Sex-biased expression of microRNAs in *Drosophila melanogaster*

Antonio Marco
School of Biological Sciences, University of Essex, Colchester CO4 3SQ, UK

1. Summary

Most animals have separate sexes. The differential expression of gene products, in particular that of gene regulators, is underlying sexual dimorphism. Analyses of sex-biased expression have focused mostly on protein-coding genes. Several lines of evidence indicate that microRNAs, a class of major gene regulators, are likely to have a significant role in sexual dimorphism. This role has not been systematically explored so far. Here, I study the sex-biased expression pattern of microRNAs in the model species *Drosophila melanogaster*. As with protein-coding genes, sex-biased microRNAs are associated with the reproductive function. Strikingly, contrary to protein-coding genes, male-biased microRNAs are enriched in the X chromosome, whereas female microRNAs are mostly autosomal. I propose that the chromosomal distribution is a consequence of high rates of de novo emergence, and a preference for new microRNAs to be expressed in the testis. I also suggest that demasculinization of the X chromosome may not affect microRNAs. Interestingly, female-biased microRNAs are often encoded within protein-coding genes that are also expressed in females. MicroRNAs with sex-biased expression do not preferentially target sex-biased gene transcripts. These results strongly suggest that the sex-biased expression of microRNAs is mainly a consequence of high rates of microRNA emergence in the X chromosome (male bias) or hitchhiked expression by host genes (female bias).

2. Introduction

Sexual dimorphism is prevalent in animal species. Sexual phenotypic differences are the consequence of a differential expression of genes between males and females [1]. During the past decade, high-throughput transcript analyses have identified many genes with a sex-biased expression pattern [2–4]. For instance, the *Drosophila* gene *paired* is expressed at a higher level in adult males than in females [5], and it encodes a transcription factor involved in the development of male accessory glands [6]. Indeed, other transcription factors have been identified as sex-biased genes [5,7], indicating that transcriptional gene regulation is tightly linked to sexual dimorphism. Post-transcriptional regulators may also have an impact in sexual dimorphism. MicroRNAs are short endogenous regulatory RNA molecules that are involved in virtually all studied biological processes [8,9]. Recently, differences between male and female microRNA expression profiles have been observed [10–12], suggesting that microRNAs have a role in sexual differentiation.

The study of sex-biased expression of gene products in the model species *Drosophila melanogaster* has produced a number of insightful observations. First, male-biased genes evolve faster than non-biased genes [3,13–15]. Second, the X chromosome is depleted of male-based genes and enriched for female-biased genes [2,3]. On the other hand, evolutionarily novel genes tend to be X-linked and highly expressed in males [16–23]. These observations suggest a movement...
of male-biased genes from the X chromosome to the autosomes, a process known as demasculinization of the X [2,20,24]. Therefore, sex-biased expression is an important factor affecting the evolutionary fate of protein-coding genes. Likewise, sex-biased expression should have an impact in microRNA evolution. However, this effect may be different to that observed for protein-coding genes as proteins and microRNAs differ in their evolutionary dynamics. For instance, gene duplication is the main mechanism by which novel protein-coding genes emerge, whereas a majority of microRNAs emerge by de novo formation within existing transcripts (reviewed in [25,26]). Consequently, novel microRNAs are more likely to be lost than protein-coding genes in a short evolutionary period. Although microRNAs have been extensively studied (average log counts per million)

Figure 1. Sex-biased microRNAs in Drosophila melanogaster. (a) Heatmap of cross-correlations of all expression datasets analysed. Different experiments are hierarchically clustered. (b) Smear plot of mature microRNA sequences. Grey lines indicate a twofold difference in expression levels between males and females. Red dots are microRNAs with a statistically significant differential expression. (c) MicroRNA transcripts with sex-biased expression, average fold change of their products and their chromosomal location. (d) Frequency plot of sex biases in expression levels for autosomes and the X chromosome.

