RESEARCH

The X chromosome of the German cockroach, Blattella germanica, is homologous to a fly X chromosome despite 400 million years divergence

Richard P Meisel[1*], Pablo J Delclos[1] and Judith R Wexler[2,3]

*Correspondence: rpmeisel@uh.edu
[1]Department of Biology and Biochemistry, University of Houston, 3455 Cullen Blvd, Houston, TX, USA 77204
Full list of author information is available at the end of the article

Abstract

Background: Sex chromosome evolution is a dynamic process that can proceed at varying rates across lineages. For example, different chromosomes can be sex-linked between closely related species, whereas other sex chromosomes have been conserved for >100 million years. Cases of long-term sex chromosome conservation could be informative of factors that constrain sex chromosome evolution. We hypothesize that the German cockroach, *Blattella germanica*, has the same X chromosome as most flies, representing an extreme case of sex chromosome conservation.

Results: To test our hypothesis, we analyzed whole genome sequence data and found that a significant excess of the same genes are on the cockroach and fly X chromosomes. We also show that the candidate X-linked cockroach genes appear to be dosage compensated in hemizygous males. In addition, three regulators of transcription and chromatin on the fly X chromosome are conserved in the cockroach genome.

Conclusions: Our results support our hypothesis that the German cockroach shares the same X chromosome as most flies. This suggests that the common ancestor of most insects had an X chromosome that resembled the extant cockroach and fly X. Cockroaches and flies diverged ∼400 million years ago, making this the longest documented conservation of a sex chromosome. Cockroaches and flies have different mechanisms of sex determination, suggesting that the X chromosome was conserved despite evolution of the sex determination pathway.

Keywords: sex chromosome; evolution; insect; arthropod genomics; dosage compensation

Background

In species with separate sexes, genetic or environmental cues initiate sexually dimorphic developmental pathways [1, 2]. If the cue is genetic, a sex determining factor may reside on a sex chromosome [3]. For example, in most therian mammals, *SRY* on the Y chromosome initiates the development of the male germline, testes, and secondary sexual traits [4]. In contrast, the dosage of the X chromosome determines the initiation of male or female development in *Drosophila melanogaster* [5–7]. In both taxa, females have the XX genotype, and males are XY. Despite the superficial similarities, the sex chromosomes and genes that initiate the sex determination pathways are not homologous between mammals and *Drosophila* [3]. In addition, some, but not all, animal taxa have evolved mechanisms to compensate for the haploid dose of the X chromosome in males, or Z chromosome in ZW females [8–11].

Sex determining pathways and sex chromosomes can evolve rapidly, often differing between closely related species [2, 3]. Evolutionary transitions in sex determination pathways are often accompanied by corresponding changes in the identity of the sex chromosomes [1, 2, 12]. Transitions in sex determining pathways and turnover of sex chromosomes are well studied across insects, where there is a diversity of sex determination mechanisms [13–16] (**Figure 1**). For example, the genetic factors that initiate sex determination in *Drosophila* do not determine sex in other flies [17–24]. In addition, the sex chromosomes of *Drosophila* are not homologous to the sex chromosomes of other flies [25–27]. The evolution of a new sex determination mechanism in the lineage leading to *Drosophila* resulted in the transition of the ancestral X chromosome into an autosome, the creation of a new X chromosome from an ancestral autosome, and the evolution of a new mechanism of X chromosome dosage compensation [27, 28].

It is most parsimonious to conclude that the ancestral sex determination system of brachyceran dipterans (which includes flies but excludes mosquitoes, craneflies, midges, gnats, etc.) consists of a Y-linked male-determining factor that regulates the splicing of *transformer* [15, 20, 24, 29–32]. The ancestral male-determining gene of brachyceran flies is yet to be identified, if it is even still present in any extant species. The ancestral brachyceran X is a heterochromatin-enriched chromosome with <100 genes that is homologous to *D. melanogaster* chromosome 4, i.e., the "dot" chromosome or Muller element F [27]. The small and heterochromatic fly X

is a notable peculiarity because most X chromosomes are gene rich and euchromatic, despite having some differences in gene content from the autosomes [33–35]. The peculiar brachyceran fly X chromosome has been conserved for >150 million years (My) in some lineages [27, 36]. The remainder of the fly genome is organized into five euchromatic chromosomes (or chromosome arms), named Muller elements A–E [37, 38].

There is some evidence that the ancestral brachyceran fly X chromosome is dosage compensated in hemizygous males. In *Drosophila*, where element F is autosomal, *Painting of fourth* (*Pof*) encodes an RNA-binding protein that localizes predominately to element F [39]. *Lucilia cuprina* (Australian sheep blowfly) has the ancestral brachyceran karyotype, with an X-linked element F [40, 41]. Expression of X-linked genes is up-regulated in *L. cuprina* males by the homolog of *Pof* [40, 42]. This dosage compensation is essential for male viability—a loss of function mutation in the *L. cuprina* the homolog of *Pof* is male lethal, but viable in females [42].

The German cockroach, *Blattella germanica*, diverged from flies ∼400 My ago (Mya) [43]. Cockroach females are XX and males are XO, i.e., one X and no Y chromosome [13, 44]. This suggests that a dosage-sensitive X-linked factor determines sex in German cockroach, analogous to, but independently evolved from, *Drosophila*. Curiously, the cockroach X chromosome is heterochromatic along most of its length [45], reminiscent of the ancestral brachyceran X chromosome. We tested the hypothesis that the cockroach and brachyceran X chromosomes are homologous despite ∼400 My divergence.

Results

Decreased sequencing coverage of element F genes in male cockroaches

We used a differential sequencing coverage approach to identify X chromosome genes in the German cockroach genome assembly. X-linked genes are expected to have half as many male-derived reads mapped to them as female-derived reads because the X chromosome is present in one copy in males and two copies in females [27]. We used available whole genome sequencing data [46] to calculate the relative coverage of male and female reads ($\log_2 \frac{M}{F}$) for each annotated cockroach gene (Additional file 1). The mode of the $\log_2 \frac{M}{F}$ distribution is at 0, consistent with a karyotype where most of the genome is autosomal (**Figure 2A**). However, there is a heavy

shoulder of genes with $\log_2 \frac{M}{F} < 0$, suggesting that X-linked genes are also in the assembly (**Figure 2A**). In total, 3,499 of the 28,141 annotated genes have female-biased coverage ($\log_2 \frac{M}{F} \leq -1$), whereas only 1,363 genes have male-biased coverage ($\log_2 \frac{M}{F} \geq 1$), consistent with a heavy shoulder of X-linked genes. Assuming the 1,363 male-biased genes represent the false-positive rate, we expect 2,136/3,499 female-biased genes to be X-linked. This is consistent with the upper-bound of the number of X-linked genes in the cockroach genome—the cockroach X is the smallest of 12 chromosomes [45], which means that fewer than 2,345 genes (28141/12) should be X-linked.

To test the hypothesis that the German cockroach X chromosome is homologous to the ancestral brachyceran fly X, we evaluated if cockroach genes with *D. melanogaster* homologs on Muller element F (the ancestral brachyceran X chromosome) have lower $\log_2 \frac{M}{F}$ than genes with homologs on the other five elements. Cockroach genes with *D. melanogaster* homologs on Muller elements A–E have distributions of $\log_2 \frac{M}{F}$ centered around 0, consistent with being autosomal (**Figure 2B**). In contrast, cockroach element F genes have a median $\log_2 \frac{M}{F} < 0$, and the average $\log_2 \frac{M}{F}$ for element F genes is significantly less than the other genes (P=10^{-10} using a Mann-Whitney U test comparing element F genes with elements A–E). If the entirety of element F were X-linked in cockroach, we would expect the median $\log_2 \frac{M}{F} = -1$ for genes with element F homologs. However, cockroach element F genes have a median $\log_2 \frac{M}{F} > -1$. Therefore, we hypothesize that a disproportionate amount of, but not all, element F genes are X-linked in German cockroach.

We next calculated the frequency of element F genes in the cockroach genome with $\log_2 \frac{M}{F} = -1$. First, we used mixtools [47] to fit a mixture of two normal distributions to the $\log_2 \frac{M}{F}$ values of element F genes. To fit a mixed distribution, we set the starting values for the means of the two distributions at -1 and 0, the expectations for X-linked and autosomal genes, respectively. The best fitting mixture consists of one distribution centered at -0.093 (close to the expectation of 0 for autosomal genes) and a second distribution centered at -0.997 (close to the expectation of -1 for X-linked genes; **Figure 2B**). The distribution centered near $\log_2 \frac{M}{F} = -1$ contains 43.6% of element F genes, suggesting that >22/51 cockroach element F genes are X-linked. A similar analysis done on cockroach genes with *D. melanogaster* homologs on elements A–E results in a mixture of two distributions

centered at $\log_2 \frac{M}{F} = 0.016$ (82.7% of genes) and $\log_2 \frac{M}{F} = -0.359$ (17.3% of genes). This suggests that the vast majority of cockroach genes with homologs on elements A–E are not X-linked.

The two distributions of cockroach element F genes are largely non-overlapping, and they seem to describe two classes of element F genes: autosomal genes with $\log_2 \frac{M}{F} > -0.5$, and X-linked genes with $\log_2 \frac{M}{F} < -0.5$ (**Figure 2B**). If there is an excess of element F genes on the cockroach X, we expect a higher frequency of element F genes to have $\log_2 \frac{M}{F} < -0.5$ than genes on the other five elements. We therefore counted the number of genes with $\log_2 \frac{M}{F} < -0.5$ on each of the six Muller elements (**Table 1**). To determine a null distribution of those genes on each chromosome, we randomly assigned the total number of genes with $\log_2 \frac{M}{F} < -0.5$ to the six elements based on the size of each Muller element (measured as the total number of cockroach genes on the element) in 1,000 bootstrap replicates of the data. A significant excess of cockroach element F genes have $\log_2 \frac{M}{F} < -0.5$ relative to our null expectation (**Figure 2C**). This provides further evidence that element F is X-linked in German cockroach.

