crosses of animals originally caught in the wild. In the summer of 2009, the sex ratio of these families was determined by relaxing the clams with $MgSO_4$ and taking a needle biopsy of gonad tissue for microscopic examination (table 1). Additional individuals from two families, which were selected for their strong sex bias, 023 (female-biased) and 025 (male-biased), were shipped to the University of Southern California, where the live animals were opened and sexed. The bodies of three males and three females from each family were frozen in liquid nitrogen for eventual preparation of cDNA libraries.

Library Preparation

RNA purification, cDNA synthesis, and Illumina library construction were performed using the protocols of Mortazavi et al. (2008), with the following modifications. Total RNA, mRNA, and DNA were quantified using a Qubit fluorometer (Invitrogen). mRNA fragmentation was performed using Fragmentation Reagent (Ambion) for a 3 min and 50 s incubation at 70 °C and subsequently cleaned through an RNA cleanup kit (Zymo Research). Additional DNA and gel purification steps were conducted using Clean and Concentrator kits (Zymo Research). Each sample individual was barcoded following the Illumina protocol. Two technical replicates were generated per individual for paired-end 71-bp reads on an Illumina Genome Analyzer II, producing a total of 2 technical replicates × 3 biological replicates × 2 sexes × 2 families = 24 samples.

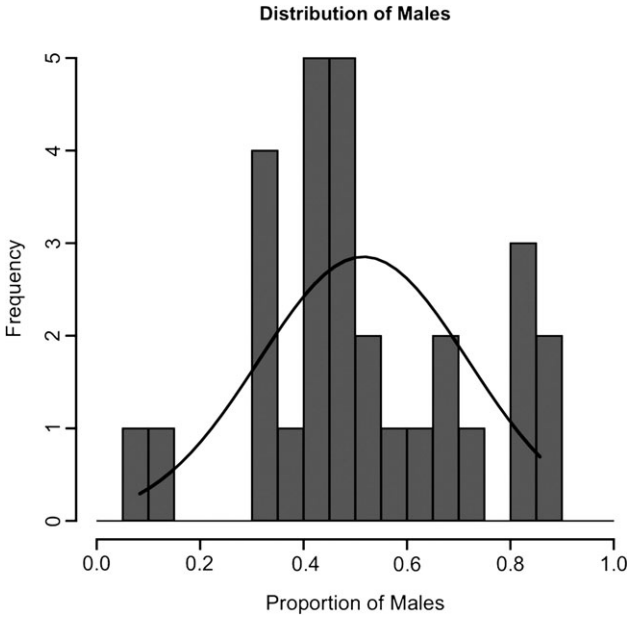

FIG. 1. Male proportion in families. The percentage of males per family ranges from 8% to 83%. The chi-square test is highly significant ($P < 0.001$), supporting the sex ratio heterogeneity across all the families.

Short-Read Sequencing and De Novo Assembly

Across all samples, 90 million (M) paired-end reads, 71 bases long, were obtained. All raw Illumina FastQ sequences are available for download at the NCBI Short Read Archive under the accession number SRA037984.1. Because de novo assembly of transcriptomes from nonmodel species lacking genome sequences can be sensitive to sequencing errors, it is critical that the reads used to generate contigs have the highest sequencing quality. Reads were removed from consideration in the de novo assembly if the read had a terminal "phred" (Ewing and Green 1998) quality value less than 15, or if the read contained more than two unknown nucleotides (i.e., *N*). Reads were also filtered if they were similar to known polymerase chain reaction primer and Illumina adapter sequences. Since the M and F mitochondrial genomes are available (GenBank accession numbers AB065374 and AB065375), additional reads were removed when aligned with six or fewer mismatches to these sequences. After these four filtering steps, 32.9 M paired-end reads and 8.1 M single-end reads were kept for further analyses (supplementary table S1, Supplementary Material online).

The 41 M retained reads were assembled with Velvet (version 1.0.15) (Zerbino and Birney 2008), in conjunction with a custom post-processing algorithm capable of retaining information from alternative splices. Velvet was run under the following settings with a kmer length of 35: -cov_cutoff auto -max_branch_length 0 -max_divergence 0 -max_gap_count 0 -read_trkg yes. Sequenced reads that were kept as pairs and not filtered out together or separately were treated as "-shortPaired" with insert length of 105 bases and standard deviation of 40 bases. Single-end reads that were not filtered out were treated as "-short." ASplice was run with default parameters. These algorithms utilize de Bruijn graphs to assemble short reads into contigs, using sequence overlap information, until the contigs can no longer be extended. The assembly resulted into 35,784 contigs clustering within 22,886 nodes. These nodes represent genes and their isoforms identified in the assembly. The assembled sequences are available on the NCBI Transcription Shotgun Assembly Database (GenBank Accession Numbers: JO101212-JO124029).

Table 2. Summary Statistics of Gene Sequences Identified De Novo.

Total number of reads	**41,031,443**
Total nucleotides	**2,913,232,453**
Contigs	
Total number of contigs	**35,784**
Median length of contig sequences	**590**
N50 length of contig sequences	**1,434**
Representative node sequences	
Total number of nodes	**22,886**
Median length of node sequences	**506**
N50 length of node sequences	**1,011**
Total length of all node sequences	**18,132,893**
Number of nodes having multiple contigs	**6,271**
Number of nodes having at least one N	**425**

With the set of de novo sequences serving as a reference, reads from each of the individual samples were mapped using the Burrows-Wheeler Alignment tool (Li and Durbin 2009). The number of reads that mapped to each gene was tabulated and normalized to calculate fragments per kilobase of exon per million fragment sequenced (FPKM). Additional normalization among all samples was performed, using the Trimmed Mean of M values protocol outlined in Robinson and Oshlack (2010), which takes into account differences in overall RNA populations across biological samples and is one of several methods used to evaluate RNA-sequencing data. Normalization was implemented using the edgeR package in R (Robinson and Smyth 2007). To detect differential expression, we used FPKM values as the dependent variable in gene-specific mixed linear models implemented in R. Sex, family, and sex-by-family interaction were fixed effects, whereas replicates were random effects. The false discovery rate was used to account for multiple testing (Benjamini and Hochberg 1995); a cutoff of 0.05 was applied to call effects significant. The significance of differential gene enrichment between groups was tested using a Wilcoxon rank sum test. All statistical analyses and graphs evaluating consistency between replicates and genotypes were produced using R v2.13.0.

Gene Functional Annotation and Classification

Blast2GO v.2 (Götz et al. 2008) and WEGO (Ye et al. 2006) were used to obtain dene ontology (GO) annotations. Genes were also annotated using a BLASTX (Altschul et al. 1990) search to the nonredundant GenBank CDS translations + PDB + SwissProt + PIR + PRF (nr) database available from GenBank (expected value $<1.00 \times 10^{-5}$). Extensive databases of sequences for Pacific oyster and blue

Table 3. Mean Expression of Genes.

	Sex-Biased Genes (±SE)	Male-Biased Genes (±SE)	Female-Biased Genes (±SE)	Sex-Unbiased Genes (±SE)	Family-Biased Genes (±SE)	Family-Unbiased Genes (±SE)
Males	84.194 (±5.710)	126.805 (±9.234)	21.941 (±2.315)	18.222 (±0.878)	77.417 (±13.971)	22.366 (±0.914)
Females	58.088 (±6.437)	16.264 (±2.879)	119.189 (±14.954)	20.220 (±0.600)	63.317 (±13.790)	22.533 (±0.714)
Family 1	72.297 (±4.893)	70.964 (±5.358)	74.245 (±9.154)	19.068 (±0.817)	59.587 (±12.199)	22.465 (±0.838)
Family 2	69.984 (±4.632)	72.106 (±5.865)	66.885 (±7.524)	19.374 (±0.550)	81.147 (±16.323)	22.435 (±0.601)
Males Family 1	84.599 (±5.797)	127.661 (±9.365)	21.689 (±2.445)	18.431 (±1.378)	60.684 (±11.410)	22.712 (±1.355)
Males Family 2	83.788 (±5.740)	125.949 (±9.296)	22.193 (±2.288)	18.013 (±0.557)	94.150 (±17.838)	22.021 (±0.652)
Females Family 1	59.995 (±7.040)	14.267 (±2.341)	126.801 (±16.643)	19.705 (±0.609)	58.491 (±14.307)	22.217 (±0.747)
Females Family 2	56.18 (±5.926)	18.262 (±3.450)	111.576 (±13.392)	20.735 (±0.615)	68.144 (±15.402)	22.849 (±0.702)
Number of genes	1,575	935	640	21,311	165	22,721

SE, standard error.

mussel are available, but the divergence time with *R. philippinarum* was estimated between 542 and 488 Ma (Plazzi and Passamonti 2010). For this reason, we allowed a higher flexibility and chose the annotation with the highest BLAST score as long as the span of the alignment was greater than 80% of the length of the gene under query. For genes that did not report any hits, we lowered the minimum span to 40% of the length, choosing the annotation with the highest BLAST score, having Expected value $<1.00 \times 10^{-5}$. The GOstat package (Beissbarth and Speed 2004) was used to identify overrepresented GO categories in groups of transcripts ($P < 0.01$). InterProScan version 4.8 (Hunter et al. 2009) was used to identify functional conserved domains of reproductive and ubiquitination genes.