3. Results

3.1. Sex-biased expression of Drosophila microRNAs

To characterize which microRNAs have a sex-biased expression pattern in D. melanogaster, 13 different small RNA sequencing experiments (including males, females, embryos, ovaries and testes) were cross-compared (see §3). Figure 1a shows the correlation among the expression profiles for all experiments, indicating that the female and male pairs of profiles are highly correlated, despite coming from independent experiments. Thus, pairs of male and female profiles were used as biological replicates to calculate differential expression between sexes. A total of 476 mature microRNAs (two sequences per microRNA precursor) were analysed. Of them, 28 and 37 mature microRNAs showed a significant expression bias in males and females, respectively (figure 1b; see §5). Table 1 includes details of sex-biased microRNAs and their fold change. The expression levels for all analysed mature microRNAs are available in the electronic supplementary material, table S1. As only reads mapping to a single microRNA were taken into account, removing reads mapping to multiple sites may influence our analysis. Hence, I compared the expression levels resulting from unique reads and from multiple matching reads. Four microRNA families were affected by multiple matching reads: mir-983, mir-281, mir-276 and mir-2. The first three did not show any differential expression between sexes. However, the fourth one included two microRNAs, mir-2a-1 and mir-13b-2, which are female-biased. To avoid biases due to multiple matches, the mir-2 family was removed from the subsequent analyses.

MicroRNA precursors potentially encode for two mature products (so-called 3 prime and 5 prime products). In agreement with this, many of the sex-biased microRNAs are pairs...
derived from the same precursor (table 1). Additionally, microRNAs are frequently clustered in the genome, and these clusters of microRNAs are often transcribed in a single RNA molecule (reviewed in reference [26]). Indeed, sex-biased microRNAs are frequently clustered, and nearly all of the microRNAs in a cluster show a consistent sex-biased expression (table 1 and figure 1c). Therefore, the observed bias in mature microRNA production is primarily a consequence of the sex-biased expression of their transcripts.

### 3.2. Male-biased microRNAs are preferentially located in the X chromosome and expressed in the testes

Figure 1c shows microRNA transcripts with sex-biased expression and their chromosomal distribution. Contrary to the observation for protein-coding genes, microRNAs expressed in males tend to be located in the X chromosome. By contrast, all female-biased microRNAs are located in autosomes, which is again the opposite observation to that which has been made for protein-coding genes. Figure 1d further explores the relationship between sex-chromosome location and sex-biased expression. The frequency distribution of fold change in expression for autosomal microRNAs shows three peaks, one large peak of unbiased expression and two smaller ones of male- and female-biased expression. However, the distribution of X-linked microRNAs is bimodal (figure 1d): they are either unbiased or highly expressed in males. Thus, male-biased microRNAs and the X chromosome are closely associated.

To further understand what it means to be sex-biased expressed, the expression profile of biased microRNAs was explored. Figure 2 plots a hierarchical tree of sex-biased expressed microRNAs and their relative expression levels in testes, ovaries and early embryos. Most male-biased microRNAs are highly expressed in the testes. This indicates that production of microRNAs in males is largely associated with the germline and the reproduction function. This is consistent with figure 1a in which adult samples were poorly correlated with young samples, perhaps because young individuals have not yet developed fully functional gonads.

### 3.3. Female-biased microRNAs are expressed in ovaries and early embryos

The expression profile in figure 2 shows that female-biased microRNAs fall into three distinct groups. First, a group of female-biased microRNAs are expressed in the somatic stem cells in the ovary, showing that microRNAs are important for the maintenance of stem cells in the ovary, in agreement with previous findings [30]. Second, some female-biased microRNAs are highly expressed in ovaries. This suggests that these microRNAs are important for the formation and maturation of *Drosophila* eggs.