We selected two element F genes with $\log_2 \frac{M}{F} < -1$ that we hypothesize are X-linked to validate using quantitative PCR (qPCR). Both genes had a significantly higher concentration in females relative to males in our qPCR assay, with an estimated female concentration that is twice the male concentration (Additional file 2). This is the expected result if both genes are X-linked. Therefore, male:female sequencing coverage and qPCR provide consistent evidence that element F is X-linked in German cockroach.

Reduced heterozygosity of element F genes in male cockroaches

German cockroach males have one copy of the X chromosome and females have two copies of the X. We therefore expect that females could be heterozygous for variants in X-linked genes, whereas males must be hemizygous (only one allele per gene). To further test if element F genes are disproportionately X-linked in cockroach, we used the available cockroach genome sequence data to calculate the number of heterozygous sequence variants in cockroach genes (Additional file 1). We then compared heterozygosity of cockroach genes across Muller elements (**Figure 3**). There is not a significant difference in female heterozygosity between genes assigned to element F

and genes on the other five elements (P=0.32 in a Mann-Whitney U test). In contrast, element F genes have significantly fewer heterozygous variants in males than genes on elements A–E (P=0.017 in a Mann-Whitney U test). This is consistent with an excess of element F genes on the German cockroach X chromosome.

We expect candidate X-linked genes with reduced $\log_2 \frac{M}{F}$ sequencing coverage to also have reduced heterozygosity in males relative to females. To test this hypothesis, we compared heterozygous variants in males and females by calculating the fraction of heterozygous variants in each gene that are in males relative to females (Additional file 1). This value ranges from zero (if a gene only has heterozygous variants in females) to one (if a gene only has heterozygous variants in males). Equal heterozygosity in both sexes has a value of 0.5. Of the 40 element F genes with sequencing coverage and heterozygosity data, 10 (25%) have both $\log_2 \frac{M}{F} < -0.5$ and fraction of male heterozygous variants < 0.5 (**Figure 3C**). This is significantly greater than the 2.5% of element A–E genes with both $\log_2 \frac{M}{F} < -0.5$ and fraction of male heterozygous variants < 0.5 (z=9.68, P=10^{-21}). This result provides further evidence that there is an excess of element F genes on the German cockroach X chromosome.

The cockroach X chromosome is dosage compensated in males

We next tested if the haploid dosage of element F genes affects their expression in cockroach males. The ideal data to test for the effects of a haploid X are expression measurements from males and females from the same tissue [10, 11]. Unfortunately, there are no available sex-matched RNA-seq gene expression datasets from German cockroach. We therefore used an alternative approach in which we compared expression in adult male heads with a mixed sex adult head sample (Additional file 1). We also compared expression in adult male heads with whole adult females (Additional file 1). If the haploid X chromosome is dosage compensated in males, we expect the distributions of $\log_2$ fold-change ($\log_2$ FC) expression between the two tissue samples to be equivalent for genes on element F and genes on elements A–E. Indeed, there is not a significant difference in the median $\log_2$ FC between element F genes and element A–E genes (P=0.15 for male head *vs* mixed sex head, P=0.30 for male head *vs* whole adult female, with both P-values from Mann-Whitney U tests; **Figure 4A-B**).

Only a subset of a element F genes is expected to be X-linked in cockroach based on $\log_2 \frac{M}{F}$ sequencing coverage (**Figure 2B**). If the X chromosome is dosage compensated in males, we expect the average $\log_2$ FC expression between tissue samples to be similar for element F genes with evidence of X-linkage ($\log_2 \frac{M}{F} < -0.5$) and element F genes that appear to be autosomal ($\log_2 \frac{M}{F} \geq -0.5$). Indeed, there is not a significant difference in $\log_2$ FC between the two subsets of element F genes (P=0.84 for male head *vs* mixed sex head, P=0.30 for male head *vs* whole adult female, with both P-values from Mann-Whitney U tests; **Figure 4C-D**). Therefore, the limited RNA-seq data that are available suggest that the German cockroach X chromosome is dosage compensated in males.

Conservation of element F transcriptional regulators in cockroach

In some fly species where element F is the X chromosome, X-linked genes are present in a single (haploid) copy in males [27]. Males of the blow fly *L. cuprina* are haploid for such an X chromosome, and their X-linked genes are up-regulated in males by an RNA-binding protein encoded by a homolog of *Drosophila Pof* [40, 42]. POF localizes nearly exclusively to element F gene bodies in *D. melanogaster* [39, 48–50]. There is a *Pof* homolog in the cockroach genome (BGER016147), which we aligned to the *D. melanogaster* protein. The most conserved region of *Pof* overlaps with a predicted RNA binding domain within the cockroach protein sequence (**Figure 5A–B**). Therefore, a key component of the molecular machinery which regulates dosage compensation on the brachyceran X is present in the German cockroach genome.

The proteins encoded by *eggless* (*egg*) and *windei* (*wde*) interact with POF to create an environment around genes on element F that resembles pericentromeric heterochromatin in *Drosophila*. Egg is a SETDB1 homolog that is responsible for di- and/or tri-methylation of lysine 9 in histone H3 in the gene-dense region of *D. melanogaster* element F [51–55]. There are two predicted homologs of *egg* in the cockroach genome (BGER011023 and BGER011024). BGER011023 has a predicted SET lysine methyltransferase domain and a methyl-CpG binding domain commonly found in histone methyltransferases. BGER011024, on the other hand, has a tudor domain, which is found proximal to the SET domain in *D. melanogaster* Egg [56]. These predicted functional domains overlap with the portions of the cockroach

proteins that are most conserved relative to *D. melanogaster* Egg (**Figure 5C–D**).

BGER011023 and BGER011024 are contiguous on a single *B. germanica* scaffold (Scaffold202; KN196692), suggesting that together they may constitute a single gene that encodes all Egg functional regions.

Wde is an essential co-factor of Egg [57]. There is one predicted homolog of *wde* in the cockroach genome annotation (BGER025676), but an independently sequenced cockroach *wde* gene (CCX34999) is longer than the *wde* homolog predicted by the automated annotation [58]. We therefore compared CCX34999 with *D. melanogaster* Wde. CCX34999 contains a predicted fibronectin type-III domain at the C-terminal end, similar to *D. melanogaster* Wde [56]. The C-terminal end of CCX34999 is also the most conserved part of the protein relative to *D. melanogaster* Wde (**Figure 5E–F**). There is a coiled-coil region of *D. melanogaster* Wde that is required to interact with Egg. That coiled-coil region of Wde, and the corresponding region of Egg that interacts with Wde, are among the most conserved regions of the *D. melanogaster* proteins when compared to the cockroach homologs (**Figure 5C,E**). Therefore, homologs of *Pof* and its two key interactors are present in the German cockroach genome, providing evidence that a similar mechanism may dosage compensate the cockroach and ancestral fly X chromosomes in hemizygous males.

Discussion

We provide two lines of evidence that the X chromosome of the German cockroach, *B. germanica*, is homologous to a fly X chromosome. First, there is reduced sequencing coverage of nearly half of the Muller element F genes in male cockroaches, consistent with a haploid dose of the X chromosome in males (**Figure 2**). Second, there is decreased heterozygosity of element F genes in male cockroaches, including those with reduced male sequencing coverage (**Figure 3**). We therefore hypothesize that element F is an ancient X chromosome that was present in the most recent common ancestor (MRCA) of flies and cockroaches, and it has been conserved as an X chromosome in the German cockroach and many fly species. An alternative explanation for the excess of element F genes on the cockroach X chromosome is that those genes independently became X-linked in both cockroaches and flies.

There are at least four lines of evidence that favor the hypothesis that element F is an ancient X chromosome retained since the MRCA of cockroaches and flies. First, an independent analysis concluded that the MRCA of flies and cockroaches had XX females and either XY or XO males [16]. Second, the *B. germanica* X chromosome stains heavily for heterochromatin [45], similar to the brachyceran fly X [59]. X chromosomes tend to be euchromatic in males [33–35], making the similarity between the *B. germanica* and brachyceran X heterochromatin notable. However, most of what we know about insect sex chromosome heterochromatin comes from cytological examination of meiotic cells from testes [60], where sex-chromosome-specific heterochromatization could differ from the normal behavior in somatic cells [61]. Additional work is necessary to investigate the chromatin state of insect sex chromosomes outside of the male germline. Third, the observed number of element F genes with evidence for X-linkage in cockroach greatly exceeds the expectation if the X chromosomes of flies and cockroaches were independently derived (**Figure 2C**). Fourth, the fraction of element F genes that appear to be X-linked in cockroach (22/51) is consistent with two separate estimates of the expected conservation of a shared X chromosome that was present in the MRCA of flies and cockroaches. We explain the two separate estimates of expected X chromosome conservation below.

