No accelerated rate of protein evolution in male biased *Drosophila pseudoobscura* genes

Muralidhar Metta, Rambabu Gudavalli, Jean-Michel Gibert & Christian Schlötterer

Institut für Tierzucht und Genetik, Veterinärmedizinische Universität Wien, Josef Baumann Gasse 1, 1210 Wien, Austria
Running head: Evolution of male biased genes in *D. pseudoobscura*

Key words: male bias, SAGE, mutation rate, selection, gene expression

Corresponding author:

Christian Schlötterer

Institut für Tierzucht und Genetik

Josef Baumann Gasse 1

1210 Wien

Austria

Tel.: +43-1-25077-5603

Fax: +43-1-25077-5693

christian.schloetterer@vu-wien.ac.at
ABSTRACT

Sexually dimorphic traits are often subject to diversifying selection. Also genes
with a male biased gene expression are probably affected by sexual selection and have a
high rate of protein evolution. We used SAGE to measure sex biased gene expression in
*Drosophila pseudoobscura*. Consistent with previous results from *D. melanogaster*, a
larger number of genes were male biased (402 genes) than female biased (138 genes).
About 34% of the genes changed the sex related expression pattern between *D.
melanogaster* and *D. pseudoobscura*. Combining gene expression with protein
divergence between both species, we observed a striking difference in rate of evolution
for genes with a male biased gene expression in one species only. Contrary to
expectations, *D. pseudoobscura* genes in this category showed no accelerated rate of
protein evolution, while *D. melanogaster* genes did. If sexual selection is driving
molecular evolution of male biased genes, our data imply a radically different selection
regime in *D. pseudoobscura*. 
INTRODUCTION

Females and males often exhibit substantial differences in behavior, physiology and morphology. Although only a small number of genes contribute to the first steps of sexual differentiation, a significant proportion of the genome is involved in the establishment and maintenance of sexual dimorphism (Hughes 2001). It is well established that such sexually dimorphic traits are often subject to directional selection (Luck and Joly 2005; Snook et al. 2005). While initial studies focused mainly on morphology, behavior and physiology (Dickens et al. 1998; Holland and Rice 1999; Hoy 1990), more recently the attention has turned to pattern of molecular evolution (Meiklejohn et al. 2003; Ranz et al. 2003). Several protein involved in male reproductive function are shown to exhibit accelerated rate of molecular evolution, presumably due to positive selection driving amino acid replacements (Begun and Lindfors 2005; Coulthart and Singh 1988; Tsaur and Wu 1997; Wagstaff and Begun 2005). However, some of these genes may also evolve rapidly due to reasons other than positive selection (Civetta et al. 2006; Rooney et al. 2000; Torgerson and Singh 2004). Interestingly, recent studies in Drosophila and mice showed that this pattern could be generalized to genes that are more highly expressed in the male than in the female (male biased genes) (Good and Nachman 2005; Swanson et al. 2001; Zhang et al. 2004). In mice, however, the accelerated rate of evolution seems to be restricted to genes with a male biased gene expression during late spermatogenesis (Good and Nachman 2005).

An important assumption that has been implicitly made for all studies determining the rate of molecular evolution is the maintenance of function/gene expression across
species. Interestingly, a comparative analysis of gene expression in *D. melanogaster* and *D. simulans*, which diverged less than 3 million years ago, indicated that for a significant number of genes (about 20%) the sex bias in gene expression has been lost or even reversed between these two species (RANZ et al. 2003). Thus, the interpretation of the observed high rate of sequence evolution of genes with a male biased gene expression in one species is less clear. It was proposed that the accelerated rate of protein evolution of male biased genes is caused by sexual selection driven either by male-male competition, male-female interaction or both (SWANSON and VACQUIER 2002). Nevertheless, to rule out other evolutionary forces, such as a rapid rate of protein evolution associated with a shift in function/gene expression, it is important to determine rates of sequence evolution in combination with patterns of gene expression in more than one species.

We performed an analysis of sex biased gene expression in *D. pseudoobscura*, the second Drosophila species for which a completely sequenced genome is available (RICHARDS et al. 2005). Combining sequence and expression data, we provide evidence that the accelerated rate of protein evolution is restricted to *D. melanogaster*, while *D. pseudoobscura* male biased genes evolve at lower rates, comparable to non-biased genes.

**MATERIAL & METHODS**

**Serial Analysis of Gene Expression (SAGE) library**

Total RNA was extracted with Trizol from 220 mg males and 370 mg females of *D. pseudoobscura* (obtained from M. Akam Lab). Only virgin individuals from both sexes between three and seven days old were used. The SAGE library construction
followed published protocols (VELCULESCU et al. 1995) at the EMBL core facility. In brief, the mRNA was isolated using a biotin-labeled Tₐ oligo and Dynal beads. Double stranded cDNA was generated with Superscript II reverse transcriptase and DNA polymerase I. The resulting cDNA library was digested with NlaIII (anchoring enzyme). 3’ restriction fragments were recovered with Dynal beads and split into two populations. Each aliquot was ligated to one of the two annealed linker pairs. After extensive washing tags were created by digestion with BsmFI (tagging enzyme). Tags were blunted with T4 polymerase and ligated to generate ditags. Ditags were PCR amplified and digested with NlaIII. Restriction fragments were separated on a 12% polyacrylamide gel and the band containing the ditags was isolated from the gel. The purified inserts were concatemerized, ligated into pZero and transformed into EDH5α electromax cells. Similar to other SAGE experiments in Drosophila (JASPER et al. 2002; JASPER et al. 2001), our design was aiming for approximately 20,000 tags per library.