Interestingly, a third group of female-biased microRNAs do not appear to be present in the ovary and they are highly expressed in young embryos (figure 2). These eggs were originally collected up to 1 h after laying [27,31], indicating that these young embryos have not yet started to have zygotic transcription [32]. This suggests that these microRNAs may be maternally deposited by the mother into the unfertilized eggs (oocytes). As a matter of fact, ongoing work in the laboratory has shown that the microRNAs mir-92a and mir-92b, and the mir-310/mir-311/mir-312/mir-313 cluster are abundant in *Drosophila* unfertilized eggs [33]. From these analyses, I conclude that both male- and female-biased microRNAs are mostly associated with the reproductive function.

### 3.4. Intronic female-biased microRNAs are associated with host gene expression

It may be possible that microRNA transcription pattern is associated with the transcription profile of their neighbouring protein-coding genes. In general, as shown in table 2, both expression patterns were not significantly associated (11 of 19 microRNA transcripts have the same expression bias as their closest neighbouring gene; \( p = 0.32 \), binomial test). In particular, there are eight microRNA transcripts with male-biased expression, and only three of their respectively closest genes show a similar bias. In the case of female-biased microRNA transcripts, eight
of 11 have their closest gene with a female-biased expression pattern. A closer inspection to the data reveals that this bias is produced mostly by microRNAs hosted within protein-coding genes (overlapping transcripts). Indeed, all six genes hosting microRNAs with female-biased expression are themselves expressed more highly in females than in males ($p = 0.016$). This shows that female-biased expression of microRNAs is highly associated with the production of microRNAs from introns of sex-biased expressed protein-coding genes.

### 3.5. Evolutionary origin of sex-biased microRNAs

There are two possible ways a gene may become sex-biased. First, a gene can acquire sex-biased expression. Second, a new gene appears (either de novo or by the duplication of an existing gene) having from the very beginning a sex-biased expression. Figure 3 shows the evolutionary origin of sex-biased microRNAs. Most male-biased microRNAs emerged within the *Drosophila* lineage, with only two exceptions: mir-993 and mir-283/304/12. These are indeed the least biased of all of the microRNAs. Thus, microRNAs with a strong male bias are evolutionarily young. By contrast, the evolutionary origin of female-biased microRNA families is diverse, and there are both old and young microRNAs. Among the old microRNAs, we have the mir-92, mir-184 and mir-9 families, which are conserved even in chordates. Interestingly, there are no *D. melanogaster*-specific microRNAs with a clear female-biased expression (contrary to the case of male-biased microRNAs). There are, however, two female-biased microRNAs which appeared in the *Drosophila* genus lineage: mir-314 and the mir-310–mir-313 cluster.

### 3.6. Targets of sex-biased microRNAs

Do sex-biased microRNAs also target sex-biased expressed gene transcripts? To explore this question, three different target prediction algorithms were used: TargetScan, miRanda and DIANA-microT (see §5). MicroRNAs were binned by their
bias level, and the expression bias of their targets was plotted in figure 4. These boxplots show that there is no tendency of sex-biased microRNAs to target sex-biased transcripts, at least not as a global pattern. I further explored the targets of seven melanogaster-subgroup-specific male-biased microRNAs. For two of them, two of the three prediction algorithms detected a significant association with sex-biased transcripts: mir-985 has a tendency to target female-biased genes, whereas mir-997 significantly targets male-biased genes (electronic supplementary material, table S2). The other associations were not significant and/or supported by only one prediction algorithm. Finally, I investigated whether recently emerged male-biased microRNAs also target evolutionarily young genes. I calculated the ratio between Drosophila-specific and conserved targeted genes for the targets predicted for the three above-mentioned algorithms. Four of seven studied microRNAs showed a tendency to target more conserved genes than expected by chance for at least two algorithms (electronic supplementary material, table S3), among them mir-985. The results here described rely heavily on target prediction algorithms and, therefore, should be taken with caution. However, they suggest that newly emerged microRNAs can potentially target conserved genes, altering regulatory relationships that have been conserved throughout evolution.