The first estimate of expected conservation of an X-linked element F draws upon rates of gene relocation between Muller elements in *Drosophila*. If element F were the ancestral X chromosome, we would expect some relocation of genes onto and off of element F as the lineages leading to cockroaches and flies diverged from their MRCA [62]. Based on the frequency of gene relocation between Muller elements in *Drosophila* [63] and the sizes of the elements *D. melanogaster*, we expect 6.4 genes to have relocated off element F in the cockroach lineage and 1.3 genes to have relocated onto element F in the fly lineage (see Methods for calculations). There are 29 *D. melanogaster* element F genes that do not have evidence for X-linkage in cockroach (**Figure 2B**). Gene movement alone can thus explain 7–8 of these apparently autosomal element F genes.

The second estimate of expected conservation of an X-linked element F extrapolates from the conservation of element F between *D. melanogaster* and the blow fly *L. cuprina*. In the *L. cuprina* genome, only 67.1% (49/73) of genes with *D. melanogaster* element F homologs are X-linked [42]. Assuming a linear relationship between

divergence time [36, 64] and conservation of element F gene content, we would expect only 11.1% of cockroach genes with element F homologs to be X-linked:

$$67.1\% \times \frac{\text{64 My since divergence between } \textit{Drosophila} \text{ and blow flies}}{\text{386.9 My since divergence between flies and cockroaches}}$$

Our estimate of >44% of X-linked element F genes greatly exceeds this expected amount of conservation. The actual number of element F genes that are X-linked in *B. germanica* (22/51) is in between the estimates predicted based on rates of gene relocation and a linear loss of gene content. Therefore, conservation of an X-linked element F from the MRCA of flies and cockroaches is consistent with the expected amount of gene turnover in the time since the MRCA.

Curiously, there is a long tail of genes with much higher sequencing coverage in females relative to males ($\log_2 \frac{M}{F} \ll -1$), regardless of the Muller element of their *D. melanogaster* homologs (**Figure 2A**). Sexually dimorphic amplification (endoreplication) of a subset of the genome has been documented in insects, such as in the chorion genes that are highly expressed in the *Drosophila* ovary [65, 66]. It is therefore possible that a subset of the cockroach genome is disproportionately amplified in females (possibly to meet the gene expression demands of oogenesis), causing the long tail of negative $\log_2 \frac{M}{F}$ values that we observe. Additional work is necessary to test this hypothesis.

Our analysis of RNA-seq data suggests that the cockroach X chromosome is dosage compensated in males—we find no evidence for reduced expression of element F genes in male cockroaches, regardless of whether the genes appear to be haploid in males (**Figure 4**). In addition, homologs of genes involved in regulating expression of element F genes in flies are present in the cockroach genome, with their functional domains conserved (**Figure 5**). This is consistent with cockroaches and flies sharing a mechanism of X chromosome dosage compensation that has been conserved since their MRCA.

Finally, our results provide evidence that X chromosomes can be conserved even though there are changes in the master regulators of sex determination. Sex in *B. germanica* is likely determined by X chromosome dosage, analogous to *Drosophila*, but different from the ancestral brachyceran sex determination system (**Figure 1**). It is unlikely that the same X-linked dosage sensitive factors determine

sex in cockroach and *Drosophila* because the X chromosome is not homologous between the two taxa (element A is the X chromosome in *Drosophila*). In addition, the master regulators of *Drosophila* sex determination almost certainly differ from the sex determiners in the MRCA of Brachycera, which likely used a Y-linked male-determiner (**Figure 1**). Moreover, the sexually dimorphic splicing of the sex determination pathway gene *transformer* differs between German cockroach and flies [67]. Therefore, we hypothesize that *B. germanica* has a homologous X chromosome with the MRCA of Brachycera, but the sex determination system is not conserved between cockroaches and flies. Our results suggest that conservation of sex chromosomes does not necessarily imply conservation of sex determination. Future work addressing this problem could inform our understanding of how evolutionary transitions in sex determination pathways can be decoupled from sex chromosome turnover [68].

Conclusions

We present evidence that the X chromosome of the German cockroach is homologous to an X chromosome shared by many fly species. We hypothesize that this X chromosome was inherited from the MRCA of cockroaches and flies >400 Mya. To the best of our knowledge, this is the longest documented conservation of an X chromosome. This ancient X chromosome may be dosage compensated in male cockroaches and flies by a conserved mechanism. The extremely long-term conservation of the X chromosome is especially remarkable because cockroaches and flies have diverged in their sex determination pathways, suggesting that sex chromosome conservation can be decoupled from the evolution of sex determination.

Methods

Assigning German cockroach genes to Muller elements

We retrieved the *D. melanogaster* homologs of *B. germanica* genes from the Baylor College of Medicine i5k Maker annotation, version 0.5.3 [46]. *Drosophila* and other fly genomes are organized into six chromosomes (or chromosome arms) known as Muller elements [25, 37, 69, 70]. Muller element F is the ancestral X chromosome of Brachycera, and elements A–E are autosomal in flies with this ancestral karyotype [27]. We assigned German cockroach genes with a single *D. melanogaster* homolog to fly Muller elements based on the element of their homolog.

Differential sequencing coverage in males and females

We tested for genes that were sequenced at different depths in males and females as a way to identify X chromosome genes [27]. First, we aligned paired-end reads from three male cockroach whole genome sequencing libraries (SRX693111, SRX693112, and SRX693113) and one female library (SRX693110) to the reference *B. germanica* genome assembly [JPZV00000000.1; 46] using BWA-MEM with default parameters [71]. We then assigned mapped read pairs to genes (from the v. 0.5.3 i5k annotation) if the first (forward) read aligned to any portion of a gene sequence. We only considered the forward read because insert sizes differ across the available sequencing libraries, which could introduce biases in gene coverage if we allowed or required both forward and reverse reads to overlap genes. Considering only the forward read should decrease the effect of these biases because read lengths are the same (101 bp) across all libraries. We summed across libraries to determine the total number of reads mapped to each gene for each sex. We next divided the number of male-derived (female-derived) reads aligned to each gene by the total number of male-derived (female-derived) reads aligned to all genes to determine a normalized mapping coverage of male-derived (female-derived) reads for each gene (Additional file 1). We used these normalized counts to calculate the $\log_2$ male:female read mapping coverage ($\log_2 \frac{M}{F}$) for each annotated cockroach gene, and we normalized the data so that the median across all genes is zero.

Quantitive PCR validation of candidate X-linked genes

We used qPCR to validate two candidate X-linked genes in German cockroach. Briefly, genomic DNA was extracted from the head and legs of five individual male and five individual female cockroaches from the Orlando Normal strain. We designed PCR primers to amplify the genomic region corresponding to each gene, as well as two control genes that we hypothesize are autosomal (sequences provided in Additional file 2). We used a StepOne Plus Real-Time PCR System (Applied Biosystems) to quantify the concentration of DNA from each of the candidate genes and the control genes in each individual cockroach. We then used a mixed effects model to assess the effect of sex on the concentration of the candidate X-linked genes. Details are provided in Additional file 2.

Differential heterozygosity in males and females

We tested for genes with reduced heterozygosity in males relative to females as an additional way to identify X chromosome genes. We used the Genome Analysis Toolkit (GATK) version 3.4-0 to identify heterozygous single nucleotide polymorphisms (SNPs) and small variants in the alignments of male and female sequencing reads described above, following the GATK best practices [72–74]. Because there is no reference variant set for cockroach, we used the following steps to extract high confidence variants [68]. First, we used Picard Tools version 1.133 to identify and remove duplicate reads, and we realigned indels with GATK. Then, we performed naïve variant calling using the GATK HaplotypeCaller with a phred-scaled confidence threshold of 20, and we selected the highest confidence SNPs from that first-pass (QD < 2.0, MQ < 40, FS > 60, SOR > 4, MQRankSum < -12.5, ReadPosRankSum < -8). We used those high quality variants to perform base recalibration, we re-input those recalibrated bases into another round of variant calling, and we extracted the highest quality variants. We repeated the process so that we had performed three rounds of recalibration, which was sufficient for convergence of variant calls. We applied GenotypeGVCFs to the variant calls from all of the Illumina libraries for joint genotyping of both males and females, and we selected only the high quality variants using VariantFiltration (FS > 30 and QD < 2). We then extracted variants that mapped to *B. germanica* genes (from the v. 0.5.3 i5k annotation).

Differential gene expression using RNA-seq data

We compared the expression of genes in adult male heads (SRX3189901 and SRX3189902) with expression in a mixed sex adult head sample (SRX682022) using available RNA-seq data [75, 76]. We also compared male head expression with expression in whole adult females [SRX2746607 and SRX2746608; 46]. We aligned the RNA-seq reads from each library to *B. germanica* transcripts (from the v. 0.5.3 i5k annotation) using kallisto [77]. The male head libraries were sequenced using single end reads, and we specified an average fragment length (-l) of 200 bp and a standard deviation (-s) of 20 bp. There is only a single transcript for each gene in the *B. germanica* annotation, and so we treated transcript-level read counts as equivalent to gene-wise counts. We also only included genes with at least 10 mapped

reads across all samples. We then used DESeq2 to estimate the $\log_2$ fold-change of expression for each gene between male heads and mixed sex heads, as well as between male heads and whole adult females [78]. All reads from a given accession were treated as belonging to a replicate (i.e., we summed read counts of different sequencing runs within each accession).

Conservation of element F regulators

We aligned the sequences of three *D. melanogaster* proteins that regulate element F gene expression (POF, Eggless, and Windei) with their *B. germanica* homologs using MUSCLE [79]. We then calculated amino acid (aa) sequence conservation in 50 aa sliding windows (with 1 aa increments) in the reference protein sequence. Gaps in the cockroach sequences were counted as mismatches, and gaps in the *D. melanogaster* sequences were ignored. Functional domains were predicted by the NCBI Conserved Domain Database [56] or retrieved from UniProt [80].