**D. melanogaster/D. simulans expression data**

The *D. melanogaster* genes in our analyses were classified as male, female or unbiased based on the gene expression data for adult males and females reported in PARISI et al (2003). We selected this data set, as it represents the largest set of genes and the adult flies covered a similar age range as our study (5-7 days). Furthermore the expression data from (RANZ et al. 2003) and (ARBEITMAN et al. 2002) were underrepresented for male biased genes. Expression intensity and sex bias were determined by averaging replicate experiments. *D. simulans* expression intensity and sex bias used in this study were obtained from RANZ et al. (2003) as provided in the online supplement. If not noted otherwise, we used a two-fold expression difference between
males and females to define a sex biased gene expression. To classify genes according to expression level (i.e.: identifying highly expressed genes) in *D. melanogaster*, we considered the used absolute intensity values (after background subtraction) in the two channels from microarray experiments of PARISI *et al.* (2003). As multiple experiments in adult flies were performed and some genes were represented more than once on each array, we averaged the absolute intensity values of each gene across sex, experiments and duplicate spots. Based on this average we classified genes according to their expression level. We note that the intensity on a two color array is not an optimal estimator of expression level (TOWNSEND 2003), nevertheless the misestimates introduced may just result in greater noise. Hence, we do not expect that our procedure result in a strong bias of our results.

**Data analysis**

**SAGE tag extraction:** Vector sequence was removed using an in house perl script. Duplicated clones were identified and removed with the xmatchdt.pl software (DINEL *et al.* 2005). 14bp SAGE tags (10bp + CATG, the NlaIII recognition sequence) were extracted with the SAGEparser software (DINEL *et al.* 2005). This software also screens for overrepresented ditags and corrects for this PCR artifact. SAGE data were submitted to GEO (GSM60839, GSM60840).

**Comparison of male and female SAGE library:** Sequencing errors could inflate the number of unique tags. Therefore, we only used tags that were identified more than once in the full data set containing male and female library. This strategy should not make difference to the results, as unique tags would not be considered as a significant difference between both libraries. The tags from the male and female library were
compared using a two-tailed Fisher’s exact test as implemented in online SAGE analysis tools (http://www.mbgproject.org/MBGP_Tools.html). Consistent with previous reports (RUIJTER et al. 2002), we obtained almost identical results when we used the Audic-Claverie statistics (AUDIC and CLAVERIE 1997) as implemented in Discoveryspace 3.2.3 (http://www.bcgsc.ca/bioinfo/software/discoveryspace/), thus we only report results based on Fisher’s exact test. We expressed the sex bias by the log₂ ratio of male to female tags. In those cases for which tags were detected in one sex only, we set the tag count to one for the other sex, in order to avoid problems in calculating the log₂ ratio.

Tag to gene mapping: We used the complete transcriptome from the D. pseudoobscura annotation 1.0 to build a database consisting of the sequence between the 3’ end of the transcript and the most 3’ NlaIII recognition sequence. Transcripts without NlaIII recognition site were not considered. This database was used to match the SAGE tags. Only perfect and unique matches were considered.

Gene conservation: We used a pre-compiled list of D. melanogaster genes that are conserved in D. pseudoobscura, which is available from Flybase upon request, to link sex biased gene expression with gene conservation.

Functional enrichment of SAGE libraries: To test whether male/female biased tags in our SAGE library are enriched for genes with sex specific function, we used GeneMerge software, which relies on Gene Ontology (GO) terms for a functional classification of genes (CASTILLO-DAVIS and HARTL 2003).

Substitution rates, codon usage, and Grantham’s distance: D. melanogaster transcripts from release 4.0 and D. pseudoobscura transcripts from release 1.0 were aligned using the gene prediction software GeneWise (BIRNEY and DURBIN 2000) and
protein sequences from *D. melanogaster* (release 4.0). We used an in house perl script to parse the GeneWise output and generate aligned fasta files. The rates of non-synonymous substitutions (\(d_N\)) were calculated from pairwise alignments of *D. melanogaster* - *D. pseudoobscura* genes using Yang & Nielsen method (*YANG* and *NIELSEN* 2000) as implemented in the program yn00 in PAML 3.14 (*YANG* 1997). We used \(d_N\) rather than the ratio of non-synonymous to synonymous substitutions (\(d_N/d_S\)), as synonymous substitutions are saturated due to the high divergence of the two species (*RICHARDS et al.* 2005). Given that the time of divergence for all genes is the same (within the limitations of lineage sorting) and Drosophila shows no male biased mutation rates (*BAUER* and *AQUADRO* 1997), non-synonymous substitution rates are sufficient to compare rates of protein evolution (*CUSACK* and *WOLFE* 2005). We note, however, that genes with a male bias in *D. melanogaster* show a trend towards a slightly higher \(d_S\) than other genes (*JAGADEESHAN* and *SINGH* 2005; *ZHANG et al.* 2004). The magnitude of the effect in our analysis is, however, too large to explained by a variation in \(d_N\). Furthermore, we would like to point out that the elevated \(d_S\) may be a result of recurrent selective sweeps (*BETANCOURT* and *PRESGRAVES* 2002), hence the \(d_N/d_S\) ratio may be overly conservative. The level of codon usage bias was measured as frequency of optimal codons (\(F_{op}\)) (*IKEMURA* 1981) and was calculated using the program codonW (*PEDEN* 1999). Rather than comparing the codon usage for each of the species separately, we focused our analysis on the difference in codon usage, as this conveniently combines the data from both species into a single statistic. Grantham distance was used as measure to calculate the amino acid similarity and was calculated by using in house perl scripts, implementing the Grantham matrix (*GRANTHAM* 1974).
RESULTS