Sequence Polymorphism Analysis

Representative transcript sequences were identified using a global multiple sequence alignment of all contig sequences for each node. For each sample, SNPs were identified with reference to the de novo assembled reference sequence, using SAMtools (Li et al. 2009). Given the nature of the assembly, the SNP data were calculated in a conservative and parsimonious way: Sites with less than 5× coverage were discarded, positions with a phred score lower than 15 were excluded, and indels were not taken into account. All assembled sequences were then aligned and analyzed with the VariScan 2.0 software (Hutter et al. 2006) in order to compute polymorphism data. A block data file was generated to specify gene boundaries in the alignment in order to calculate the statistics for each gene. The program was run under the following settings: RunMode = 12, UseMuts = 1, CompleteDeletion = 0, FixNum = 1, NumNuc = 9, SlidingWindow = 0. This configuration reported the number of segregating sites (S), total number of mutations (η), the number of singletons, nucleotide diversity (π), Watterson's estimator of nucleotide diversity per site (θ), Tajima's D statistic, Fu & Li's D^* and F^*. Sequence polymorphism was analyzed between the following differentially expressed gene categories: family-biased genes versus family-unbiased genes, sex-biased genes versus sex-unbiased genes, male-biased genes versus female-biased genes, and reproductive genes versus male-biased genes. The reproductive gene group included genes annotated by Blast2GO under the "Reproduction" category. We used R

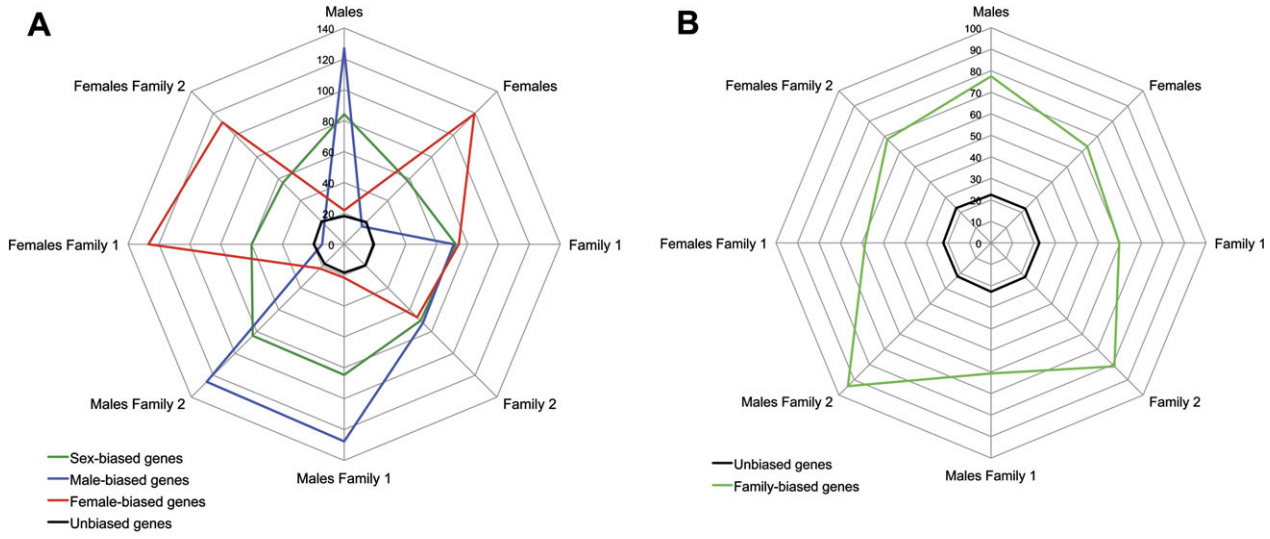

Fig. 2. Radar plots of mean gene expression. (*A*) Sex-biased genes are more highly expressed in males than in females ($P < 2.2 \times 10^{-16}$; table 4). In males, sex-biased genes are 4.7 times more expressed than unbiased genes, whereas in females, the ratio is 2.9 (see table 3). Male-biased genes in females of Family 2 (which produces more males) show higher transcription in comparison to females of Family 1 (which produces more females) ($P = 7.9 \times 10^{-3}$; table 4). Males show higher transcription of female-biased genes than females of male-biased genes ($P < 2.2 \times 10^{-16}$; table 4). (*B*) Family-biased transcripts are more highly expressed in males and females of Family 2, with males having a higher expression than females. In Family 1, the ratio between family-biased and family-unbiased genes is 2.6, whereas it is 3.6 in Family 2 (see table 3).

Table 4. Statistic Significance of Transcription Enrichment Comparison between Groups.

A	B	*P* value (H_1: A > B)[a]
Sex-biased genes	**Sex-biased genes**	
Males	Females	$<2.2 \times 10^{-16}$
Family 1	Family 2	ns
Male-biased genes	**Male-biased genes**	
Family 1	Family 2	ns
M fam1	M fam2	ns
F fam2	F fam1	7.9×10^{-3}
Female-biased genes	**Female-biased genes**	
Family 1	Family 2	ns
M fam1	M fam2	ns
F fam1	F fam2	ns
Female-biased genes	**Male-biased genes**	
Males	Females	$<2.2 \times 10^{-16}$
Females	Males	ns
Family 1	Family 1	ns
Family 2	Family 2	ns
M fam1	F fam1	$<2.2 \times 10^{-16}$
M fam2	F fam2	$<2.2 \times 10^{-16}$
Family-biased genes	**Family-biased genes**	
Males	Females	1.505×10^{-5}
Family 2	Family 1	2.475×10^{-8}
M fam2	M fam1	6.295×10^{-8}
M fam1	F fam1	ns
F fam2	F fam1	2.627×10^{-6}
M fam2	F fam2	2.50×10^{-4}

[a] *P* value of the Wilcoxon rank sum test for alternative hypothesis (H_1): column A > column B.

v2.13.0 to obtain Kernel density plots of *D*, *D**, and *F** and to calculate Wilcoxon rank sum tests of polymorphism between groups. We identified SNPs that were sex or family specific (specific SNP; All individuals within a group have the SNP and no individuals outside of this group have it). We annotated genes containing specific SNPs with Blast2GO.

Results

Sex Ratio-Biased Families

We aimed at comparing the transcriptomes of males and females from families comprising mostly male or female progeny. We secured 951 samples representing 29 families; 578 animals responded to relaxation or were sacrificed and could be sexed. The overall sex ratio of the population was balanced (male/female ratio=0.52), but the percentage of males per family ranged from 8% to 83% (see table 1). The contingency chi-square test is highly significant ($P < 0.001$), supporting sex ratio heterogeneity among families (fig. 1). Female-biased family 023 (males=17%) and male-biased family 025 (males=82%) were chosen for transcriptomic analyses. Here, we refer to the female-biased family as "Family 1" and to the male-biased family as "Family 2."

Short-Read Sequencing and De Novo Assembly

The 35,784 cDNA sequence contigs found within 22,886 nodes represent isoforms, ranging in lengths from 300 to 20,197 bp, with median and N50 lengths of 590 and 1,434 bp, respectively. Contigs within a node can be collapsed into a single "representative node sequence," with median and N50 lengths of 506 and 1,011 bp, respectively. Of the 22,886 node sequences, 6,271 (27%) contain multiple contigs, which potentially correspond to different isoforms. The assembly produced a substantial number of long node sequences: 11,593 (51%) are >500 bp and 4,746 (21%) are >1,000 bp. The total length of all node sequences is 18.1 Mb (table 2).

Expression Bias

We identified 1,575 genes that differed in overall expression levels between males and females (sex-biased genes), 165 genes that differed between the two families (family-biased genes), and 47 genes that had a sex–family interaction effect. Among sex-biased genes, 935 are male-biased, whereas 640 are female-biased (table 3). A radar plot of the mean gene expression within each group is shown in figure 2 (numeric values in table 3), and the statistical significance of the differential transcript enrichment in all analyzed groups is shown in table 4. Sex-biased genes are more highly expressed in males than in females ($P < 2.2 \times 10^{-16}$). Male-biased genes in females of Family 2 (which produces more males) show higher transcription in comparison to females of Family 1 (which produces more females) ($P = 7.9 \times 10^{-3}$). Males show higher transcription of female-biased genes than females of male-biased genes ($P < 2.2 \times 10^{-16}$). Family-biased transcripts are more represented in males and in Family 2 (fig. 2*B*; table 4): Specifically, they are more highly expressed in males and females of Family 2, with males having a higher expression than females.