Expected conservation of element F

We performed calculations to estimate the number of genes relocated onto and off of element F in the lineages leading to cockroach and flies. First, the expected number of genes relocated from element F to the other elements in the lineage leading to the German cockroach was estimated from the observed number of X-to-autosome relocations in the lineage leading to *D. melanogaster* since the divergence with *Drosophila pseudoobscura* [24; 63], the fraction of genes on element F ($86/14237 = 0.006$) and element A (the *Drosophila* X chromosome, $2274/14237 = 0.16$) in *D. melanogaster* [81], the divergence time between *D. melanogaster* and *D. pseudoobscura* [54.9 My; 82], and the divergence time between flies and cockroaches [386.9 My; 43]. We assumed that the rate of relocation from the ancestral X chromosome to the autosomes in the lineage leading to cockroach is the same as the rate from the *Drosophila* X to autosomes. We then calculated the expected number of genes relocated from element F to other elements in the lineage leading to the German cockroach as:

$$(24 \text{ genes})\left(\frac{0.006}{0.16}\right)\left(\frac{386.9 \text{ My}}{54.9 \text{ My}}\right) = 6.4 \text{ genes}$$

Second, to estimate the number of genes relocated onto element F from other elements in the lineage leading to *D. melanogaster*, we included an estimate of the number of autosome-to-X relocations in the lineage leading to *D. melanogaster* since the divergence with *D. pseudoobscura* [5; 63]. We treated element F as an X chromosome in the entire lineage leading from the MRCA of flies and cockroach, which it was for most of that time (332/387 My). We then calculated the expected number of genes relocated onto element F in the lineage leading to *D. melanogaster* as:

$$(5 \text{ genes})\left(\frac{0.006}{0.16}\right)\left(\frac{386.9 \text{ My}}{54.9 \text{ My}}\right) = 1.3 \text{ genes}$$

Competing interests

The authors declare that they have no competing interests.

Declarations

The data analyzed in the current study are available from the NCBI accessions provided in the Methods.

Author's contributions

RPM and JRW conceived the project. RPM analyzed the genomic data. JRW collected biological materials, extracted DNA, and designed primers for qPCR validation. PJD collected and analyzed the qPCR data. RPM wrote the paper, with assistance from PJD and JRW. All authors read and approved the final manuscript

Acknowledgements

This work was supported by National Institutes of Health grant R35GM122592 to Artyom Kopp and National Institute of Food and Agriculture funds from the University of California. Computational analyses were performed on the Maxwell cluster provided by the Core facility for Advanced Computing and Data Science at the University of Houston.

Author details

[1]Department of Biology and Biochemistry, University of Houston, 3455 Cullen Blvd, Houston, TX, USA 77204.
[2]Department of Ecology and Evolution, University of California, Davis, One Shields Ave, Davis, CA, USA 95616.
[3]Department of Entomology, University of Maryland, 4112 Plant Sciences Building, College Park, MD, USA 20742.

References

1. Bull, J.J.: Evolution of Sex Determining Mechanisms. Benjamin/Cummings, Menlo Park, CA (1983)
2. Beukeboom, L., Perrin, N.: The Evolution of Sex Determination. Oxford University Press, New York, NY (2014)
3. Bachtrog, D., Mank, J.E., Peichel, C.L., Kirkpatrick, M., Otto, S.P., Ashman, T.-L., Hahn, M.W., Kitano, J., Mayrose, I., Ming, R., Perrin, N., Ross, L., Valenzuela, N., Vamosi, J.C.: Sex determination: why so many ways of doing it? PLoS Biol **12**(7), 1001899 (2014). doi:10.1371/journal.pbio.1001899
4. Goodfellow, P.N., Lovell-Badge, R.: *SRY* and sex determination in mammals. Ann. Rev. Genet. **27**(1), 71–92 (1993). doi:10.1146/annurev.ge.27.120193.000443. PMID: 8122913. http://dx.doi.org/10.1146/annurev.ge.27.120193.000443
5. Baker, B.S., Belote, J.M.: Sex determination and dosage compensation in *Drosophila melanogaster*. Ann. Rev. Genet. **17**, 345–393 (1983)
6. Cline, T.W.: The *Drosophila* sex determination signal: how do flies count to two? Trends Genet. **9**(11), 385–390 (1993). doi:10.1016/0168-9525(93)90138-8

7. Erickson, J.W., Quintero, J.J.: Indirect effects of ploidy suggest X chromosome dose, not the X:A ratio, signals sex in Drosophila. PLoS Biology **5**, 332 (2007)
8. Ohno, S.: Sex Chromosomes and Sex-linked Genes. Springer, ??? (1967)
9. Mank, J.E.: Sex chromosome dosage compensation: definitely not for everyone. Trends Genet. **29**(12), 677–683 (2013). doi:10.1016/j.tig.2013.07.005
10. Chandler, C.H.: When and why does sex chromosome dosage compensation evolve? Ann. N. Y. Acad. Sci. **1389**(1), 37–51 (2017). doi:10.1111/nyas.13307. https://nyaspubs.onlinelibrary.wiley.com/doi/pdf/10.1111/nyas.13307
11. Gu, L., Walters, J.R.: Evolution of sex chromosome dosage compensation in animals: A beautiful theory, undermined by facts and bedeviled by details. Genome Biol. Evol. **9**(9), 2461–2476 (2017). doi:10.1093/gbe/evx154
12. van Doorn, G.S.: Patterns and mechanisms of evolutionary transitions between genetic sex-determining systems. Cold Spring Harbor Perspectives in Biology **6**(8), 017681 (2014). doi:10.1101/cshperspect.a017681. http://cshperspectives.cshlp.org/content/6/8/a017681.full.pdf+html
13. Kaiser, V.B., Bachtrog, D.: Evolution of sex chromosomes in insects. Ann. Rev. Genet. **44**, 91–112 (2010). doi:10.1146/annurev-genet-102209-163600
14. Gempe, T., Beye, M.: Function and evolution of sex determination mechanisms, genes and pathways in insects. Bioessays **33**(1), 52–60 (2011). doi:10.1002/bies.201000043
15. Bopp, D., Saccone, G., Beye, M.: Sex determination in insects: variations on a common theme. Sex. Dev. **8**(1-3), 20–28 (2014)
16. Blackmon, H., Ross, L., Bachtrog, D.: Sex determination, sex chromosomes, and karyotype evolution in insects. J. Hered. **108**(1), 78–93 (2017). doi:10.1093/jhered/esw047
17. Sievert, V., Kuhn, S., Traut, W.: Expression of the sex determining cascade genes *Sex-lethal* and doublesex in the phorid fly *Megaselia scalaris*. Genome **40**(2), 211–214 (1997)
18. Meise, M., Hilfiker-Kleiner, D., Dübendorfer, A., Brunner, C., Nothiger, R., Bopp, D.: *Sex-lethal*, the master sex-determining gene in *Drosophila*, is not sex-specifically regulated in *Musca domestica*. Development **125**(8), 1487–1494 (1998). http://dev.biologists.org/content/125/8/1487.full.pdf+html
19. Saccone, G., Peluso, I., Artiaco, D., Giordano, E., Bopp, D., Polito, L.C.: The *Ceratitis capitata* homologue of the *Drosophila* sex-determining gene *sex-lethal* is structurally conserved, but not sex-specifically regulated. Development **125**(8), 1495–1500 (1998)
20. Pane, A., Salvemini, M., Delli Bovi, P., Polito, C., Saccone, G.: The *transformer* gene in *Ceratitis capitata* provides a genetic basis for selecting and remembering the sexual fate. Development **129**(15), 3715–3725 (2002)
21. Hall, A.B., Basu, S., Jiang, X., Qi, Y., Timoshevskiy, V.A., Biedler, J.K., Sharakhova, M.V., Elahi, R., Anderson, M.A.E., Chen, X.-G., Sharakhov, I.V., Adelman, Z.N., Tu, Z.: A male-determining factor in the mosquito *Aedes aegypti*. Science **348**(6240), 1268–1270 (2015). doi:10.1126/science.aaa2850. http://www.sciencemag.org/content/348/6240/1268.full.pdf
22. Criscione, F., Qi, Y., Tu, Z.: GUY1 confers complete female lethality and is a strong candidate for a male-determining factor in *Anopheles stephensi*. eLife **5**, 19281 (2016). doi:10.7554/eLife.19281
23. Krzywinska, E., Dennison, N.J., Lycett, G.J., Krzywinski, J.: A maleness gene in the malaria mosquito *Anopheles gambiae*. Science **353**(6294), 67–69 (2016). doi:10.1126/science.aaf5605. http://science.sciencemag.org/content/353/6294/67.full.pdf
24. Sharma, A., Heinze, S.D., Wu, Y., Kohlbrenner, T., Morilla, I., Brunner, C., Wimmer, E.A., van de Zande, L., Robinson, M.D., Beukeboom, L.W., Bopp, D.: Male sex in houseflies is determined by *Mdmd*, a paralog of the generic splice factor gene *CWC22*. Science **356**(6338), 642–645 (2017). doi:10.1126/science.aam5498. http://science.sciencemag.org/content/356/6338/642.full.pdf
25. Foster, G.G., Whitten, M.J., Konovalov, C., Arnold, J.T.A., Maffi, G.: Autosomal genetic maps of the Australian sheep blowfly, *Lucilia cuprina dorsalis* R.-D. (Diptera: Calliphoridae), and possible correlations with the linkage maps of *Musca domestica* L. and *Drosophila melanogaster* (Mg.). Genet. Res. **37**, 55–69 (1981). doi:10.1017/S0016672300020012
26. Baker, R.H., Wilkinson, G.S.: Comparative genomic hybridization (CGH) reveals a neo-X chromosome and

biased gene movement in stalk-eyed flies (genus *Teleopsis*). PLoS Genet. **6**(9), 1001121 (2010)