We generated two SAGE libraries for *D. pseudoobscura*, one for males and one for females. Each library contained at least 23400 tags. Conservatively, we restricted our analyses to tags that were identified more than once in the two libraries, resulting in 3618 and 3481 unique tags (genes) for the male and female library, respectively. The total number of unique tags from both libraries was 4450, which corresponds to 42.3% of the annotated genes in *D. pseudoobscura*. Consistent with the gene expression pattern in *D. melanogaster*, we detected more male biased genes (402) than female biased ones (138) and this difference was robust to changes in the significance thresholds (Fig. 1 and 2). The variance in gene expression among female biased genes was significantly higher than in male biased genes (*P* < 0.001, Levene’s test). This contrasts with *D. melanogaster*, for which a higher variance was observed for male biased genes expressed in adults (Parisi *et al.* 2003).

We used the release 1.0 of the *D. pseudoobscura* genome annotation (Richards *et al.* 2005) to map the SAGE tags of both libraries to the corresponding genes (see Material & Methods for more details). Consistent with previous SAGE experiments (Lee *et al.* 2005; Pleasance *et al.* 2003), a significant proportion of the SAGE tags (e.g. 45-66% in *D. melanogaster* [Jasper *et al.* 2002; Jasper *et al.* 2001; Lee *et al.* 2005]) could not be unambiguously mapped to a gene. On average, we mapped 1183 (27%) of the SAGE tags to a single *D. pseudoobscura* gene. Interestingly, the proportion of genes that could be mapped differed between male biased, female biased and unbiased genes. Only 19 (14%) of the female biased tags could be mapped, while 32% and 27% were mapped
for male biased and unbiased tags, respectively. The difference between male and female biased tags that could be mapped is highly significant (P< 0.001, Fisher’s exact test).

After mapping the SAGE tags to *D. pseudoobscura* genes, we identified genes for which the available information was consistent with a sex biased expression pattern (Supplementary Table 1). The female SAGE library contained several genes with a typical female specific function. We used GeneMerge for a formal test of significant overrepresentation and found the GO terms oogenesis, insect chorion formation, vitelline membrane formation and protein biosynthesis to be overrepresented (e<0.05). The male library contained several ubiquitin and microtubules associated genes. Nevertheless, GeneMerge did not find a significant overrepresentation of these classes.

We further corroborated our tag to gene mapping using the a method called Generation of Longer cDNA fragments from serial analysis of gene expression tags for Gene Identification (GLGI) to generate longer cDNA fragments based on the poly A-tail and the tag sequence (CHEN et al. 2000). Out of 21 tags haphazardly selected from different classes of sex bias, 20 identified the correct gene. One tag resulted in a cDNA sequence that could not be identified in the entire *D. pseudoobscura* genome. Hence, we estimate the fraction of inaccurately mapped tags to fall into the interval of 0.001 to 0.237 (95% confidence interval, binomial distribution). Overall, these results indicate a reasonable accuracy of our tag to gene mapping.

Contrary to the case in *D. melanogaster*, the ratio of male biased genes to unbiased genes did not differ between X-chromosomes and autosomes (P= 0.393, Fisher’s exact test) in *D. pseudoobscura*. Furthermore, no pronounced difference could be detected between chromosomes XL and XR (which corresponds to 3L in *D.*
melanogaster, see Fig. 3A & 3B). In total, 1102 genes with expression data for males and females in both species, D. melanogaster and D. pseudoobscura, were available. About 34% of the genes changed the sex bias of the expression pattern between both species. Eight classes of gene expression could be distinguished (Table 1). We measured the rate of evolution by the proportion of non-synonymous substitutions. Consistent with an accelerated rate of evolution of male biased genes, we found the highest rate of non-synonymous substitutions for genes with a male biased gene expression in both species. Genes with a male bias in D. melanogaster and no sex bias in D. pseudoobscura also showed a high rate of sequence evolution. The complementary group of genes with male biased gene expression in D. pseudoobscura and no sex bias in D. melanogaster, however, did not have a high rate of non-synonymous changes. The difference between these two groups was statistically significant (P < 0.001, Mann-Whitney U test). Overall, female biased genes showed a trend towards a lower rate of protein evolution, but the sample sizes were too small to provide robust estimates. Interestingly, a similar pattern emerged when we used the dN/dS ratio rather than dN. The difference between D. melanogaster and D. pseudoobscura for genes with a male expression pattern in one species only was also highly significant based on the dN/dS ratio (P < 0.001, Mann-Whitney U test). We also note that the differences in rate of evolution remain when more stringent criteria were used to define sex-biased genes (Fig. 4).