Annotation

8,473 genes, corresponding to 37% of the entire data set (fig. 3), were annotated with Blast2GO. Contig sequences were also aligned using a BLASTX search to the nr protein database available from GenBank; 12,915 nodes (56%) had a hit when the length of the alignment was required to be greater than 80% of the length of the query. For 4,176 nodes that did not have any acceptable hits, we were able to find local regions of similarity (40% of the length), which could be an indication of a conserved domain. Overall, 17,091 genes (75%) were annotated with BLASTX, providing 8,713 hits in addition to the GO annotations. In total, 17,186 genes were annotated (8,473 with GO and 8,713 with BLASTX), and 5,700 genes (25%) were not annotated (fig. 3). The proportions of GO-annotated genes are almost identical in sex-biased and family-biased transcripts, the lowest proportion being in male-biased genes (32%) and the highest proportion in female-biased genes (44%). The highest percentage of nonannotated genes is in female-biased ones (20%) (fig. 3). The distribution of GO terms (Level 2) is shown in figure 4: In the biological process domain, the most represented terms are cellular process (23%), metabolic process (17%), and biological regulation (10%), whereas developmental process constitutes 7% and reproduction 1.5%. Binding (53%) and catalytic activity (34%) are the principal terms for molecular function. The cellular component domain shows an abundance of the organelle term (32%).

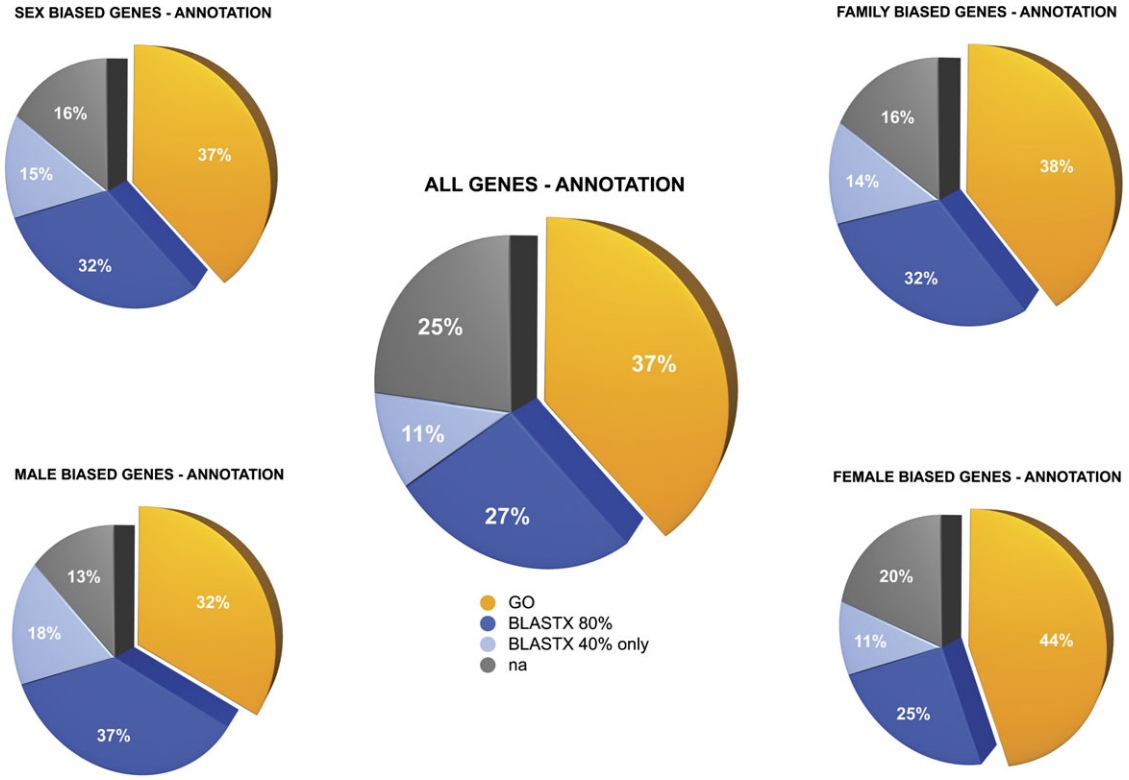

Fig. 3. Proportion of annotated genes. All genes: 8,473 genes, corresponding to the 37% of all the data set (orange color), were annotated with GO. Contig sequences were also aligned using a BLASTX search to the nr protein database available from GenBank: 12,915 nodes resulted in a hit when the length of the alignment was required to be greater than 80% of the length of the query. For 4,176 nodes that did not report any acceptable hits, we were able to find local regions of similarity (40% of the length; light blue color). Overall, 17,091 genes (75%) were annotated with BLASTX, providing 8,713 extra hits in addition to the GO annotation, thus, the nonannotated genes were 5,700, that is, the 25% of the data set (gray color). Dark blue color indicates genes annotated by BLASTX (80%) but not with GO.

DUI is believed to be linked with sex determination (Breton et al. 2007; Passamonti and Ghiselli 2009; Breton, Beaupre, et al. 2011; Breton, Ghiselli, et al. 2011); for this reason, we focused further analyses on the reproductive biological process. In addition, we analyzed transcripts related to the ubiquitination process since it is involved in mitochondrial inheritance (sperm mitochondria degradation in mammals: see Sutovsky et al. 2000). Annotation, function, and expression ratio of biased genes are reported in table 5 (reproductive genes) and table 6 (ubiquitination genes). Most of the overrepresented GO terms (supplementary table S2, Supplementary Material online) belong to male-biased genes and are related to the reproductive process. We also identified functionally conserved domains of genes involved in sex determination and ubiquitination (supplementary fig. S1, Supplementary Material online).

Sequence Polymorphism

The number of SNPs identified for each individual ranged from 14,740 to 27,666 (supplementary table S3, Supplementary Material online). The scatter plot of the coverage against the number of SNPs shows no correlation between them (fig. 5). We calculated the values of S, η, number of singletons, π, θ, Tajima's D, Fu & Li's D^* and F^* for 13,441 genes; for the remaining genes, analysis was not possible due to the absence of polymorphism (i.e., $S = 0$) or because of gaps or missing data in the alignment. Only loci for which at least 9 of the 12 individuals had data were included. Overall, 13,342 family-unbiased genes, 99 family-biased genes, 12,381 sex-unbiased genes, 1,031 biased genes, 495 female-biased genes, 533 male-biased genes, and 29 reproductive genes were analyzed.

For each category, Kernel density plots of Tajima's D, Fu & Li's D^*, and Fu & Li's F^* were obtained (fig. 6), showing notably different frequency distribution between biased and unbiased genes. The Wilcoxon rank sum test between biased and unbiased genes of each category is significant in all cases (P values in table 7), showing that sex- and family-biased genes have higher polymorphism compared with unbiased genes, and that male-biased genes have higher polymorphism than female-biased genes. Reproductive genes appear to be the most variable among the sex-biased genes (table 7; fig. 6).

The specific SNPs are 131 in males, 35 in females, 15 in Family 1, and 6 in Family 2, whereas the number of SNP-containing genes is 103, 30, 14, and 6, respectively (table 4). Genes containing specific SNPs were annotated with Blast2GO (supplementary table S4). The GO annotation was successful for 23% of the male-specific SNP genes and for 30% of the female genes. We identified six reproduction-associated genes containing male-specific SNPs: Four are involved in sperm motility and two in the