27. Vicoso, B., Bachtrog, D.: Reversal of an ancient sex chromosome to an autosome in *Drosophila*. Nature **499**(7458), 332–335 (2013)
28. Kuroda, M.I., Hilfiker, A., Lucchesi, J.C.: Dosage compensation in *Drosophila*—a model for the coordinate regulation of transcription. Genetics **204**(2), 435–450 (2016). doi:10.1534/genetics.115.185108. http://www.genetics.org/content/204/2/435.full.pdf
29. Pane, A., De Simone, A., Saccone, G., Polito, C.: Evolutionary conservation of *Ceratitis capitata transformer* gene function. Genetics **171**(2), 615–624 (2005). doi:10.1534/genetics.105.041004. http://www.genetics.org/content/171/2/615.full.pdf+html
30. Bopp, D.: About females and males: continuity and discontinuity in flies. J Genet **89**(3), 315–323 (2010)
31. Scott, M.J., Pimsler, M.L., Tarone, A.M.: Sex determination mechanisms in the Calliphoridae (blow flies). Sex. Dev. **8**(1-3), 29–37 (2014). doi:10.1159/000357132
32. Meccariello, A., Salvemini, M., Primo, P., Hall, B., Koskinioti, P., Dalikova, M., Gravina, A., Gucciardino, M.A., Forlenza, F., Gregoriou, M.E., Ippolito, D., Monti, S.M., Petrella, V., Perrotta, M.M., Schmeing, S., Ruggiero, A., Scolari, F., Giordano, E., Tsoumani, K.T., Marec, F., Windbichler, N., Nagaraju, J., Arunkumar, K.P., Bourtzis, K., Mathiopoulos, K.D., Ragoussis, J., Vitagliano, L., Tu, Z., Papathanos, P.A., Robinson, M.D., Saccone, G.: *Maleness-on-the-Y* (*MoY*) orchestrates male sex determination in major agricultural fruit fly pests. bioRxiv, 533646 (2019). doi:10.1101/533646. https://www.biorxiv.org/content/early/2019/02/07/533646.full.pdf
33. Vicoso, B., Charlesworth, B.: Evolution on the X chromosome: unusual patterns and processes. Nat. Rev. Genet. **7**, 645–653 (2006)
34. Gurbich, T.A., Bachtrog, D.: Gene content evolution on the X chromosome. Curr. Opin. Genet. Dev. **18**, 493–498 (2008)
35. Wright, A.E., Dean, R., Zimmer, F., Mank, J.E.: How to make a sex chromosome. Nat. Commun. **7**, 12087 (2016)
36. Wiegmann, B.M., Trautwein, M.D., Winkler, I.S., Barr, N.B., Kim, J.-W., Lambkin, C., Bertone, M.A., Cassel, B.K., Bayless, K.M., Heimberg, A.M., Wheeler, B.M., Peterson, K.J., Pape, T., Sinclair, B.J., Skevington, J.H., Blagoderov, V., Caravas, J., Kutty, S.N., Schmidt-Ott, U., Kampmeier, G.E., Thompson, F.C., Grimaldi, D.A., Beckenbach, A.T., Courtney, G.W., Friedrich, M., Meier, R., Yeates, D.K.: Episodic radiations in the fly tree of life. Proc. Natl. Acad. Sci. U.S.A. **108**(14), 5690–5695 (2011). doi:10.1073/pnas.1012675108. http://www.pnas.org/content/108/14/5690.full.pdf+html
37. Muller, H.J.: Bearings of the 'Drosophila' work on systematics. In: Huxley, J. (ed.) The New Systematics, pp. 185–268. Clarendon Press, Oxford, Oxford (1940)
38. Schaeffer, S.W., Bhutkar, A., McAllister, B.F., Matsuda, M., Matzkin, L.M., O'Grady, P.M., Rohde, C., Valente, V.L.S., Aguade, M., Anderson, W.W., Edwards, K., Garcia, A.C.L., Goodman, J., Hartigan, J., Kataoka, E., Lapoint, R.T., Lozovsky, E.R., Machado, C.A., Noor, M.A.F., Papaceit, M., Reed, L.K., Richards, S., Rieger, T.T., Russo, S.M., Sato, H., Segarra, C., Smith, D.R., Smith, T.F., Strelets, V., Tobari, Y.N., Tomimura, Y., Wasserman, M., Watts, T., Wilson, R., Yoshida, K., Markow, T.A., Gelbart, W.M., Kaufman, T.C.: Polytene chromosomal maps of 11 Drosophila species: the order of genomic scaffolds inferred from genetic and physical maps. Genetics **179**, 1601–1655 (2008). doi:10.1534/genetics.107.086074
39. Johansson, A.-M., Stenberg, P., Allgardsson, A., Larsson, J.: POF regulates the expression of genes on the fourth chromosome in *Drosophila melanogaster* by binding to nascent RNA. Mol. Cell Biol. **32**, 2121–2134 (2012)
40. Linger, R.J., Belikoff, E.J., Scott, M.J.: Dosage compensation of X-linked Muller element F genes but not X-linked transgenes in the Australian sheep blowfly. PLoS ONE **10**(10), 0141544 (2015). doi:10.1371/journal.pone.0141544
41. Vicoso, B., Bachtrog, D.: Numerous transitions of sex chromosomes in Diptera. PLoS Biol. **13**(4), 1002078 (2015). doi:10.1371/journal.pbio.1002078
42. Davis, R.J., Belikoff, E.J., Scholl, E.H., Li, F., Scott, M.J.: *no blokes* is essential for male viability and X chromosome gene expression in the Australian sheep blowfly. Curr. Biol. **28**(12), 1987–1992 (2018). doi:10.1016/j.cub.2018.05.005

43. Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A.J., Aspöck, U., Aspöck, H., Bartel, D., Blanke, A., Berger, S., Böhm, A., Buckley, T.R., Calcott, B., Chen, J., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S., Huang, Y., Jermiin, L.S., Kawahara, A.Y., Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H., Li, Y., Li, Z., Li, J., Lu, H., Machida, R., Mashimo, Y., Kapli, P., McKenna, D.D., Meng, G., Nakagaki, Y., Navarrete-Heredia, J.L., Ott, M., Ou, Y., Pass, G., Podsiadlowski, L., Pohl, H., von Reumont, B.M., Schütte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song, W., Su, X., Szucsich, N.U., Tan, M., Tan, X., Tang, M., Tang, J., Timelthaler, G., Tomizuka, S., Trautwein, M., Tong, X., Uchifune, T., Walzl, M.G., Wiegmann, B.M., Wilbrandt, J., Wipfler, B., Wong, T.K.F., Wu, Q., Wu, G., Xie, Y., Yang, S., Yang, Q., Yeates, D.K., Yoshizawa, K., Zhang, Q., Zhang, R., Zhang, W., Zhang, Y., Zhao, J., Zhou, C., Zhou, L., Ziesmann, T., Zou, S., Li, Y., Xu, X., Zhang, Y., Yang, H., Wang, J., Wang, J., Kjer, K.M., Zhou, X.: Phylogenomics resolves the timing and pattern of insect evolution. Science **346**(6210), 763–767 (2014). doi:10.1126/science.1257570. http://science.sciencemag.org/content/346/6210/763.full.pdf

44. Ross, M.H., Cochran, D.G.: Genetics of the German cockroach. Comp. Biochem. Physiol. A., Comp. Physiol. **94**(4), 551–554 (1989)

45. Keil, C.B., Ross, M.H.: C-banded meiotic karyotype of *Blattella germanica* from prophase I cells. J. Hered. **75**(3), 185–190 (1984). doi:10.1093/oxfordjournals.jhered.a109909

46. Harrison, M.C., Jongepier, E., Robertson, H.M., Arning, N., Bitard-Feildel, T., Chao, H., Childers, C.P., Dinh, H., Doddapaneni, H., Dugan, S., Gowin, J., Greiner, C., Han, Y., Hu, H., Hughes, D.S.T., Huylmans, A.-K., Kemena, C., Kremer, L.P.M., Lee, S.L., Lopez-Ezquerra, A., Mallet, L., Monroy-Kuhn, J.M., Moser, A., Murali, S.C., Muzny, D.M., Otani, S., Piulachs, M.-D., Poelchau, M., Qu, J., Schaub, F., Wada-Katsumata, A., Worley, K.C., Xie, Q., Ylla, G., Poulsen, M., Gibbs, R.A., Schal, C., Richards, S., Belles, X., Korb, J., Bornberg-Bauer, E.: Hemimetabolous genomes reveal molecular basis of termite eusociality. Nat. Ecol. Evol. **2**, 557–566 (2018). doi:10.1038/s41559-017-0459-1

47. Benaglia, T., Chauveau, D., Hunter, D.R., Young, D.S.: mixtools: an R package for analyzing mixture models. J Stat Softw **32**(6), 6 (2009)

48. Larsson, J., Chen, J.D., Rasheva, V., Rasmuson-Lestander, .A.s., Pirrotta, V.: Painting of fourth, a chromosome-specific protein in Drosophila. Proc. Natl. Acad. Sci. U.S.A. **98**, 6273–6278 (2001). doi:10.1073/pnas.111581298