We further substantiated the different evolutionary behavior of genes with a male biased expression pattern in D. melanogaster by calculating Grantham’s distance (GRANTHAM 1974) between D. melanogaster and D. pseudoobscura genes (Table 1). Grantham’s distance considers the carbon-composition, polarity and volume of amino
acids to determine their exchangeability. Thus, a higher Grantham distance indicates more radical amino acid changes. Our results of the Grantham distances nicely parallel the pattern of non-synonymous substitutions. Highest Grantham’s distances were found for genes with a male biased gene expression in both species. Again, genes with a male biased gene expression in *D. pseudoobscura* and no sex bias in *D. melanogaster* had a significantly lower Grantham distance (P=0.019, Mann-Whitney U test) than the complementary group of genes with no sex bias in *D. pseudoobscura* and male bias in *D. melanogaster*.

We also determined the influence of gene expression on codon usage by comparing the frequency of the optimal codon (*F_{op}* between the two species. Consistent with the previously reported similarity in codon usage between *D. melanogaster* and *D. pseudoobscura* (Moriyama and Powell 1997), unbiased genes have a very similar optimal codon frequency (Table 1). Pronounced differences, however, were found for those genes with a male biased gene expression pattern in *D. melanogaster* (Table 1). Comparing the difference in codon usage (*F_{op (pseudoobscura)} - F_{op (melanogaster)}*) between the two categories of genes with male biased gene expression in one species, but no sex bias in the other species, we detected a significant difference (P<0.001, Mann-Whitney U test) with only *D. melanogaster* male biased genes showing a lower codon usage bias.

We calculated Cohen’s *d* (Cohen 1988) to know whether the effect size is sufficiently large that the statistical significances reported are biologically relevant. A moderate effect size for Grantham’s distance (*d* = 0.3) and high effect size for codon bias (*d* = 0.47) were observed. These results suggest that about 21% of the observations of
Grantham’s distance and 33% of the observations of codon bias are not overlapping (Cohen 1988).

Rates of evolution were based on all genes to which we could unambiguously map a SAGE tag. As these genes may not be completely conserved between D. melanogaster and D. pseudoobscura we repeated our analyses using only those genes for which we could align the complete protein between both species, but the same trends were observed. Furthermore, to exclude that a differential coverage of low expressed genes in the SAGE and microarray experiments affected our interpretation of the results, we repeated the analysis using only the 20% and 40% most highly expressed genes from the microarray experiments in D. melanogaster. However, the same trend was observed (data are available from authors upon request).

We also compared D. melanogaster data from Ranz et al (2003) with D. pseudoobscura expression data, to see whether similar pattern can be observed. About 50% of the genes that are included in the analysis with Parisi et al (2003) are missing in this comparison. Comparing genes with a male biased expression in D. pseudoobscura and unbiased expression in D. melanogaster (12 genes rather than 77 genes in the Parisi et al (2003) data) and the complementary class (161 genes rather than 187 genes in the Parisi et al (2003) data), the later class manifests higher non-synonymous substitution rates (0.13 against 0.07). However, this difference is lacking sufficient statistical power, which may be due to smaller sample sizes.

The XR chromosome of D. pseudoobscura is originated by translocation from an autosome (3L) in D. melanogaster. Hence, we were interested if we could detect evidence for different rates of evolution for genes located on XL, XR and autosomes.
Nevertheless, within the limitations of a moderate number of genes, we observed similar rates of evolution within each class of sex-biased genes across the chromosomes (data are available from authors upon request).

DISCUSSION

In this study we compared both sequence and expression data for males and females in two different Drosophila species. We measured gene expression in *D. pseudoobscura* using SAGE and compared it to microarray based expression data in *D. melanogaster*. While the comparison of gene expression data generated by different experimental systems could potentially lead to some bias, we note that SAGE and microarray data were found to be highly correlated (ISHII et al. 2000). Furthermore, we do not compare the expression data directly, but we compare the gene expression differences between males and females. Thus, the relative gene expression is determined for each species separately using a consistent experimental design, which further minimizes the risk of a possible bias (DRAGHICI et al. 2006). Assuming that both SAGE and microarray experiments measure the expression level accurately, the only possible bias could arise from genes with a low expression level that are not included in the SAGE data. We have therefore repeated our analysis using only those genes that are highly expressed in *D. melanogaster*, but obtained the same results. Also a different significance level for identifying sex biased genes in *D. pseudoobscura* did not change the overall picture (Fig. 1). Hence we are confident that the comparison of SAGE and microarray data did not bias our results.
Impact of tag to gene mapping on the comparison of male and female gene expression

SAGE allows the analysis of sex biased gene expression independent of an annotated genome. SAGE provides a set of tags, which are overrepresented in males or females. A genome annotation is, however, required when these tags should be associated with the corresponding gene (tag to gene mapping). Thus, the proportion of SAGE tags mapped depends on the accuracy of genome annotation. Three major factors could affect the efficiency of our tag to gene mapping: 1) as the annotation of *D. pseudoobscura* relied heavily on genes known in *D. melanogaster* the sequence conservation of *D. melanogaster* genes in *D. pseudoobscura* is a major determinant of the tag to gene mapping efficiency 2) many SAGE tags are located in the 3’ UTR (Pleasance et al. 2003), but this region is very poorly annotated in *D. pseudoobscura* 3) genes specific to *D. pseudoobscura* and not yet discovered genes in *D. melanogaster*.

Our analyses indicated that a lower proportion of female biased tags could be mapped to *D. pseudoobscura* genes than unbiased or male biased tags. In the following we discuss to what extent this result could be an artifact of the incomplete genome annotation available for *D. pseudoobscura*.