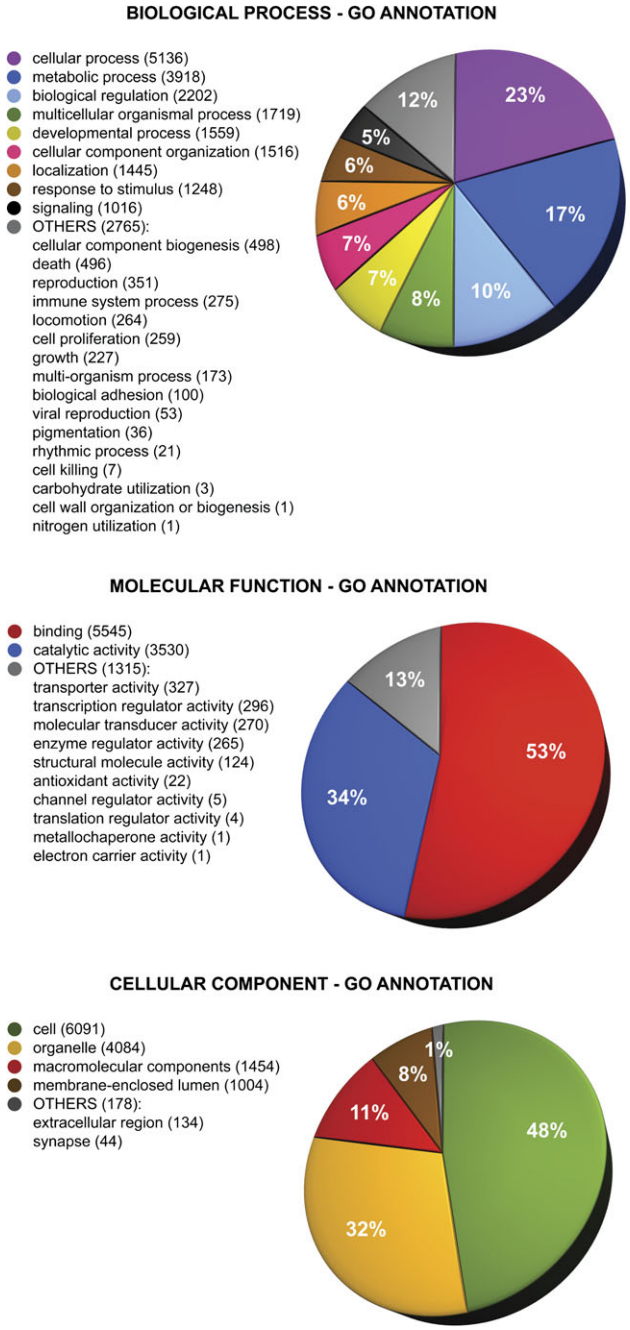

Fig. 4. Distribution of GO terms (Level 2). Biological process domain: The most represented terms are cellular process (23%), metabolic process (17%), and biological regulation (10%), whereas developmental process constitutes 7% and reproduction 1.5%. Molecular function domain: Binding (53%) and catalytic activity (34%) are the principal terms. Cellular component domain: An abundance of the organelle term is present (32%).

ubiquitination process. Genes containing female- or family-specific SNPs do not show any obvious direct involvement in reproduction.

Discussion

Sex-Biased and Family-Biased Gene Expression

We found 1,575 genes showing a significant differential expression between sexes. This is a substantial number considering the lack of secondary sexual characters and sexual dimorphism in bivalves, and that sex-specific function of reproductive genes in these organisms is limited to gonad development, gametogenesis, and fertilization.

As reported for other species (Meiklejohn et al. 2003; Ranz et al. 2003; Ellegren and Parsch 2007), most of the genes contributing to differential expression are male-biased. Males show a higher expression of sex-biased genes (male-biased+female-biased) than females ($P < 0.001$, table 4; fig. 2*A*), and the expression of female-biased genes in males is higher than the expression of male-biased genes in females ($P < 0.001$, table 4). This is consistent with female-biased genes having a higher proportion of essential functions (thus shared by the two sexes) than male-biased genes (Zhang et al. 2004; Proschel et al. 2006; Clark et al. 2007; Ellegren and Parsch 2007; Larracuente et al. 2008).

In order to find genes involved in the skewed sex ratio, we investigated the differences in gene expression between families showing an opposite sex bias (85% of females in Family 1; 82% of males in Family 2) and identified 165 genes with a distinct family-biased expression. We also found interactions between sex and family, meaning that being a male (female) in one family is not the same as being male (female) in the other family, from a transcriptional point of view. Transcriptional activity of family-biased genes seems to follow the trend of the sex towards which expression is biased, with the male-biased family being more transcriptionally active than the female-biased family ($P < 0.001$, table 4; fig. 2*B*), suggesting that upregulation of transcriptionally biased genes is a typical male feature (see also Connallon and Knowles 2005). The most interesting observation is that females of the male-biased family express more male-biased genes than females of the female-biased family ($P < 0.01$, table 4); if their eggs contain a greater amount of male-biased transcripts, a role in male embryo development could be proposed. The process by which maternal factors in the egg influence the early embryonic stages of the progeny is called preformation (reviewed in Extavour and Akam 2003). Preformation was observed in *R. philippinarum* germ line-specific RNA helicase Vasa (Milani et al. submitted) and in the bivalve *Crassostrea gigas* (Fabioux et al. 2004).

Having obtained the expression patterns, we proceeded with annotation. Thirty-seven percent (8,473) of all genes was annotated with GO, and the proportion of identified transcripts increased to 75% (17,186) when the BLASTX annotation was included. Considering the absence of a reference genome and the lack of genetic information from species related to *R. philippinarum*, we find these results gratifying. The highest proportion of GO-annotated sex-biased genes is among female-biased transcripts (44%), whereas male-biased ones have the lowest (32%) (fig. 3). This could be explained by the faster evolution of male-biased genes (thus, the lower percentage of orthologs found through Blast2GO). Quite surprisingly, the opposite situation is observed with nonannotated genes: 20% of female-biased genes are unidentified, against 13% of the

Table 5. Annotation, Function, and Expression Ratio of Biased Reproductive Genes.

GO Annotation	Function	M/F	Fam2/Fam1	M Fam2/M Fam1	F Fam2/F Fam1
Male-biased genes					
Sperm-associated antigen 6	Flagellar protein	10.27***	1.04	1.00	1.61
Kelch-like 10	Spermiogenesis and male fertility	99.31***	1.17	1.16	3.47
14-3-3 protein	Signal transduction and development of spermatozoa	1.85*	1.12	1.05	1.28
Nucleoside diphosphate kinase homolog 5	Spermiogenesis and elimination of reactive oxygen species	4.21***	1.51	1.93**	0.57
Axonemal dynein light chain p33	Sperm motility	3.61***	0.90	0.97	0.68
14-3-3 protein	Signal transduction and development of spermatozoa	1.84*	1.34	1.71**	0.87
Pyruvate dehydrogenase E1 mitochondrial (*pdhe1-a*)	Sperm maturation and capacitation	4.96**	1.29	1.38	0.94
Boule protein	Male fertility and sperm development	4.90**	0.99	1.03	0.82
Forkhead box I1	Sperm maturation	49.52***	0.88	0.87	1.58
cAMP-responsive element-binding protein 1	Transcription	5.01***	1.00	1.02	0.93
DC-STAMP domain containing 1	Acrosome. Fertilization and male fertility	72.82***	1.03	1.04	0.49
Baculoviral IAP repeat-containing 4 (*birc4*)	Antiapoptosis testis specific (E3 ubiquitin protein ligase)	5.03**	4.14*	5.47**	1.53
Meiotic recombination protein REC8 homolog	Meiosis, gametogenesis	59.19***	0.53	0.52**	3.66
Spermatogenesis-associated 4 (*spata4*)	Spermatogenesis, apoptosis	7.64***	1.03	1.08	0.76
inx-1 protein	Gap junctions regulation of fertilization	136.61***	1.16	1.15	(I)**
Forkhead Box j1	Ciliogenesis	15.55***	1.14	1.13	1.22
Muts protein homolog 4	Meiosis, gametogenesis	14.41***	0.90	0.91	0.70
Muts homolog 4	Meiosis, gametogenesis	12.97*	0.97	0.96	1.01
Kelch-like 10	Spermiogenesis and male fertility	111.84**	1.37	1.37	1.32
Synaptonemal complex protein 3	Spermiogenesis and male fertility	19.61***	1.17	1.18	1.15
Centrin-1	Cilia axonemes beating. Sperm centrosome	3.69*	1.21	1.25	1.06
cGMP-gated cation channel alpha-1	Sperm chemosensation and chemotaxis	20.70***	0.63	0.63*	0.64
SRY (sex determining region y)-box 30 (*SOX30*)	Differentiation of developing male cells	60.74*	0.78	0.76	7.60*
ls27 protein	Oocyte triggering	126.16***	1.27	1.29	(II)
Centrosome protein 4	Fertilization	17.64***	0.76	0.74	1.28
Axonemal heavy chain dynein type 3	Sperm motility	15.51***	0.70	0.69*	0.88
Heavy chain 8	Sperm motility	112.51***	0.96	0.98	(II)
Female-biased genes					
30S ribosomal protein S12	Ribosomal protein	0.23***	0.90	0.85	0.91
Translation initiation factor eIF-2B subunit beta	Ovary development	0.22*	1.10	1.18	1.08
Polyspecific ribonuclease PARN	Oogenesis and transcripts maturation	0.06***	1.07	0.69	1.09
Transient receptor potential cation subfamily 2	Sperm fertilization, mechanosensation	0.18***	0.58	0.76	0.54**
Rac GTPase-activating protein 1	Cytokinesis	0.21*	1.60	1.22	1.70
Family-biased genes					
Baculoviral IAP repeat-containing 4 (*birc4*)	Antiapoptosis testis specific (E3 ubiquitin protein ligase)	5.03**	4.14*	5.47**	1.53
14-3-3 protein	Signal transduction and development of spermatozoa	6.11	27.67**	(III)**	3.03
cytosolic phospholipase A2	Acrosome reaction, fertility	7.72	15.11*	22.11**	3.83

NOTE.—(I) expressed only by females of Family 2, (II) expressed only by females of Family 1, and (III) expressed only by males of Family 2.
*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

Table 6. Annotation, Function, and Expression Ratio of Biased Ubiquitination-related Genes.