49. Johansson, A.-M., Stenberg, P., Bernhardsson, C., Larsson, J.: Painting of fourth and chromosome-wide regulation of the 4th chromosome in *Drosophila melanogaster*. EMBO J. **26**, 2307–2316 (2007)

50. Johansson, A.-M., Stenberg, P., Pettersson, F., Larsson, J.: POF and HP1 bind expressed exons, suggesting a balancing mechanism for gene regulation. PLoS Genet. **3**, 209 (2007)

51. Seum, C., Reo, E., Peng, H., Rauscher, F.J. III, Spierer, P., Bontron, S.: *Drosophila* SETDB1 is required for chromosome 4 silencing. PLoS Genet. **3**(5), 76 (2007). doi:10.1371/journal.pgen.0030076

52. Tzeng, T.-Y., Lee, C.-H., Chan, L.-W., Shen, C.-K.J.: Epigenetic regulation of the *Drosophila* chromosome 4 by the histone H3K9 methyltransferase dSETDB1. Proc. Natl. Acad. Sci. U.S.A. **104**(31), 12691–12696 (2007). doi:10.1073/pnas.0705534104. http://www.pnas.org/content/104/31/12691.full.pdf

53. Brower-Toland, B., Riddle, N.C., Jiang, H., Huisinga, K.L., Elgin, S.C.R.: Multiple SET methyltransferases are required to maintain normal heterochromatin domains in the genome of *Drosophila melanogaster*. Genetics **181**(4), 1303–1319 (2009). doi:10.1534/genetics.108.100271. http://www.genetics.org/content/181/4/1303.full.pdf+html

54. Figueiredo, M.L.A., Philip, P., Stenberg, P., Larsson, J.: HP1a recruitment to promoters is independent of H3K9 methylation in *Drosophila melanogaster*. PLoS Genet. **8**(11), 1003061 (2012). doi:10.1371/journal.pgen.1003061

55. Lundberg, L.E., Stenberg, P., Larsson, J.: HP1a, Su(var)3-9, SETDB1 and POF stimulate or repress gene expression depending on genomic position, gene length and expression pattern in *Drosophila melanogaster*. Nucl. Acids Res. **41**(8), 4481–4494 (2013). doi:10.1093/nar/gkt158. http://nar.oxfordjournals.org/content/41/8/4481.full.pdf+html

56. Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Wang, Z., Yamashita,

R.A., Zhang, D., Zheng, C., Geer, L.Y., Bryant, S.H.: CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. **45**(D1), 200–203 (2017). doi:10.1093/nar/gkw1129

57. Koch, C.M., Honemann-Capito, M., Egger-Adam, D., Wodarz, A.: Windei, the *Drosophila* homolog of mAM/MCAF1, is an essential cofactor of the H3K9 methyl transferase dSETDB1/Eggless in germ line development. PLoS Genet. **5**(9), 1000644 (2009). doi:10.1371/journal.pgen.1000644

58. Herraiz, A., Belles, X., Piulachs, M.-D.: Chorion formation in panoistic ovaries requires windei and trimethylation of histone 3 lysine 9. Exp. Cell Res. **320**(1), 46–53 (2014). doi:10.1016/j.yexcr.2013.07.006

59. Boyes, J.W., Van Brink, J.M.: Chromosomes of calyptrate diptera. Can. J. Genet. Cytol. **7**(4), 537–550 (1965). doi:10.1139/g65-073. http://dx.doi.org/10.1139/g65-073

60. White, M.J.D.: Animal Cytology and Evolution. Cambridge Univ. Press, ??? (1973)

61. Lifschytz, E., Lindsley, D.L.: The role of X-chromosome inactivation during spermatogenesis. Proc. Natl. Acad. Sci. U.S.A. **69**, 182–186 (1972). doi:10.1073/pnas.69.1.182

62. Bhutkar, A., Russo, S.M., Smith, T.F., Gelbart, W.M.: Genome-scale analysis of positionally relocated genes. Genome Res. **17**, 1880–1887 (2007). doi:10.1101/gr.7062307

63. Meisel, R.P., Han, M.V., Hahn, M.W.: A complex suite of forces drives gene traffic from *Drosophila* X chromosomes. Genome Biol. Evol. **1**, 176–188 (2009)

64. Ding, S., Li, X., Wang, N., Cameron, S.L., Mao, M., Wang, Y., Xi, Y., Yang, D.: The phylogeny and evolutionary timescale of Muscoidea (Diptera: Brachycera: Calyptratae) inferred from mitochondrial genomes. PLoS ONE **10**(7), 0134170 (2015). doi:10.1371/journal.pone.0134170

65. Spradling, A.C., Mahowald, A.P.: Amplification of genes for chorion proteins during oogenesis in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U.S.A. **77**(2), 1096–1100 (1980). doi:10.1073/pnas.77.2.1096

66. Claycomb, J.M., Orr-Weaver, T.L.: Developmental gene amplification: insights into DNA replication and gene expression. Trends Genet. **21**(3), 149–162 (2005). doi:10.1016/j.tig.2005.01.009

67. Wexler, J., Delaney, E.K., Belles, X., Schal, C., Wada-Katsumata, A., Amicucci, M., Kopp, A.: Hemimetabolous insects elucidate the origin of sexual development via alternative splicing. bioRxiv, 587964 (2019). doi:10.1101/587964. https://www.biorxiv.org/content/early/2019/03/26/587964.full.pdf

68. Meisel, R.P., Gonzales, C.A., Luu, H.: The house fly Y Chromosome is young and minimally differentiated from its ancient X Chromosome partner. Genome Res. **27**(8), 1417–1426 (2017). doi:10.1101/gr.215509.116. http://genome.cshlp.org/content/27/8/1417.full.pdf+html

69. Sturtevant, A.H., Novitski, E.: The homologies of the chromosome elements in the genus Drosophila. Genetics **26**, 517–541 (1941)

70. Powell, J.R.: Progress and Prospects in Evolutionary Biology: The Drosophila Model. Oxford University Press, New York (1997)

71. Li, H.: Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv **1303.3997v2** (2013). 1303.3997

72. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A.: The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. **20**(9), 1297–1303 (2010). doi:10.1101/gr.107524.110. http://genome.cshlp.org/content/20/9/1297.full.pdf+html

73. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., Daly, M.J.: A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. **43**(5), 491–498 (2011)

74. Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K.V., Altshuler, D., Gabriel, S., DePristo, M.A.: From fastq data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. Curr. Protoc. Bioinformatics. **43**, 1–33 (2013). doi:10.1002/0471250953.bi1110s43

75. Drinnenberg, I.A., deYoung, D., Henikoff, S., Malik, H.S.: Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. eLife **3**, 03676 (2014). doi:10.7554/eLife.03676

76. Robertson, H.M., Baits, R.L., Walden, K.K.O., Wada-Katsumata, A., Schal, C.: Enormous expansion of the chemosensory gene repertoire in the omnivorous german cockroach *Blattella germanica*. J. Exp. Zool. B **330**(5),

265–278 (2018). doi:10.1002/jez.b.22797. https://onlinelibrary.wiley.com/doi/pdf/10.1002/jez.b.22797

77. Bray, N.L., Pimentel, H., Melsted, P., Pachter, L.: Near-optimal probabilistic RNA-seq quantification. Nat. Biotech. **34**, 525–527 (2016)

78. Love, M., Huber, W., Anders, S.: Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. **15**(12), 550 (2014). doi:10.1186/s13059-014-0550-8

79. Edgar, R.C.: MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res. **32**, 1792–1797 (2004). doi:10.1093/nar/gkh340

80. The UniProt Consortium: UniProt: the universal protein knowledgebase. Nucl. Acids Res. **45**(D1), 158–169 (2017). doi:10.1093/nar/gkw1099

81. Gramates, L.S., Marygold, S.J., Santos, G.d., Urbano, J.-M., Antonazzo, G., Matthews, B.B., Rey, A.J., Tabone, C.J., Crosby, M.A., Emmert, D.B., Falls, K., Goodman, J.L., Hu, Y., Ponting, L., Schroeder, A.J., Strelets, V.B., Thurmond, J., Zhou, P., the FlyBase Consortium: FlyBase at 25: looking to the future. Nucl. Acids Res. **45**(D1), 663–671 (2017). doi:10.1093/nar/gkw1016

82. Tamura, K., Subramanian, S., Kumar, S.: Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks. Mol. Biol. Evol. **21**, 36–44 (2004)

Figures

Figure 1 Insect phylogeny and sex chromosomes. Evolutionary relationships and sex chromosome karyotypes of major insect groups. The phylogenetic topology and time to common ancestor are shown [43], but the relative branch lengths are not drawn to scale. Information on insect sex chromosomes and sex determination are reviewed elsewhere [3, 13, 16, 27].

Figure 2 Reduced male-to-female sequence coverage of Muller element F genes. (A) The distribution of $\log_2 \frac{M}{F}$ for all annotated genes in the *B. germanica* genome is shown, truncated to not include extreme values. **(B)** Boxplots show the distributions of $\log_2 \frac{M}{F}$ for *B. germanica* with homologs on one of the six *D. melanogaster* Muller elements. The red dashed line indicates the expectation of $\log_2 \frac{M}{F} = -1$ for X-linked genes. Each element F gene is shown as a red dot on top of the box plot. The two normal distributions in the mixed distribution of element F genes is shown next to the boxplot. **(C)** The percent of *B. germanica* genes with $\log_2 \frac{M}{F} < -0.5$ that have *D. melanogaster* homologs on each of the six Muller elements is plotted. The 95% confidence intervals (CIs) of the expected percent of genes for each Muller element are shown by the error bars. Observed percentages that lie outside the CI indicate an excess or deficiency of genes on an element with moderately female-biased coverage.