In principle, the low tag to gene mapping efficiency of female biased genes could be the result of a lower sequence conservation of female biased genes. We tested this hypothesis by determining the proportion of *D. melanogaster* genes that are conserved in *D. pseudoobscura*. We found that 83 % of the genes with a female biased gene expression in *D. melanogaster* were conserved in *D. pseudoobscura*. 79 % of the unbiased and only 63 % of the male biased genes were conserved. Hence, based on *D.
*melanogaster* gene expression data a lower proportion of male biased genes are expected, but not a lower proportion of female biased genes.

The gene expression data from *D. melanogaster* and *D. pseudoobscura* are based on different methods. In contrast to the microarray data for *D. melanogaster*, the dynamic range of a SAGE analysis depends on the number of tags sequenced. As the depth of our SAGE sequencing was not extensive, we probably covered only the more highly expressed genes. To exclude that the differential coverage of low expressed genes affected our results, we repeated our analyses restricting the microarray data to the 20%/40% most highly expressed genes in *D. melanogaster*. Nevertheless, we still found approximately the same degree of conservation. Hence, we conclude that the lower efficiency of tag to gene mapping for female biased tags in *D. pseudoobscura* is not an artifact of a different dynamic range of SAGE and microarray data.

Alternative splicing could result in transcripts with different SAGE tags. If genes with a female biased gene expression in *D. pseudoobscura* are more frequently alternatively spliced, this could result in a lower tag to gene mapping efficiency. We explored this possibility and determined the proportion of genes that would result in different SAGE tags due to alternative splicing. However, only a very small proportion of alternatively spliced genes in *D. melanogaster* (4%) yielded alternative SAGE tags. No significant difference was detected between male and female biased genes. Unless this pattern has dramatically changed in *D. pseudoobscura*, we could discount this explanation for the low efficiency of tag to gene mapping for female biased genes in *D. pseudoobscura*. 
Tag to gene mapping is also affected by polymorphism (Ng et al. 2005). As we used a different *D. pseudoobscura* strain than the sequenced one, it is possible that sequence polymorphism reduces the tag to gene mapping efficiency. The lower efficiency to map female biased tags to *D. pseudoobscura* genes could be explained if female biased genes have more sequence polymorphism than male biased ones. Currently, too few polymorphism data are available in *D. pseudoobscura* to address this question.

Tag to gene mapping may be affected by the incomplete 3’UTR annotation in *D. pseudoobscura*. If female biased genes have a longer 3’UTR than male biased genes, this could lead to an under-representation of female biased genes. We compared the length of 3’UTRs in *D. melanogaster* and found that male biased genes on average have shorter 3’UTRs (293bp) than female biased genes (333 bp). The longest 3’UTRs, however, were observed for the unbiased genes with 433 bp on average. As we mapped a higher proportion of unbiased tags than female biased, we conclude that UTR length is unlikely to be the sole reason for the low efficiency of tag to gene mapping of female biased SAGE tags.

Finally, a high proportion of hitherto unrecognized genes could explain the low tag to gene mapping efficiency for female biased genes. Either the genes have not yet been annotated in *D. melanogaster* (e.g.: (Lee et al. 2005)) or they are present in *D. pseudoobscura* only. In depth sequencing of cDNA clones obtained from female *D. pseudoobscura* will be highly instrumental to address this question further.

We further tried estimating the error rate in tag to gene mapping. Given the limitations of the current genome annotation of *D. pseudoobscura*, it is possible that a certain fraction of the tags were mapped to the wrong gene. If male biased genes in *D.
pseudoobscura would evolve at a high rate, similar to D. melanogaster, the lower \( d_N \) value in our data could be caused by an erroneous inclusion of non-male biased genes.

We estimated the proportion of falsely assigned genes (with a lower \( d_N \) value) by iteratively removing the gene with the lowest \( d_N \) value from the genes that were male biased in D. melanogaster, but unbiased in D. pseudoobscura. Fig. 5 shows that about 80% of the genes would need to be wrongly assigned to obtain a similar \( d_N \) value as for genes with male biased gene expression in both species.

Finally, the best support for the accuracy of our tag to gene mapping procedure is provided by the experimental validation using the GLGI method: only one out of 21 tags analyzed did not confirm our tag to gene mapping. Hence, we conclude that errors in tag to gene mapping could not explain our results.

**Stability of male biased gene expression**

We found that about 34% of the genes changed the sex related expression pattern between D. melanogaster and D. pseudoobscura. A recent comparison between D. melanogaster and D. simulans found a very similar proportion of genes with a change in expression pattern between males and females despite a much closer phylogenetic relationship (RANZ et al. 2003). Hence, it may be possible that the sex related expression of several genes changed multiple times. As male biased genes were particularly instable (RANZ et al. 2003), we were interested if genes with a male biased gene expression in D. melanogaster and D. pseudoobscura maintained their sex bias during evolution. We used the limited set of genes, for which expression data are available in three species, D. melanogaster, D. simulans and D. pseudoobscura. Interestingly, all seven genes with a male biased gene expression in D. melanogaster and D. pseudoobscura were also male
biased in *D. simulans* (Table 2A). Nevertheless, we would like to point out that we used a different criterion to define male bias (2-fold difference in expression), than Ranz et al (2003). Using the significance threshold of Ranz et al (2003), we found that one out of 21 genes with a male biased gene expression in *D. melanogaster* and *D. pseudoobscura*, was not male biased in *D. simulans* (Table 2B). Given that about 20% of the male biased genes in *D. melanogaster* changed their expression pattern in *D. simulans*, this high concordance may suggest that only a subset of the male biased genes changes its expression pattern, while others remain male biased. The analysis of sex biased gene expression in further species will be instrumental to determine their evolutionary stability.