4. Discussion

In this study, I have shown that sex-biased microRNAs are mainly associated with the reproductive function: male micro-RNAs are expressed in testes, and female micro-RNAs are abundant in ovaries and oocytes. However, their evolutionary

Table 2. Sex-biased microRNA transcripts and their closest neighbouring genes.

<table>
<thead>
<tr>
<th>microRNA/cluster</th>
<th>fold change a</th>
<th>distance to closest gene b</th>
<th>closest gene</th>
<th>fold change a</th>
</tr>
</thead>
<tbody>
<tr>
<td>mir-972 – mir-979</td>
<td>−5.2</td>
<td>overlapping</td>
<td>Grip84</td>
<td>2.3</td>
</tr>
<tr>
<td>mir-982 – mir-984</td>
<td>−3.5</td>
<td>overlapping</td>
<td>CG3626</td>
<td>0.4</td>
</tr>
<tr>
<td>mir-959 – mir-964</td>
<td>−2.8</td>
<td>overlapping</td>
<td>CG31646</td>
<td>−2.6</td>
</tr>
<tr>
<td>mir-283 – mir-12</td>
<td>−1.1</td>
<td>overlapping</td>
<td>Gmap</td>
<td>−0.3</td>
</tr>
<tr>
<td>mir-985</td>
<td>−6.7</td>
<td>19344</td>
<td>disco</td>
<td>−1.5</td>
</tr>
<tr>
<td>mir-997</td>
<td>−5.3</td>
<td>737</td>
<td>D1</td>
<td>1.2</td>
</tr>
<tr>
<td>mir-2498 – mir-992</td>
<td>−4.7</td>
<td>817</td>
<td>CG32532</td>
<td>0.0</td>
</tr>
<tr>
<td>mir-993</td>
<td>−1.2</td>
<td>11072</td>
<td>Ama</td>
<td>2.0</td>
</tr>
<tr>
<td>mir-92a – mir-92b</td>
<td>4.6</td>
<td>overlapping</td>
<td>jigr1</td>
<td>2.6</td>
</tr>
<tr>
<td>mir-995</td>
<td>3.0</td>
<td>overlapping</td>
<td>cdc2c</td>
<td>3.9</td>
</tr>
<tr>
<td>mir-9c – mir-9b</td>
<td>2.2</td>
<td>overlapping</td>
<td>grp</td>
<td>3.1</td>
</tr>
<tr>
<td>mir-184</td>
<td>1.1</td>
<td>overlapping</td>
<td>CG44206</td>
<td>0.0</td>
</tr>
<tr>
<td>mir-308</td>
<td>1.0</td>
<td>overlapping</td>
<td>Rps23</td>
<td>1.2</td>
</tr>
<tr>
<td>mir-998 – mir-11</td>
<td>0.4</td>
<td>overlapping</td>
<td>E2f</td>
<td>1.4</td>
</tr>
<tr>
<td>mir-989</td>
<td>8.6</td>
<td>2739</td>
<td>Rd1</td>
<td>0.9</td>
</tr>
<tr>
<td>mir-994 – mir-318</td>
<td>8.3</td>
<td>249</td>
<td>Lrp-1B</td>
<td>−0.3</td>
</tr>
<tr>
<td>mir-310 – mir-313</td>
<td>4.1</td>
<td>1624</td>
<td>gsm</td>
<td>−0.7</td>
</tr>
<tr>
<td>mir-279 – mir-996</td>
<td>1.1</td>
<td>2892</td>
<td>Ef1gamma</td>
<td>1.4</td>
</tr>
<tr>
<td>mir-314</td>
<td>0.2</td>
<td>182</td>
<td>Tim13</td>
<td>−7.5</td>
</tr>
</tbody>
</table>

aLogarithm of fold change between male and female expression levels.
bIn nucleotides.
origin is different. Male-biased microRNAs tend to be evolutionarily young (dipteran/Drosophila-specific; figure 3) and they often emerge in the X chromosome. Contrary to microRNAs, male-biased protein-coding genes appear to be generally under-represented in the X chromosome in flies [2,3], and a movement of male genes out of the X, or demasculinization of the sex chromosome, has been suggested [2,20]. However, novel genes tend to be X-linked, and male-expressed and older genes may have moved outside the X chromosome [23,17,21]. An enrichment in the X chromosome for microRNAs with male-biased expression has also been reported in mammals [12,34–36].