Figure 3 Reduced male heterozygosity in Muller element F genes. (A-B) Boxplots show the distributions of heterozygous variants per Mb in males and females within genes assigned to each Muller element on a $\log_{10}$ scale. Each element F gene is shown as a red dot on top of the box plot. **(C)** The points in the scatterplot show the $\log_2 \frac{M}{F}$ sequencing coverage and fraction of heterozygous variants in males for genes assigned to Muller elements, with element A–E genes in gray and element F genes in red. The standard deviations of $\log_2 \frac{M}{F}$ coverage and fraction of heterozygous variants in males are shown for element A–E genes in black and element F genes in red.

Figure 4 No reduced male expression of element F genes. (A-B) Boxplots show the distributions of $\log_2 \mathrm{FC}$ of expression between either male and mixed sex heads or male heads and female whole adults for genes with *D. melanogaster* homologs on each Muller element. Each element F gene is shown as a red dot on top of the box plot. **(C-D)** The points in the scatterplots show the $\log_2 \frac{M}{F}$ sequencing coverage and $\log_2 \mathrm{FC}$ of expression for genes assigned to Muller elements, with element A–E genes in gray and element F genes in red. The standard deviations of $\log_2 \frac{M}{F}$ coverage and $\log_2 \mathrm{FC}$ expression are shown for element A–E genes in black and element F genes in red.

Figure 5 Three key regulators of element F transcription and chromatin are conserved in cockroach. Lines show the percent amino acid sequence conservation. The coordinates of the following predicted functional domains are shown as gray boxes in each graph: **(A-B)** RNA binding domain (RBD); **(C-D)** coiled-coil domain (CC), tudor domain (T), methyl-CpG-binding domain (MBD), and SET domain; **(E-F)** CC domain, and fibronectin type III repeats (F-III). **(C-D)** Predicted pre-SET domains are shown as white boxes next to SET domains. **(C)** The region of *D. melanogaster* Egg that interacts with Wde is shown by a black box, **(E)** as is the region of Wde that interacts with Egg.

Tables

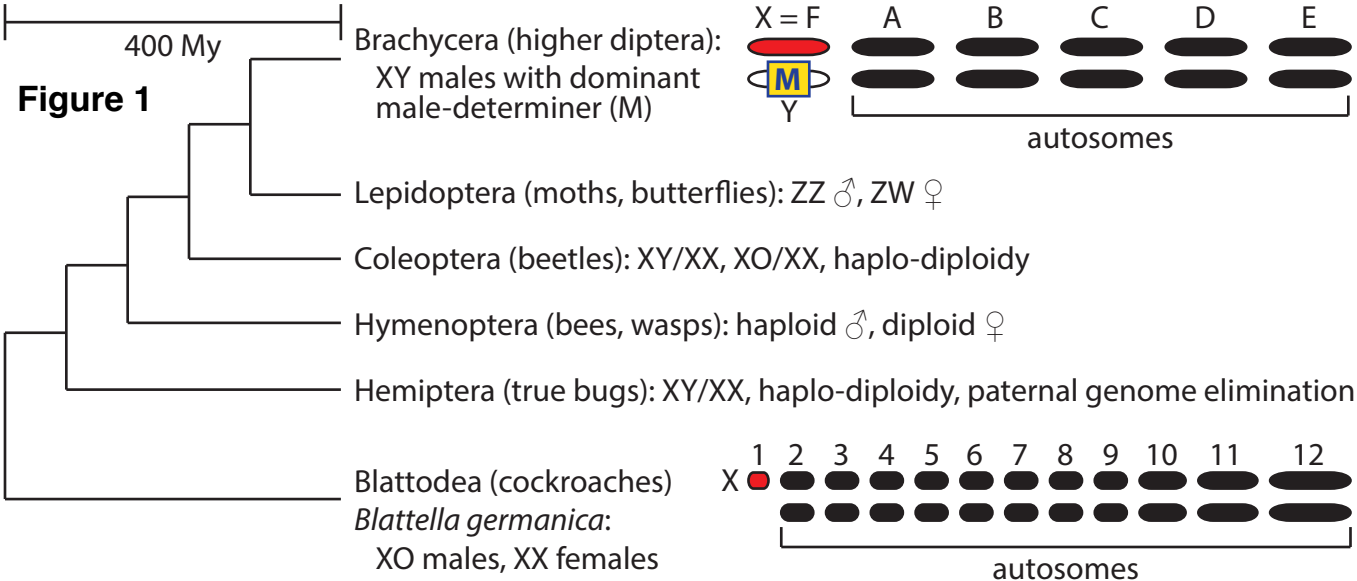
Table 1 Genes with female-biased sequencing coverage and a *D. melanogaster* homolog on each Muller element.

	Muller element						
	A	B	C	D	E	F	total
# genes	898	999	1138	1159	1413	51	5658
female-biased ($\log_2 \frac{M}{F} < -0.5$)	53	40	66	58	65	20	302
% female-biased	5.90	4.00	5.80	5.00	4.60	39.22	5.34

Additional Files

Additional file 1 — Gene coverage, heterozygosity, and expression
Data table showing read mapping coverage, heterozygosity, and gene expression analysis for each cockroach gene.

Additional file 2 — qPCR validation of 2 candidate X-linked genes
Methods and results of qPCR validation of cockroach genes with evidence of X-linkage.


400 My
Figure 1
Brachycera (higher diptera):
XY males with dominant
male-determiner (M)
X = F
A
B
C
D
E
M
Y
autosomes
Lepidoptera (moths, butterflies): ZZ ♂, ZW ♀
Coleoptera (beetles): XY/XX, XO/XX, haplo-diploidy
Hymenoptera (bees, wasps): haploid ♂, diploid ♀
Hemiptera (true bugs): XY/XX, haplo-diploidy, paternal genome elimination
Blattodea (cockroaches)
Blattella germanica:
XO males, XX females
X
1 2 3 4 5 6 7 8 9 10 11 12
autosomes

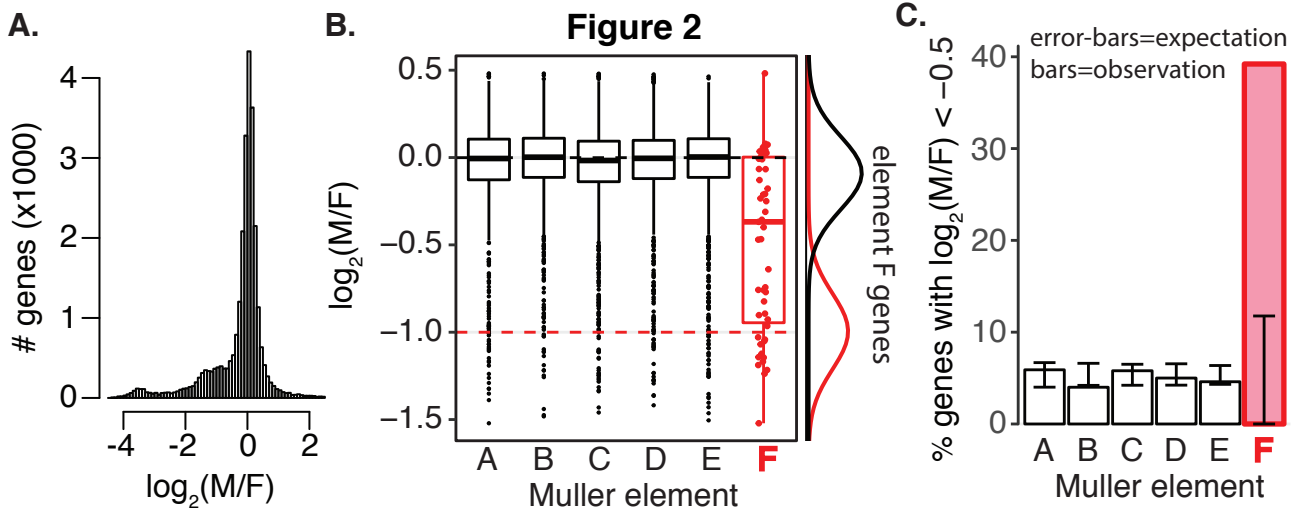


Figure 2
A.
genes (x1000)
4
3
2
1
0
-4
-2
0
2
log2(M/F)
B.
log2(M/F)
0.5
0.0
−0.5
−1.0
−1.5
A
B
C
D
E
F
Muller element
element F genes
C.
% genes with log2(M/F) < −0.5
error-bars=expectation
bars=observation
40
30
20
10
0
A
B
C
D
E
F
Muller element