**No increase in d\textsubscript{N} by change in sex biased gene expression**

An analysis of sex biased gene expression in *D. melanogaster* and *D. simulans* indicated a rapid change in expression pattern, in particular for male biased genes (Ranz et al. 2003). Assuming that such changes in expression pattern reflect functional shifts, it is interesting to determine if a change in gene expression is associated with an increased rate of protein evolution. Two different evolutionary scenarios would predict an accelerated rate of evolution: 1) A change in gene expression relaxes the functional constraint leading to more amino acid replacements. 2) The change in gene expression is associated with an expansion/modulation of the functional repertoire. Thus, directional selection is expected to drive the required changes in the protein.

Of the possible changes in expression pattern, three categories contained more than 50 genes; two categories containing male biased – unbiased genes and one category with a female biased gene expression in *D. melanogaster* but unbiased in *D.*
pseudoobscura. Contrary to expectations, none of the categories with a change in gene expression had a higher d\textsubscript{N} than genes with a male biased gene expression in both species. In fact, genes with a female biased gene expression in *D. melanogaster* and no sex bias in *D. pseudoobscura* had a d\textsubscript{N} lower than all categories with no change in gene expression pattern. This indicates that a change in sex biased gene expression pattern does not necessarily increase the non-synonymous substitution rate.

**Male biased genes in *D. pseudoobscura* do not evolve fast**

It is well documented that genes with a male biased gene expression have an accelerated rate of protein evolution ([Meiklejohn et al. 2003](#)). Nevertheless, most of the inference is based on the expression pattern in one of the species only. Consistent with the previous results, we also found that genes in *D. melanogaster* and *D. pseudoobscura* with a male biased gene expression were evolving at the highest rate in our data set. Nevertheless, the comparison to genes that changed their expression pattern is not consistent with an unconditionally higher rate of protein evolution for male biased genes. If male biased genes would evolve faster, genes with a change in expression should have an intermediate rate of evolution. Moreover, this pattern should be independent of which species shows the male bias. Genes with a male biased gene expression in *D. melanogaster* only evolve fast and with a very similar rate as genes that are male biased in both species. However, contrary to expectation, those genes with a male biased gene expression in *D. pseudoobscura* only are not evolving fast and have a similar rate of protein evolution as unbiased genes. This pattern is better explained by a change in the evolution rate of male biased genes between *D. pseudoobscura* and *D. melanogaster*. 
It was previously suggested that the accelerated rate of evolution of male biased genes is driven by positive selection (Swanson and Vacquier 2002). Consistent with this hypothesis, we also found a higher Grantham distance and a lower codon usage for genes with a male biased gene expression pattern in both species. Interestingly, genes with a male biased gene expression in *D. pseudoobscura* only did not show this effect, further emphasizing the difference of male biased genes between *D. melanogaster* and *D. pseudoobscura*. Nevertheless, we note that relaxed constraint, rather than directional selection, would result in a similar pattern for the male biased genes in *D. melanogaster*.

Our analyses were based on genes that are conserved between *D. melanogaster* and *D. pseudoobscura*. Thus, rapidly evolving genes will be missed in our analysis. We note that it is not possible to estimate the proportion of those rapidly evolving genes that are still functional in *D. pseudoobscura* and whether or not they have changed their sex bias. For this reason we focused our attention to those genes that are male biased in one species only. Our analysis is expected to be unbiased unless we would have preferentially missed fast evolving genes with a male bias in *D. pseudoobscura*. However, we do not think that this applies to our data. Assuming that genes with a male biased gene expression evolve so fast that they are difficult to detect in *D. pseudoobscura*, this would result in a downward bias in mutation rate affecting all genes with a male expression bias. Hence, the most dramatic effect would be seen for genes with a male biased gene expression in both species. Genes with a male bias in only one of the species are expected to be affected to a lesser extent and should affect both species to the same extent (assuming that the change of male biased gene expression is randomly distributed over the divergence between *D. melanogaster* and *D. pseudoobscura*). Contrary to these
predictions the highest rate of evolution is observed for genes with a male bias in both species. Genes with a male bias in *D. melanogaster* only are evolving at slightly lower rate, while genes with a male bias in *D. pseudoobscura* only show no evidence for high rate of evolution.

Assuming that the accelerated rate of protein evolution, low codon usage and high Grantham distance of male biased genes in *D. melanogaster* are driven by positive selection, probably associated with sexual selection, the important question rises how *D. pseudoobscura* differs from *D. melanogaster*. Is there any evidence that sexual selection is reduced in *D. pseudoobscura* males?

In the following, we discuss some differences between *D. melanogaster* and *D. pseudoobscura* that could, potentially be related to sexual selection. Nevertheless, it is not clear if these differences could explain the pattern observed in this report. Both species, *D. melanogaster* (*GRIFFITHS et al. 1982; HARSHMAN and CLARK 1998; IMHOF et al. 1998*) and *D. pseudoobscura* (*ANDERSON et al. 1987; COBBS 1977*) remate frequently, but the variation within a given species across studies is too large to identify possible differences between species. Nevertheless, *D. melanogaster* and *D. pseudoobscura* differ in their tendency to remate. While *D. melanogaster* has a refractory period of two hours (*BUNDGAARD and CHRISTIANSEN 1972*), *D. pseudoobscura* starts remating after 12 hours (*BECKENBACH 1981*). Furthermore, *D. pseudoobscura* females approach males more often than *D. melanogaster* females do (*GOWATY et al. 2003*), suggesting that *D. pseudoobscura* females are choosier than *D. melanogaster* ones. While it is conceivable that a longer refractory period and choosier females could reduce some aspects of sexual selection in male *D. pseudoobscura*, it is not clear if this is sufficient to explain the
observed differences in molecular evolution. Further work is required to test if these differences between *D. melanogaster* and *D. pseudoobscura* could explain the contrasting evolution of male biased genes.