GO Annotation	Function	M/F	Fam2/Fam1	MFam2/MFam1	FFam2/FFam1
Male-biased genes					
Ubiquitin-conjugating enzyme E2R2	Activated ubiquitin/E3 ubiquitin ligase binding	8.16***	1.10	1.08	1.27
Fizzy cell division cycle 20	Fertility, embryo development	11.25***	0.68	0.69**	0.62*
Cyclin F	F-box domain. SCF complex. E3 cullin-RING ligase	16.66***	1.41	1.37*	2.31*
Cullin 3 (*cul-3*)	Subunit of E3 cullin-RING ligase	2.08**	1.22	1.22	1.23
Baculoviral IAP repeat-containing 4 (*birc4*)	Antiapoptosis testis specific (E3 ubiquitin protein ligase)	5.03**	4.14*	5.47**	1.53
E3 ubiquitin-protein ligase RNF19A	Ubiquitin ligase (E3)	3.17**	0.70	0.47**	2.33
Ubiquitin specific peptidase 2 (*usp2*)	Fertilization, sperm motility. Sex differentiation	76.19***	0.87	0.85	(I)
E3 ubiquitin-protein ligase TRIM33	E3 ubiqitin protein ligase	30.17***	0.80	0.80	0.72
Female-biased genes					
Anaphase-promoting complex subunit 5	Subunit of E3 cullin-RING ligase	0.18**	0.91	1.40	0.85
Axin isoform CRA-a	Wnt signaling. Sexually dimorphic gonad development	0.19***	0.96	1.13	0.93
Peptidylprolyl isomerase 2	Ubiquitin-mediated proteolysis. Germ line determination	0.13***	0.94	0.66	0.98
BMI1 polycomb ring finger	Ubiquitin ligase. Spermatogonia and stem cells proliferation	0.02**	0.65	0.26	0.66
Family-biased genes					
Proteasome subunit alpha 6	Protein degradation. Sperm differentiation	1.00	2.05*	1.91**	2.21**
Baculoviral IAP repeat-containing 4 (*birc4*)	Antiapoptosis testis specific (E3 ubiquitin protein ligase)	5.03**	4.14*	5.47**	1.53
FAD-dependent oxidoreductase 2	Chaperonin	1.38	6.26*	6.34**	6.15**
Ubiquitin-activating enzyme (*uba-1*)	E1-activating enzyme	0.66	9.46**	9.24**	9.58**

NOTE.—(I) expressed only by females of Family 2.
*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

FIG. 5. Scatter plot of reads/numbers of SNPs. Number of reads (FKPM) plotted against number of SNPs (number of SNPs/gene length, per 1,000 bp) for each gene. Black dots: unbiased genes; blue dots: male-biased genes; red dots: female-biased genes. The number of SNPs does not increase at higher coverage.

male-biased ones. We hypothesize that this could be due to a substantial percentage of unknown genes among the female-biased ones (thus the higher proportion of unknown transcripts), but further investigation is needed to clarify this point. A more rapid evolution of reproductive genes, especially of male-enriched transcripts, is known (Reinke et al. 2004; Zhang et al. 2004; Cutter and Ward 2005; Clark et al. 2006). A shared feature of those genes is the significant excess of orphans, that is, genes with no sequence similarity with orthologs from closely related species. That female-biased genes had 12% more Blast2GO matches is indicative of their higher conservation in comparison with the male-biased ones. Testis-biased genes represent the largest class of tissue-specific genes in *D. melanogaster* (Chintapalli et al. 2007), and the observation that expression in sex-limited tissues drives the rapid evolution of sex-biased genes suggests that sex-biased expression may be an adequate predictor of evolutionary rate.

Polymorphism of Transcriptionally Biased Genes

We performed a polymorphism analysis to test if variability patterns in our data set were consistent with the above-discussed observations. The efficiency in detecting polymorphism can be affected by coverage. We did not find a correlation between gene expression and number of SNPs that were identified (fig. 5). To assess polymorphism among differentially transcribed genes, we calculated the values of S, η, number of singletons, π, θ, Tajima's D, Fu & Li's D^* and F^* for 13,441 genes. Kernel density plots of Tajima's D, Fu & Li's D^* and F^* in the first two rows of figure 6 show that both family- and sex-biased genes (red lines) are more variable compared with unbiased genes (black lines; statistical significance in table 7). The last row includes the comparisons

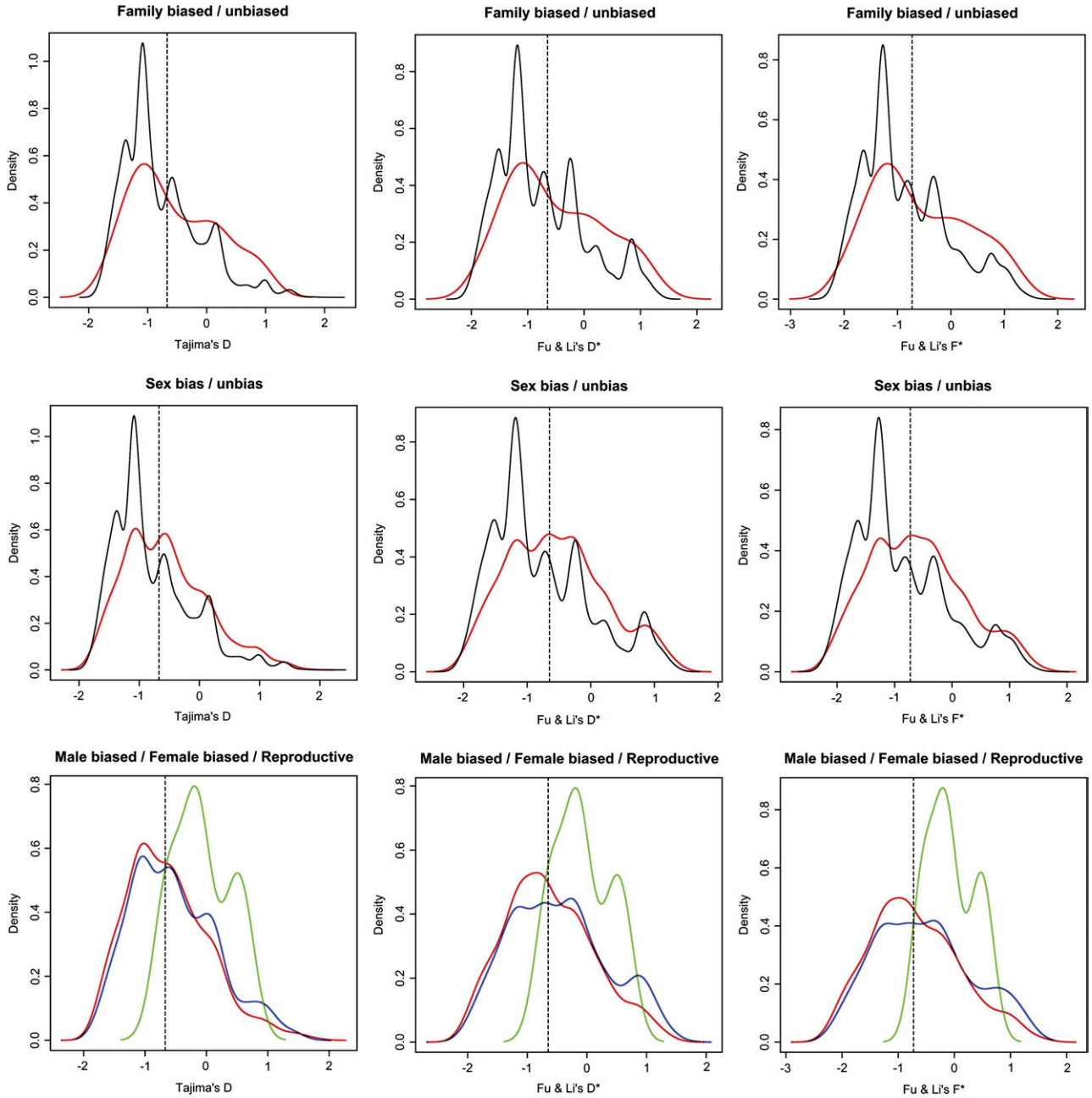

Fig. 6. Kernel density plots of Tajima's *D*, Fu & Li's *D** and *F** values. First two lanes: Kernel density plots for each category showed notably different frequency distribution between biased (red lines) and unbiased genes (black lines). The Wilcoxon rank sum test between biased an unbiased genes is significant in all the cases (*P* values in table 7), and it shows that sex- and family-biased genes have a higher polymorphism compared with unbiased genes. Third lane: Male-biased genes (blue lines) have a higher value than female-biased genes (red lines) and reproductive genes (green lines) appear to be the most variable among the sex-biased genes. Dashed lines indicate mean value.

between female- (red lines) and male-biased genes (blue lines), showing a greater polymorphism in the latter. As discussed above, female-biased genes might have a higher degree of pleiotropy (Mank et al. 2008; Mank and Ellegren 2009; Meisel 2011), thus having few opportunities for neutral and adaptive evolution (e.g., Fisher 1958; Kimura 1983; Connallon and Clark 2011). Green lines depict the variability of biased reproductive genes included in table 5, which represent the most polymorphic subset of the analysis. A higher polymorphism could depend on lower selective constraints (with the accumulation of mutation with no sensible effects on fitness) or positive selection. Reproductive traits are subject to sex-specific natural selection, which affects the fitness of the individual and, therefore, its reproductive success. They are also subject to sexual selection, which acts on fertilization success, mating preference, and sperm competition. Most of the sex-biased reproductive genes included in our analysis are involved in spermiogenesis, male fertility, fertilization, gamete recognition, and sperm motility. These functions play a central role in male reproductive fitness in a broadcast spawning organism (Levitan 1998) and could likely be under positive selection. Our whole-transcriptome

Table 7. Significance of Polymorphism Comparisons between Groups of Differentially Expressed Genes.