Figure 3

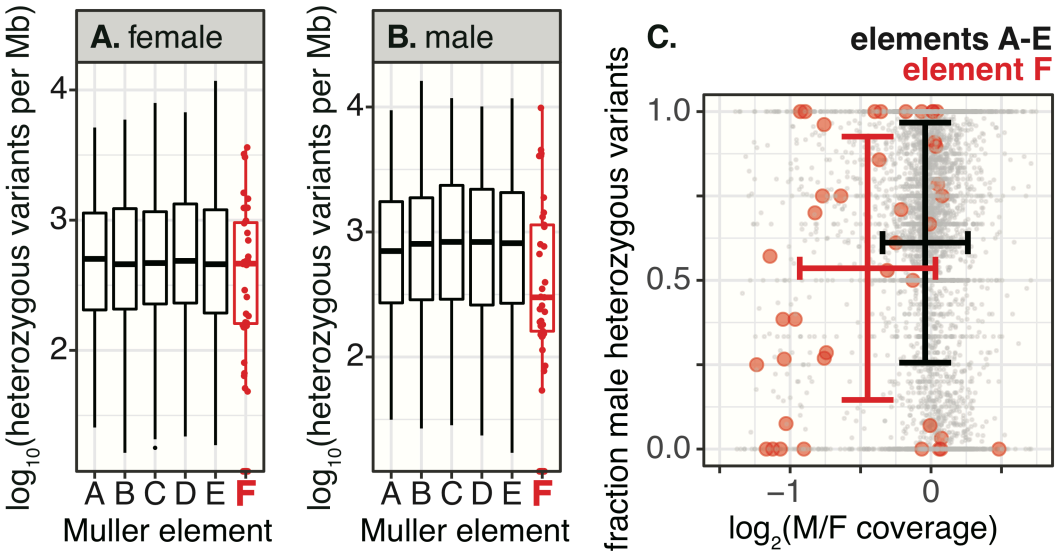

A. female
B. male
C.
elements A-E
element F
log10(heterozygous variants per Mb)
fraction male heterozygous variants
Muller element
A B C D E F
log2(M/F coverage)
4
3
2
1.0
0.5
0.0
−1
0

Figure 4

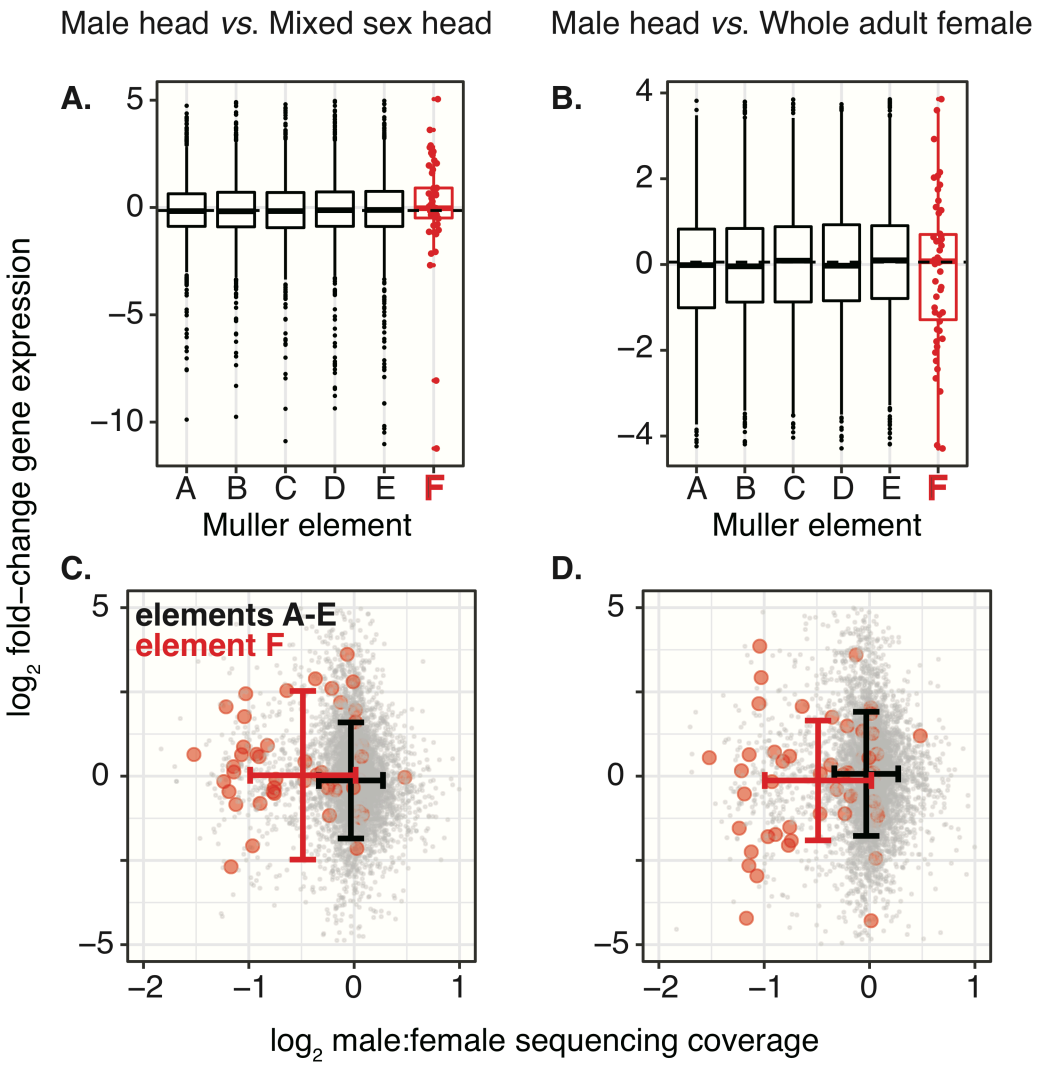
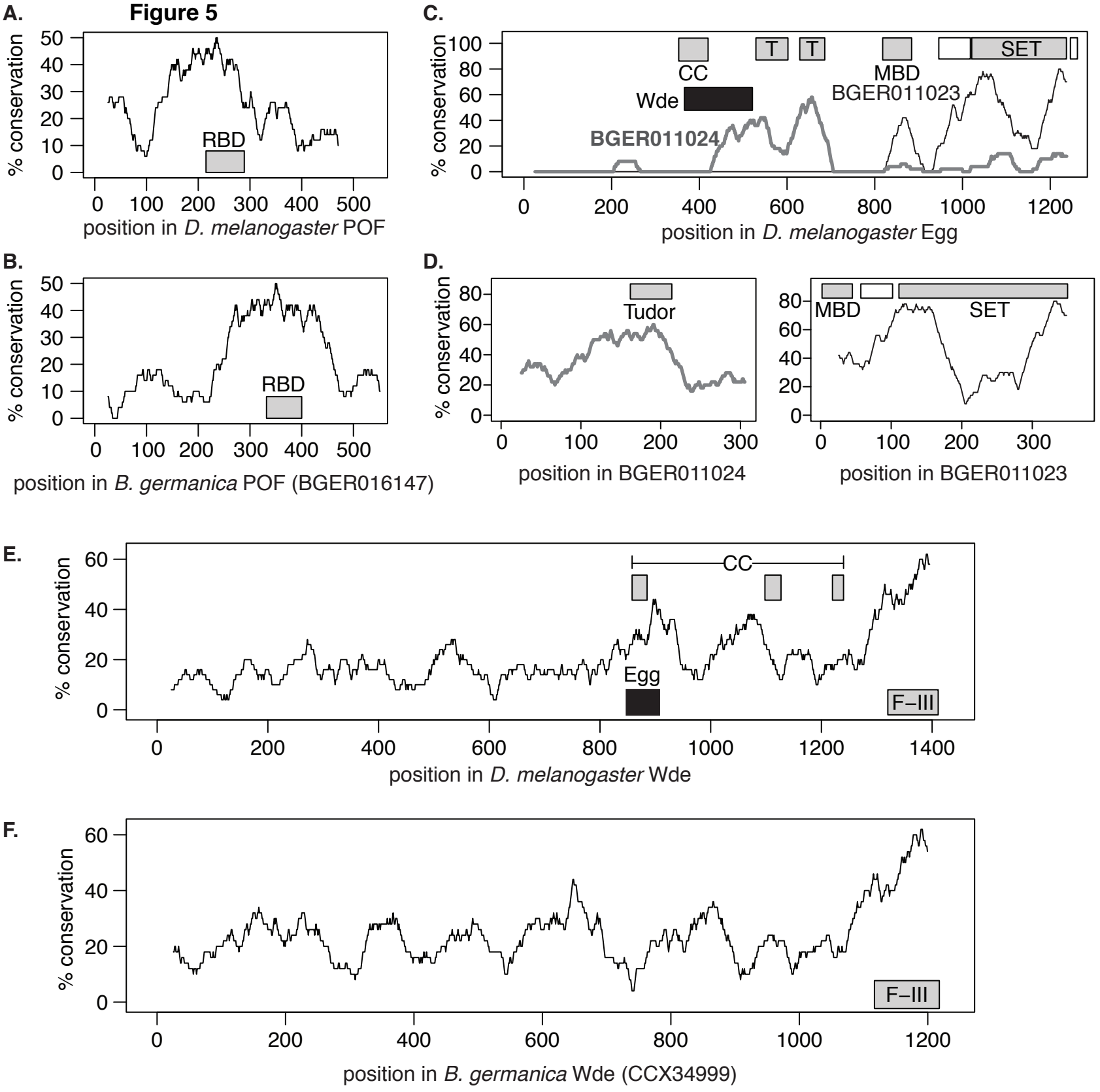

Male head vs. Mixed sex head
Male head vs. Whole adult female
A.
B.
C.
D.
log2 fold−change gene expression
Muller element
A B C D E F
elements A-E
element F
log2 male:female sequencing coverage

Figure 5


A.
% conservation
RBD
position in *D. melanogaster* POF
B.
% conservation
RBD
position in *B. germanica* POF (BGER016147)
C.
% conservation
CC
T
T
MBD
SET
Wde
BGER011023
BGER011024
position in *D. melanogaster* Egg
D.
% conservation
Tudor
position in BGER011024
MBD
SET
position in BGER011023
E.
% conservation
CC
Egg
F–III
position in *D. melanogaster* Wde
F.
% conservation
F–III
position in *B. germanica* Wde (CCX34999)