ACKNOWLEDGMENTS

We are thankful to D. Ibberson and V. Benes for the construction of the SAGE library and many helpful discussions. Many thanks to the M. Akam lab for providing the *D. pseudoobscura* strain. M. Schäfer and C. Vogl provided helpful comments. V. Nolte assisted in GLGI experiments. This work is funded through ‘Fonds zur Förderung der wissenschaftlichen Forschung’ grants to C.S.

LITERATURE CITED


Table 1. Rate of evolution, Grantham distance and codon usage for genes in different categories of sex biased expression pattern

<table>
<thead>
<tr>
<th>Expression bias</th>
<th>No. of genes</th>
<th>( d_N )</th>
<th>( d_S )</th>
<th>( d_N/d_S )</th>
<th>Grantham distance</th>
<th>( F_{op} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>D. pseudoobscura</td>
</tr>
<tr>
<td>Male biased</td>
<td>34</td>
<td>0.257</td>
<td>2.795</td>
<td>0.109</td>
<td>67.990</td>
<td>0.601</td>
</tr>
<tr>
<td>Male biased</td>
<td>77</td>
<td>0.095</td>
<td>2.061</td>
<td>0.049</td>
<td>61.880</td>
<td>0.592</td>
</tr>
<tr>
<td>Male biased</td>
<td>6</td>
<td>0.059</td>
<td>2.425</td>
<td>0.026</td>
<td>52.860</td>
<td>0.712</td>
</tr>
<tr>
<td>Unbiased</td>
<td>187</td>
<td>0.216</td>
<td>2.698</td>
<td>0.087</td>
<td>64.803</td>
<td>0.594</td>
</tr>
<tr>
<td>Unbiased</td>
<td>681</td>
<td>0.106</td>
<td>2.295</td>
<td>0.054</td>
<td>61.091</td>
<td>0.560</td>
</tr>
<tr>
<td>Unbiased</td>
<td>99</td>
<td>0.078</td>
<td>1.910</td>
<td>0.043</td>
<td>56.041</td>
<td>0.599</td>
</tr>
<tr>
<td>Female biased</td>
<td>7</td>
<td>0.041</td>
<td>1.383</td>
<td>0.027</td>
<td>46.572</td>
<td>0.608</td>
</tr>
<tr>
<td>Female biased</td>
<td>11</td>
<td>0.096</td>
<td>1.517</td>
<td>0.056</td>
<td>61.140</td>
<td>0.704</td>
</tr>
</tbody>
</table>
Table 2. Comparison of genes with male biased expression in *D. pseudoobscura* and *D. melanogaster* with *D. simulans*

A. Based on fold difference

<table>
<thead>
<tr>
<th>Gene</th>
<th>Log$_2$ expression ratio (M/F)</th>
<th>Gene</th>
<th>Log$_2$ expression ratio (M/F)$^1$</th>
<th>Gene</th>
<th>Log$_2$ expression ratio (M/F)$^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA10438-RA</td>
<td>2.32</td>
<td>CG10616</td>
<td>1.21</td>
<td></td>
<td>1.13</td>
</tr>
<tr>
<td>GA13209-RA</td>
<td>2.00</td>
<td>CG14735</td>
<td>5.34</td>
<td></td>
<td>5.07</td>
</tr>
<tr>
<td>GA15417-RA</td>
<td>1.67</td>
<td>CG2668</td>
<td>5.90</td>
<td></td>
<td>6.10</td>
</tr>
<tr>
<td>GA15920-RA</td>
<td>3.32</td>
<td>CG3092</td>
<td>5.01</td>
<td></td>
<td>4.38</td>
</tr>
<tr>
<td>GA18342-RA</td>
<td>2.81</td>
<td>CG4669</td>
<td>4.76</td>
<td></td>
<td>4.49</td>
</tr>
<tr>
<td>GA18974-RA</td>
<td>3.32</td>
<td>CG5565</td>
<td>4.46</td>
<td></td>
<td>3.49</td>
</tr>
<tr>
<td>GA20544-RA</td>
<td>2.00</td>
<td>CG7722</td>
<td>2.88</td>
<td></td>
<td>2.52</td>
</tr>
</tbody>
</table>