A careful dissection of the evolutionary origin of male-biased genes in Drosophila demonstrated that de novo originated genes tend to be X-linked and male-biased, and that there may also be an ongoing demasculinization process in the X chromosome [23]. In addition the study suggested that this demasculinization may also be happening in microRNAs. Their analysis showed that there is about a 12-fold enrichment of evolutionarily young microRNAs in the X chromosome [23,17,21]. An enrichment in the X chromosome for microRNAs with male-biased expression has also been reported in mammals [12,34–36].

There is an ongoing debate in the scientific literature about sex chromosome demasculinization. Although demasculinization has been generally considered one of the prominent features of Drosophila X chromosome evolution, recent work shows that the observed paucity of male-biased genes in the X chromosome may be artefactual [37–39]. Indeed, several groups suggest that demasculinization does not happen in Drosophila and propose that there is no global meiotic sex chromosome inactivation (MSCI) [40,41]. The movement of male-biased genes out of the X chromosome is often explained as a response to MSCI. This discussion is not settled, and evidence both for demasculinization and for MSCI is still reported [23,42–44]. Interestingly, most X-linked microRNAs escape MSCI [45]. These observations imply that X-chromosome demasculinization caused by MSCI might not happen during microRNA evolution. Even if there is an ongoing demasculinization process affecting protein-coding genes, microRNAs seem not to be affected.

Female microRNAs are generally older than male-biased microRNAs, and they are frequently encoded within other female-expressed genes. For instance, mir-995 is highly expressed in females (figure 1) and it is associated with oocytes (figure 2). This microRNA is encoded within the first intron of cdc2c, a gene involved in cell proliferation during development [46]. Hence, the presence of mir-995 in oocytes may be a by-product of the host gene expression. In addition, mir-995 can be identified in the same intron of the orthologous cdc2c gene in other insects [47], showing a deep conservation of the microRNA/host gene association. Interestingly, mir-92a is encoded within a gene (jig1) whose product is maternally deposited in the oocyte [48], and the microRNA is highly expressed in oocytes (figure 2) and detected in unfertilized eggs [33]. The presence of mir-92a in the developing egg may be a by-product of being intrinsic to a sex-biased expressed gene. As a matter of fact, mir-92a is associated with leg morphological differences between Drosophila species [49], a role (in principle) unrelated to any function in the early developing egg.

Among the microRNAs with a female-biased expression pattern, there are microRNAs associated with the gametic function. Recently, mir-989 has been discovered to be involved in cell migration in the ovary [50]. Indeed, the 3' arm of mir-989 is highly expressed in ovaries (figure 2). The analysis of female mutants also reveals that mir-9c (present in ovaries; figure 2) is somewhat involved in the control of the number of germ cells [51]. Predictably, other female-biased microRNAs here reported, such as mir-994/318, could have a role in gametic function. Strikingly, the mir-310/311/312/313, which is somewhat involved in the control of the number of germ cells [51], is involved in the development of male gonads [52]. This emphasizes that genes with sex-biased expression can also have other functions, even in the opposite sex.