A	B	*P* value (H_1: A > B)[a]
Family-biased genes		
Tajima's *D* biased	Tajima's *D* unbiased	5×10^{-4}
Fu & Li's *D** biased	Fu & Li's *D** unbiased	5×10^{-4}
Fu & Li's *F** biased	Fu & Li's *F** unbiased	5×10^{-4}
Sex-biased genes		
Tajima's *D* biased	Tajima's *D* unbiased	$<2.2 \times 10^{-16}$
Fu & Li's *D** biased	Fu & Li's *D** unbiased	$<2.2 \times 10^{-16}$
Fu & Li's *F** biased	Fu & Li's *F** unbiased	$<2.2 \times 10^{-16}$
Tajima's *D* M biased	Tajima's *D* F biased	1×10^{-3}
Fu & Li's *D** M biased	Fu & Li's *D** F biased	2×10^{-4}
Fu & Li's *F** M biased	Fu & Li's *F** F biased	2×10^{-4}
Reproductive genes		
Tajima's *D* reproductive	Tajima's *D* M biased	6.956×10^{-5}
Fu & Li's *D** reproductive	Fu & Li's *D** M biased	8×10^{-4}
Fu & Li's *F** reproductive	Fu & Li's *F** M-biased	2×10^{-4}

[a] *P* value of the Wilcoxon rank sum test for alternative hypothesis (H_1): column A > column B.

scan of polymorphism provides an indication of the evolutionary intraspecific trend of biased genes and appears to be consistent with the patterns observed in other species (Reinke et al. 2004; Zhang et al. 2004; Cutter and Ward 2005; Clark et al. 2006). Establishing positive selection as explanatory for our data would require a deeper analysis that must include sequences from closely related species to perform divergence tests.

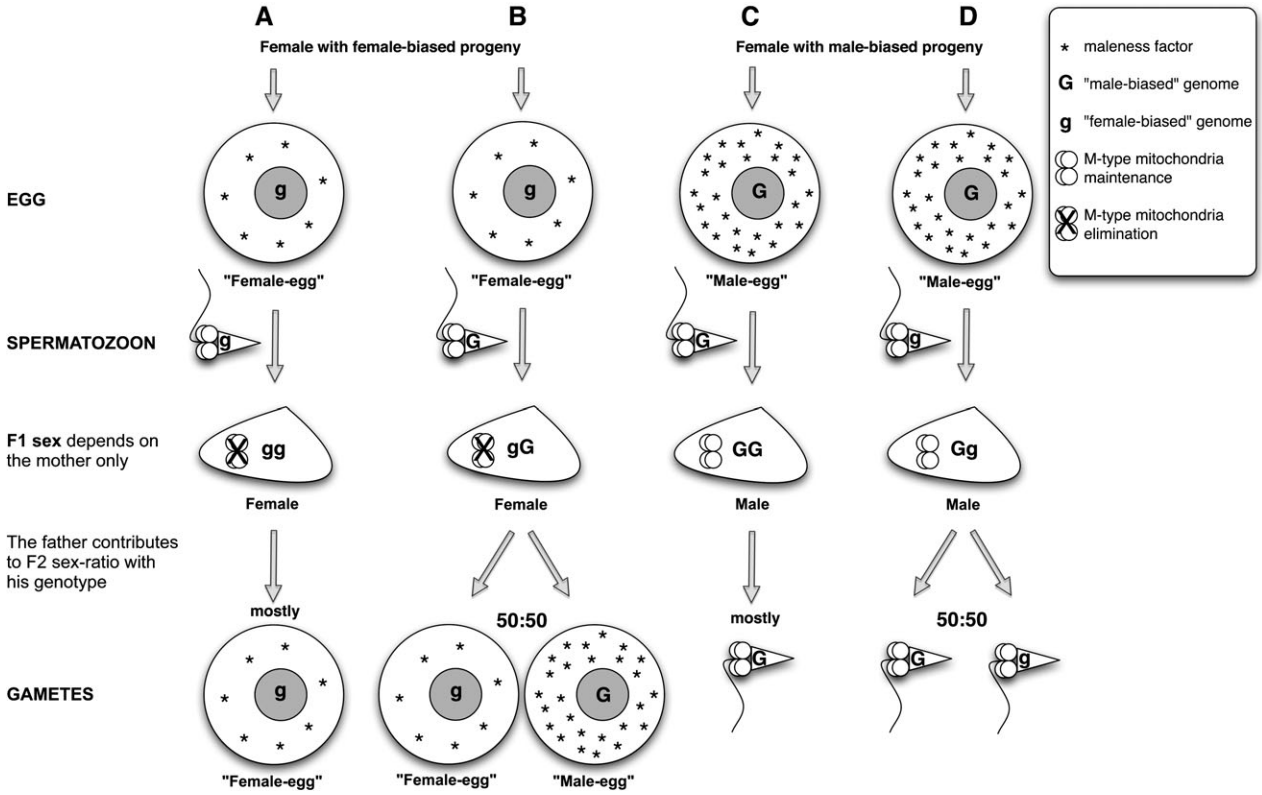

Fig. 7. A simplified model for DUI and sex determination. Transcription factors (e.g., ubiquitination genes) stored in female oocytes would activate sex–gene expression in early embryonic developmental stages, and male development would require the crossing of a critical threshold of masculinizing transcripts. The sperm genotype contributes to F2 sex bias. (*A*, *B*) A "female egg" will produce a female regardless the genotype of the spermatozoon. (*A*) If it is fertilized by a spermatozoon with a "female-biased" genotype (g), the F1 female will produce mostly female eggs. (*B*) If it is fertilized by a spermatozoon with a "male-biased" genotype (G), the F1 female will produce both egg types (50:50). (*C*, *D*) A "male egg" will produce a male regardless the genotype of the spermatozoon. (*C*) If it is fertilized by a spermatozoon with a "male-biased" genotype (G), the F1 male will produce sperm carrying a male-biased genotype (G). (*D*) If it is fertilized by a spermatozoon with a "female-biased" genotype (g), the F1 male will produce both sperm types (50:50). Some ubiquination factors could also be involved in mitochondrial inheritance, and their differential expression could be responsible for the different fate of sperm mitochondria in the two families: degradation (*A*, *B*) or maintenance (*C*, *D*). Note that the genomic sex-determining factors (G and g) probably comprise more than one gene; recombination among these genes and environmental factors could account for the nearly continuous distribution of sex ratios among families (table 1; fig. 1).

Expression Bias in Reproductive and Ubiquitination Genes: Insights on Preformation, Sex Determination, and Mitochondrial Inheritance

Analyzing the different expression patterns of annotated genes that showed sex bias, family bias, and sex–family interactions, we get indications about their implications in *R. philippinarum* sex determination and DUI mechanisms. We focused our analysis on Reproduction and "Ubiquitination" GO categories because of their likely direct relationship with sex determination and DUI. Among the genes annotated in the Reproduction GO category, male-biased are the most represented, both as number of GO terms ($P < 0.01$; supplementary table S2, Supplementary Material online) and number of genes (table 5). Indeed, 27 of 32 sex-biased reproductive genes are male-biased, despite the lower proportion of male-biased genes among annotated transcripts (fig. 3). This points to a higher abundance of reproduction-specific genes among those with a male bias. Moreover, the vast majority of these genes are involved in spermiogenesis and sperm motility (table 5), on which the reproductive success of a broadcast-spawning organism is strongly dependent (Levitan 1998). We also identified about 200 genes involved in ubiquitination, of which 8 show a male-biased transcription, indicating their potential role in spermatogenesis (table 6). The expression pattern observed in the two categories highlights genes that could be involved in the preformation process. Among differentially expressed reproductive genes (table 5), a good candidate is *SRY* (sex determining region y)-box 30 (*SOX30*), which is a transcription factor implicated in the differentiation of developing male germ cells (Wallis et al. 2008). Other than being male-biased, it is 7.6 times more expressed in females of the male-biased family compared with females of the female-biased family. Females of Family 2 could express this gene in the eggs to bias the development of the future embryo towards maleness. Among ubiquitination genes, ubiquitin activating enzyme 1 (*uba-1*) and proteasome subunit alpha 6 (*psa-6*) are both biased towards Family 2 (table 6); This would be in line with a function in male embryo preformation too. The family-biased expression of ubiquitination genes can be easily linked also to the fate of sperm mitochondria (degradation vs. maintenance). This is an important datum for the DUI system since the involvement of ubiquitination was hypothesized on the basis of what was observed in mammals (Sutovsky et al. 2000; Sutovsky 2003) but we still lack direct evidence. Our findings are consistent with a relationship between ubiquitination, sex bias, and mitochondrial inheritance and provide candidate genes for further investigation.