$^1$ Ranz *et al*, 2003
## B. Based on level of significance

<table>
<thead>
<tr>
<th>Gene</th>
<th>Log₂ expression ratio (M/F)</th>
<th>Gene</th>
<th>Log₂ expression ratio (M/F)</th>
<th>Gene</th>
<th>Log₂ expression ratio (M/F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA10438-RA</td>
<td>2.32</td>
<td>CG10616</td>
<td>1.21</td>
<td></td>
<td>1.13</td>
</tr>
<tr>
<td>GA10736-RA</td>
<td>2.32</td>
<td>CG11064</td>
<td>0.73</td>
<td></td>
<td>1.50</td>
</tr>
<tr>
<td>GA10887-RA</td>
<td>2.32</td>
<td>CG11280</td>
<td>0.31</td>
<td></td>
<td>0.38</td>
</tr>
<tr>
<td>GA11127-RA</td>
<td>0.76</td>
<td>CG11661</td>
<td>0.77</td>
<td></td>
<td>0.68</td>
</tr>
<tr>
<td>GA12802-RA</td>
<td>2.32</td>
<td>CG1417</td>
<td>0.47</td>
<td></td>
<td>0.74</td>
</tr>
<tr>
<td>GA13209-RA</td>
<td>2.00</td>
<td>CG14735</td>
<td>5.34</td>
<td></td>
<td>5.07</td>
</tr>
<tr>
<td>GA14194-RA</td>
<td>1.26</td>
<td>CG16884</td>
<td>0.71</td>
<td></td>
<td>0.49</td>
</tr>
<tr>
<td>GA14224-RA</td>
<td>2.32</td>
<td>CG16932</td>
<td>0.28</td>
<td></td>
<td>0.18</td>
</tr>
<tr>
<td>GA14923-RA</td>
<td>2.00</td>
<td>CG18408</td>
<td>0.64</td>
<td></td>
<td>0.44</td>
</tr>
<tr>
<td>GA15417-RA</td>
<td>1.67</td>
<td>CG2668</td>
<td>5.90</td>
<td></td>
<td>6.10</td>
</tr>
<tr>
<td>GA15456-RA</td>
<td>1.74</td>
<td>CG2765</td>
<td>0.68</td>
<td></td>
<td>0.70</td>
</tr>
<tr>
<td>GA15920-RA</td>
<td>3.32</td>
<td>CG3092</td>
<td>5.01</td>
<td></td>
<td>4.38</td>
</tr>
<tr>
<td>GA18342-RA</td>
<td>2.81</td>
<td>CG4669</td>
<td>4.76</td>
<td></td>
<td>4.49</td>
</tr>
<tr>
<td>GA18974-RA</td>
<td>3.32</td>
<td>CG5565</td>
<td>4.46</td>
<td></td>
<td>3.49</td>
</tr>
<tr>
<td>GA19329-RA</td>
<td>0.72</td>
<td>CG6058</td>
<td>0.81</td>
<td></td>
<td>0.71</td>
</tr>
<tr>
<td>GA20243-RA</td>
<td>2.32</td>
<td>CG7292</td>
<td>0.84</td>
<td></td>
<td>0.83</td>
</tr>
<tr>
<td>GA20345-RA</td>
<td>2.00</td>
<td>CG7430</td>
<td>0.51</td>
<td></td>
<td>0.50</td>
</tr>
<tr>
<td>GA20544-RA</td>
<td>2.00</td>
<td>CG7722</td>
<td>2.88</td>
<td></td>
<td>2.52</td>
</tr>
<tr>
<td>GA20881-RA</td>
<td>1.00</td>
<td>CG8189</td>
<td>0.76</td>
<td></td>
<td>-0.07²</td>
</tr>
<tr>
<td>Gene</td>
<td>Value 1</td>
<td>Gene</td>
<td>Value 2</td>
<td>Value 3</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>---------</td>
<td>----------</td>
<td>---------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>GA20964-RA</td>
<td>2.58</td>
<td>CG8295</td>
<td>0.76</td>
<td>0.59</td>
<td></td>
</tr>
<tr>
<td>GA22030-RA</td>
<td>1.74</td>
<td>CG9779</td>
<td>0.39</td>
<td>0.42</td>
<td></td>
</tr>
</tbody>
</table>

1 Ranz et al., 2003

2 Unbiased expression in *D. simulans*
**Figure legends**

Figure 1. Number of male and female biased tags based on different levels of significance (Fisher’s exact test)

Figure 2. Gene expression of *D. pseudoobscura* males and females measured by SAGE tag counts. The open diamonds indicate SAGE tags that differ significantly (*P*≤ 0.05, Fisher’s exact test) between both sexes. The diamonds above the diagonal are male biased and below the diagonal are female biased. Light grey crosses indicate unbiased tags. A value of –1 was assigned to tags with no counts for one of the sexes.

Figure 3. Density of male biased, female biased and unbiased genes (A) On XL, XR and the autosomes in *D. pseudoobscura* (B) On X chromosome and autosomes in *D. melanogaster*.

Figure 4. Average dN values between *D. melanogaster* and *D. pseudoobscura* at different significance level cutoffs. MM: genes with male biased expression in both the species, MU: genes with male biased expression in *D. pseudoobscura* and unbiased expression in *D. melanogaster*, UM: genes with unbiased expression in *D. pseudoobscura* and male biased expression in *D. melanogaster*.

Figure 5. Mean dN value of genes with unbiased gene expression in *D. pseudoobscura* and a male biased gene expression in *D. melanogaster* (solid line). The X-axis gives the proportion of genes with low dN value that were deleted from the full data set. Note that the genes were sorted according to dN value and iteratively the gene with the lowest dN value was removed. The dotted line provides the mean dN value for genes with a male biased gene expression in both species (no genes were deleted).
Figure 1
Figure 2
Figure 3.

A.

![Gene density (Genes/MB) for different chromosomes with bias categories]

B.

![Gene density (Genes/MB) for different chromosomes with bias categories]
Figure 4.
Figure 5.