We recently characterized sex-biased microRNAs in the parasitic Schistosoma mansoni and reported that one of the microRNA clusters (mir-71/mir-2) has two copies, one in the sexual chromosome with no detectable bias and another copy in an autosome with sex-biased expression. The duplication of the cluster happened more or less at the same time as sexual dimorphism appeared in this genus (Schistosoma). We suggested that this may be a case of escaping sex conflict, in which genes involved in sex dimorphism tend to be out of the X chromosome [10,53]. However, this is likely to be an exception to the rule in microRNAs, as their evolutionary dynamics is primarily dominated by high levels of emergence and a low probability of non-tandem duplication.

In summary, I conclude that sex-biased expression of microRNAs is a consequence of a high rate of microRNA de novo emergence. Novel microRNAs tend to appear in the X chromosome and to be expressed in the testes. Conversely, male-biased microRNAs are evolutionarily young and also show a high rate of loss. On the other hand, many
female-biased microRNA emerged within the intron of female-biased host genes. They are generally conserved suggesting that the female gametic function may be more constrained, and purifying selection could eliminate emerging microRNAs impairing ovary/oocyte development. This scenario suggests that positive/adaptive selection may have no more than little contribution to determining the sex-biased expression of microRNAs.

5. Methods

*Drosophila melanogaster* microRNA sequences are from miRBase version 19 [54]. Expression datasets were downloaded from Gene Expression Omnibus at http://www.ncbi.nlm.nih.gov/geo/, with accession numbers: GSM286602 and GSM399107 (adult males); GSM286603 and GSM399106 (adult females); GSM280082 (ovaries); GSM909277 and GSM909278 (testes); GSM385822 and GSM385744 (ovary somatic sheet); GSM180330 and GSM286613 (early embryos); GSM692232 and GSM69224 (young males and females) [31, 55–59]. Reads from these experiments were mapped to *D. melanogaster* microRNA hairpins with Bowtie v. 0.12.7 [60], allowing no mismatches nor multiple matches. Differential expression of microRNAs was estimated with EDGER [61]. In short, read matches nor multiple matches. Differential expression of the same. Expression data for implemented in DESEQ [64], but the results remained overall by an exact test controlling the false discovery rate [63]. Addition-M-values (TMM) method [62]. Then, the variation within counts were first normalized with the trimmed mean of genes were obtained with modENCODE, available at www.flybase.org [58, 65]. All statistical analyses and figures 1 and 2 were done with R [66]. Evolutionary age of microRNAs and microRNA families was estimated as previously described [67]. In brief, I compiled microRNA sequences with detectable similarity to *D. melanogaster* microRNAs with BLAST [68], using a sensitive set of parameters to detect homologous microRNAs (−w 4, −q −3, −r 2). I also included additional sequences described elsewhere to ensure that all known microRNA families with a common evolutionary origin are taken into account [54, 67, 69–71]. MicroRNA hairpins were aligned with CLUSTALX v. 2.0 [72], manually refining the alignments with RALEEE [73], and phylogenetic trees were reconstructed using the neighbour-joining and maximum-likelihood routines with default parameter as implemented in MEGA5 [74]. MicroRNA age estimates were also compared with those obtained by Mohammed et al. [75] using a different approach: they analysed whole genome alignments of 12 *Drosophila* genomes [76]. Age estimations were fully congruent between both datasets. MicroRNA targets were retrieved from our previous study [77]. In short, 3′-UTRs were downloaded from FlyBase (genome version BDGP 5.25), and the microRNA targets were predicted with three different programs based on different algorithmic approaches: TargetScan [78], DIANA-microT [79] and miRanda [80], with default parameters. The evolutionary conservation of targeted genes was inferred from the gene family tree available at TreeFam 9 [81].

Acknowledgements. I am greatly indebted to Steve Dorus, Kirill Borziak and two anonymous reviewers for useful comments and critical reading of the manuscript. I also thank Maria Vibrationoki and Colin Meiklejohn for pointing me to relevant papers, and to Haldane’s Sieve (http://haldanesieve.org) for promoting the dissemination and discussion of preprints (including the arXiv version of this work).