In both the reproductive and ubiquitination categories, baculoviral IAP (inhibitor of apoptosis) repeat-containing 4 (*birc4*) shows a strong sex–family interaction: It is male and Family 2 biased (tables 5 and 6). Mutation of an IAP protein (Birc6) in mouse leads to mitochondrial apoptosis (Ren et al. 2005). The fact that *birc4* is 5.47 times more highly transcribed in males of the male-biased family might indicate a role of this specific gene in sperm mitochondria heredity in DUI species. Accordingly, it is the only gene showing a transcriptional bias among the 20 baculoviral IAP genes annotated.

A Model for DUI and Sex Determination

Here, we propose that the preformation process stands at the basis of both sex determination and DUI. A simplified scheme of the model is shown in figure 7. Transcription factors (e.g., ubiquitination genes) stored in female oocytes during gametogenesis would activate sex–gene expression in the early embryonic developmental stages, and the sex differentiation process would be multifactorial and quantitative. Male development would require the crossing of a critical threshold of masculinizing transcripts (see also Kenchington et al. 2009), and genes containing male-exclusive SNPs could be among those responsible for maleness. Supporting that, we observed that a substantial proportion of genes with male-specific SNPs are involved in sperm functionality (supplementary table S4, Supplementary Material online). Other than having a role in sex determination, some ubiquination factors could also be involved in mitochondrial inheritance, and their differential expression could be responsible for the different fate of sperm mitochondria in the two families. Family 2-biased ubiquination factors could protect sperm mitochondria from degradation and allow them to actively participate in male germ line development (Breton et al. 2007; Passamonti and Ghiselli 2009). Supporting that, a link between DUI and gonochorism (as opposed to hermaphroditism) was demonstrated by Breton, Ghiselli, et al. (2011) in unionid bivalves. Otherwise, DUI could be a side effect of the mechanism of sex determination, as recently proposed by Kenchington et al. (2009): According to them, paternal mitochondrial DNA (mtDNA) and maleness are co-inherited but not causally linked. They observed that in Mytilus hybrid crosses, maternally determined sex bias is disrupted (though only in female-biased mothers), that triploid individuals were males and that some of them did not carry the M-type mtDNA. In the light of our model, transcription factors present in the egg would not function correctly (or be less effective) in hybrids because of regulatory incompatibilities, and the fact that triploids are males would be ascribable to a dosage effect.

Supplementary Material

Supplementary tables S1–S4 and figure S1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We would like to thank Joseph P. Dunham, Daniel Campo-Falgueras, Matthew D. Dean, and two anonymous reviewers for their precious help and suggestions. This work was supported by MRI NSF 0923513 and NIH RGM076643 awarded to S.V.N.; the University and Research Italian Ministry (MIUR PRIN07, grant number 2007NSHJL8_002

to M.P.); and the "Canziani Bequest" fund (University of Bologna, grant number A.31.CANZELSEW to M.P.). Sequencing was done at the sequencing core facility at the University of Oregon in Eugene in the Molecular Biology Institute. Author contributions: F.G. carried out the molecular sequencing and performed statistical analysis of the data; L.M. and F.G. characterized the transcriptome; P.L.C. assembled and characterized the de novo transcriptome and performed statistical analysis of the data; D.H. and J.P.D. participated in the collection of the data; F.G., L.M., and P.L.C. wrote the manuscript; M.P. and S.V.N. participated in the design of the study and coordinated it. All authors read and commented on the manuscript drafts and approved the final manuscript.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215:403–410.

Arbeitman MN, Furlong EE, Imam F, Johnson E, Null BH, Baker BS, Krasnow MA, Scott MP, Davis RW, White KP. 2002. Gene expression during the life cycle of *Drosophila melanogaster*. *Science* 297:2270–2275.

Arnold AP, van Nas A, Lusis AJ. 2009. Systems biology asks new questions about sex differences. *Trends Endocrinol Metab.* 20:471–476.

Bayrer JR, Zhang W, Weiss MA. 2005. Dimerization of doublesex is mediated by a cryptic ubiquitin-associated domain fold: implications for sex-specific gene regulation. *J Biol Chem.* 280:32989–32996.

Beissbarth T, Speed TP. 2004. GOstat: find statistically over-represented Gene Ontologies within a group of genes. *Bioinformatics* 20:1464–1465.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol.* 57:289–300.

Borsa P, Thiriot-Quiévreux C. 1990. Karyological and allozymic characterization of *Ruditapes philippinarum, R. aureus* and *R. decussatus* (Bivalvia, Veneridae). *Aquaculture* 90:209–227.

Boutet I, Moraga D, Marinovic L, Obreque J, Chavez-Crooker P. 2008. Characterization of reproduction-specific genes in a marine bivalve mollusc: influence of maturation stage and sex on mRNA expression. *Gene* 407:130–138.

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

Breton S, Ghiselli F, Passamonti M, Milani L, Stewart DT, Hoeh WR. 2011. Evidence for a fourteenth mtDNA-encoded protein in the female-transmitted mtDNA of marine mussels (Bivalvia: Mytilidae). *PLoS One* 6:e19365.

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

Cao L, Kenchington E, Zouros E. 2004. Differential segregation patterns of sperm mitochondria in embryos of the blue mussel (*Mytilus edulis*). *Genetics* 166:883–894.

Chang PL, Dunham JP, Nuzhdin SV, Arbeitman MN. 2011. Somatic sex-specific transcriptome differences in *Drosophila* revealed by whole transcriptome sequencing. *BMC Genomics* 12:364.

Chintapalli VR, Wang J, Dow JAT. 2007. Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat Genet.* 39:715–720.

Clark NL, Aagaard JE, Swanson WJ. 2006. Evolution of reproductive proteins from animals and plants. *Reproduction* 131:11–22.

Clark NL, Findlay GD, Yi X, Maccoss MJ, Swanson WJ. 2007. Duplication and selection on abalone sperm lysin in an allopatric population. *Mol Biol Evol.* 24:2081–2090.

Cogswell AT, Kenchington ELR, Zouros E. 2006. Segregation of sperm mitochondria in two- and four-cell embryos of the blue mussel *Mytilus edulis*: implications for the mechanism of doubly uniparental inheritance of mitochondrial DNA. *Genome* 49:799–807.

Connallon T, Clark AG. 2011. Association between sex-biased gene expression and mutations with sex-specific phenotypic consequences in *Drosophila*. *Genome Biol Evol.* 3:151–155.

Connallon T, Knowles LL. 2005. Intergenomic conflict revealed by patterns of sex-biased gene expression. *Trends Genet.* 21:495–499.

Craft JA, Gilbert JA, Temperton B, Dempsey KE, Ashelford K, Tiwari B, Hutchinson TH, Chipman JK. 2010. Pyrosequencing of *Mytilus galloprovincialis* cDNAs: tissue-specific expression patterns. *PLoS One* 5:e8875.

Cutter AD, Ward S. 2005. Sexual and temporal dynamics of molecular evolution in *C. elegans* development. *Mol Biol Evol.* 22:178–188.

Devauchelle N. 1990. Sexual development and maturity of Tapes philippinarum. In: *Tapes philippinarum*, biologia e sperimentazione. Treviso (Italy): Ente Sviluppo Agricolo Veneto (ESAV). p. 47–62.

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

Ellegren H, Parsch J. 2007. The evolution of sex-biased genes and sex-biased gene expression. *Nat Rev Genet.* 8:689–698.

Ewing B, Green P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 8:186–194.

Extavour CG, Akam M. 2003. Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. *Development* 130:5869–5884.

Fabioux C, Huvet A, Lelong C, Robert R, Pouvreau S, Daniel JY, Minguant C, Le Pennec M. 2004. Oyster vasa-like gene as a marker of the germline cell development in *Crassostrea gigas*. *Biochem Biophys Res Commun.* 320:592–598.

Fisher RA. 1958. The genetical theory of natural selection, 2nd ed. New York: Dover Publications Inc.

Foley B, Chenoweth SF, Nuzhdin SV, Blows MW. 2007. Natural genetic variation in cuticular hydrocarbon expression in male and female *D. melanogaster*. *Genetics* 175:1465–1477.

Fujiwara Y, Hatano K, Hirabayashi T, Miyazaki JI. 1994. Ubiquitin C-terminal hydrolase as a putative factor involved in sex differentiation of fish (temperate wrasse, *Halichoeres poecilopterus*). *Differentiation* 56:13–20.

Ginalski K, Rychlewski L, Baker D, Grishin NV. 2004. Protein structure prediction for the male-specific region of the human Y chromosome. *Proc Natl Acad Sci U S A.* 101:2305–2310.

Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36:3420–3435.

Hansen D, Pilgrim D. 1999. Sex and the single worm: sex determination in the nematode *C. elegans*. *Mech Dev.* 83:3–15.

Hedgecock D, Gaffney PM, Goulletquer P, Guo XM, Reece K, Warr GW. 2005. The case for sequencing the Pacific oyster genome. *J Shellfish Res.* 24:429–441.

Hodgkin J. 1987. A genetic analysis of the sex-determining gene, *tra-1*, in the nematode *Caenorhabditis elegans*. *Genes Dev.* 1:731–745.

Hunter S, Apweiler R, Attwood TK, et al. (38 co-authors). 2009. InterPro: the integrative protein signature database. *Nucleic Acids Res.* 37(Database Issue):D211–D215.

Hutter S, Vilella AJ, Rozas J. 2006. Genome-wide DNA polymorphism analyses using VariScan. *BMC Bioinformatics* 7:409.

Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gurgel G, Gibson G. 2001. The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nat Genet.* 29:389–395.

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

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

Kimura M. 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.

Kulkarni M, Smith HE. 2008. E1 ubiquitin-activating enzyme UBA-1 plays multiple roles throughout *C. elegans* development. *PLoS Genet.* 4:e1000131.

Larracuente AM, Sackton TB, Greenberg AJ, Wong A, Singh ND, Sturgill D, Zhang Y, Oliver B, Clark AG. 2008. Evolution of protein coding genes in *Drosophila*. *Trends Genet.* 24:114–123.

Levitan DR. 1998. Sperm limitation, sperm competition and sexual selection in external fertilizers. In: Birkhead T, Møller A, editors. Sperm competition and sexual selection. San Diego (CA): Academic Press. p. 173–215.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* 25:1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map (SAM) format and SAMtools. *Bioinformatics* 25:2018–2079.

Long AD, Mullaney SL, Reid LA, Fry JD, Langley CH, Mackay TF. 1995. High-resolution mapping of genetic factors affecting abdominal bristle number in *Drosophila melanogaster*. *Genetics* 139:1273–1291.

Mank JE, Ellegren H. 2009. Are sex-biased genes more dispensable? *Biol Lett.* 5:409–412.

Mank JE, Hultin-Rosenberg L, Zwahlen M, Ellegren H. 2008. Pleiotropic constrain hampers the resolution of sexual antagonism in vertebrate gene expression. *Am Nat.* 171:35–43.

McIntyre LM, Bono LM, Genissel A, Westerman R, Junk D, Telonis-Scott M, Harshman L, Wayne ML, Kopp A, Nuzhdin SV. 2006. Sex-specific expression of alternative transcripts in *Drosophila*. *Genome Biol.* 7:R79.

Meiklejohn CD, Parsch J, Ranz JM, Hartl DL. 2003. Rapid evolution of male-biased gene expression in *Drosophila*. *Proc Natl Acad Sci U S A.* 100:9894–9899.

Meisel RP. 2011. Towards a more nuanced understanding of the relationship between sex-biased gene expression and rates of protein coding sequence evolution. *Mol Biol Evol.* 28:1893–1900.

Mikkelsen PM, Bieler R, Kappner I, Rawlings TA. 2006. Phylogeny of Veneroidea (Mollusca: Bivalvia) based on morphology and molecules. *Zool J Linn Soc.* 148:439–521.

Milan M, Coppe A, Reinhardt R, et al. (11 co-authors). 2011. Transcriptome sequencing and microarray development for the Manila clam, *Ruditapes philippinarum*: genomic tools for environmental monitoring. *BMC Genomics* 12:234.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 5:621–628.

Muratani M, Kung C, Shokat KM, Tansey WP. 2005. The F box protein Dsg1/Mdm30 is a transcriptional coactivator that stimulates Gal4 turnover and cotranscriptional mRNA processing. *Cell* 120:887–899.

Nuzhdin SV, Pasyukova EG, Dilda CL, Zeng ZB, Mackay TF. 1997. Sex-specific quantitative trait loci affecting longevity in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A.* 94:9734–9739.

Obata M, Komaru A. 2005. Specific location of sperm mitochondria in mussel *Mytilus galloprovincialis* zygotes stained by Mito-Tracker. *Dev Growth Differ.* 47:255–263.

Parisi M, Nuttall R, Naiman D, Bouffard G, Malley J, Andrews J, Eastman S, Oliver B. 2003. Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. *Science* 299:697–700.

Passamonti M, Ghiselli F. 2009. Doubly uniparental inheritance: two mitochondrial genomes, one precious model for organelle DNA inheritance and evolution. *DNA Cell Biol.* 28:79–89.

Paz M, Torrado M, Korochkin LI, Mikhailov AT. 2005. Esterase-like and fibronectin-like polypeptides share similar sex-cell-biased patterns in the gonad of hermaphroditic and gonochoric species of bivalve mollusks. *Cell Tissue Res.* 322:475–489.

Plazzi F, Passamonti M. 2010. Towards a phylogeny of mollusks: bivalves' early evolution as revealed by mitochondrial genes. *Mol Phylogenet Evol.* 57:641–657.

Proschel 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.

Reinke V, Gil IS, Ward S, Kazmer K. 2004. Genome-wide germline-enriched and sex-biased expression profiles in *Caenorhabditis elegans*. *Development* 131:311–323.

Ren J, Shi M, Liu R, Yang QH, Johnson T, Skarnes WC, Du C. 2005. The Birc6 (Bruce) gene regulates p53 and the mitochondrial pathway of apoptosis and is essential for mouse embryonic development. *Proc Natl Acad Sci U S A.* 102:565–570.

Richards S, Liu Y, Bettencourt BR, et al. (52 co-authors). 2005. Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. *Genome Res.* 15:1–18.

Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11:R25.

Robinson MD, Smyth GK. 2007. Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* 23:2881–2887.

Saavedra C, Bachère E. 2006. Bivalve genomics. *Aquaculture* 256:1–14.

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

Salghetti SE, Caudy AA, Chenoweth JG, Tansey WP. 2001. Regulation of transcriptional activation domain function by ubiquitin. *Science* 293:1651–1653.

Sastry AN. 1979. Pelecypoda (excluding Ostreidae). In: Giese AC, Pearse JS, editors. Reproduction of marine invertebrates. Vol. 5. New York: Academic Press. p. 113–292.

Shimada M, Kanematsu K, Tanaka K, Yokosawa H, Kawahara H. 2006. Proteasomal ubiquitin receptor RPN-10 controls sex determination in *Caenorhabditis elegans*. *Mol Biol Cell.* 17:5356–5371.

Skibinski DO, Gallagher C, Beynon CM. 1994a. Mitochondrial DNA inheritance. *Nature* 368:817–818.

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

Starostina NG, Lim JM, Schvarzstein M, Wells L, Spence AM, Kipreos ET. 2007. A CUL-2 ubiquitin ligase containing three FEM proteins degrades TRA-1 to regulate *C. elegans* sex determination. *Dev Cell.* 13:127–139.

Sun J, Shang X, Tian Y, Zhao W, He Y, Chen K, Cheng H, Zhou R. 2008. Ubiquitin C-terminal hydrolase-L1 (Uch-L1) correlates with gonadal transformation in the rice field eel. *FEBS J.* 275:242–249.

Sutovsky P. 2003. Ubiquitin-dependent proteolysis in mammalian spermatogenesis, fertilization, and sperm quality control: killing three birds with one stone. *Microsc Res Tech.* 61:88–102.

Sutovsky P, Moreno RD, Ramalho-Santos J, Dominko T, Simerly C, Schatten G. 2000. Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos. *Biol Reprod.* 63:582–590.

Wallis MC, Waters PD, Graves JA. 2008. Sex determination in mammals—before and after the evolution of SRY. *Cell Mol Life Sci.* 65:3182.

Ye J, Fang L, Zheng H, et al. (11 co-authors). 2006. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* 34(Web Server issue):W293–7.

Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18:821–829.

Zhang Z, Hambuch TM, Parsch J. 2004. Molecular evolution of sex-biased genes in *Drosophila*. *Mol Biol Evol.* 21:2130–2139.

Zouros E, Oberhauser Ball A, Saavedra C, Freeman KR. 1994a. An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. *Proc Natl Acad Sci U S A.* 91:7463–7467.

Zouros E, Oberhauser Ball A, Saavedra C, Freeman KR. 1994b. Mitochondrial DNA inheritance. *Nature* 368:818.