

RESEARCH ARTICLE

Open Access



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. Delclos¹ and Judith R. Wexler^{2,3}

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. Cytological similarities between the X chromosomes of the German cockroach (*Blattella germanica*) and most flies suggest that they may be homologous—possibly representing an extreme case of long-term conservation.

Results: To test the hypothesis that the cockroach and fly X chromosomes are homologous, we analyzed whole-genome sequence data from cockroaches. We found evidence in both sequencing coverage and heterozygosity that a significant excess of the same genes are on both the cockroach and fly X chromosomes. We also present evidence that the candidate X-linked cockroach genes may be dosage compensated in hemizygous males. Consistent with this hypothesis, 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 may represent the convergent evolution of the X chromosome in the lineages leading to cockroaches and flies. Alternatively, the common ancestor of most insects may have had an X chromosome that resembled the extant cockroach and fly X. Cockroaches and flies diverged ~ 400 million years ago, which would be the longest documented conservation of a sex chromosome. Cockroaches and flies have different mechanisms of sex determination, raising the possibility that the X chromosome was conserved despite the 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].

*Correspondence: rmeisel@uh.edu

¹Department of Biology and Biochemistry, University of Houston, 3455 Cullen Blvd., Houston, TX 77204, USA

Full list of author information is available at the end of the article



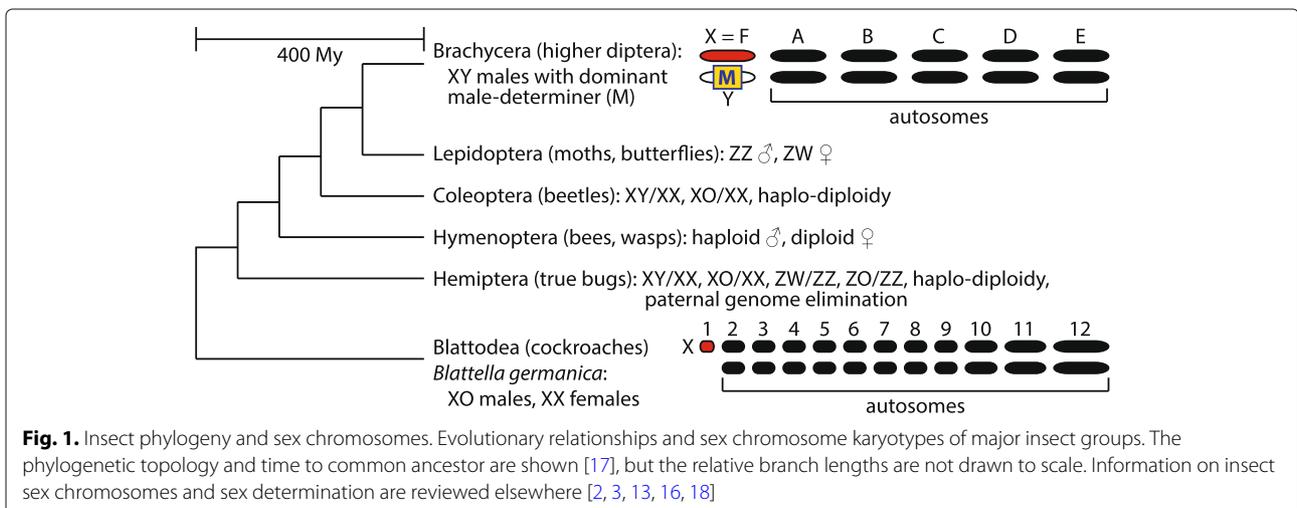
Sex-determining pathways and sex chromosomes can evolve rapidly, often differing between closely related species [2, 3]. Evolutionary transitions in sex determination pathways are often accompanied by corresponding changes in the identity of the sex chromosomes [1, 2, 12]. Transitions in sex-determining pathways and turnover of sex chromosomes are well studied across insects, where there is a diversity of sex determination mechanisms [13–16] (Fig. 1). For example, the genetic factors that initiate sex determination in *Drosophila* do not determine sex in other flies [19–26]. In addition, the sex chromosomes of *Drosophila* are not homologous to the sex chromosomes of other flies [18, 27, 28]. 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 [18, 29].

It is most parsimonious to conclude that the ancestral sex determination system of brachyceran dipterans (which includes flies but excludes mosquitoes, crane flies, midges, gnats) consists of a Y-linked male-determining factor that regulates the splicing of the *transformer* (*tra*) gene product [15, 22, 26, 30–33]. 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 chromosome is known as Muller element F [18]. Element F has reverted to an autosome in *D. melanogaster*, where it is also known as chromosome 4 or the “dot” chromosome. The dot chromosome is enriched for heterochromatin and has fewer than 100 genes [34]. Element F is notable because most X chromosomes are gene rich and euchromatic, despite having some differences in gene

content from the autosomes [35–37]. This peculiar element F X chromosome has been conserved for >150 million years (My) in some fly lineages, but it reverted to an autosome in *Drosophila* when a different chromosome became X-linked [18, 38]. The remainder of the fly genome is organized into 5 euchromatic chromosomes (or chromosome arms), named Muller elements A–E [39, 40]. Element A is the X chromosome in *D. melanogaster*.

There is some evidence that the X-linked element F is dosage compensated in hemizygous males. In *D. melanogaster*, where element F is autosomal, *Painting of fourth* (*Pof*) encodes an RNA-binding protein that localizes predominantly to element F [41]. *Lucilia cuprina* (Australian sheep blowfly) has the ancestral brachyceran karyotype, with an X-linked element F [42, 43]. Expression of X-linked genes is upregulated in *L. cuprina* males by the homolog of *Pof* [42, 44]. This dosage compensation is essential for male viability—a loss of function mutation in the *L. cuprina* homolog of *Pof* is male lethal, but viable in females [44].

The German cockroach, *Blattella germanica*, diverged from flies ~ 400 My ago (Mya) [17]. Female cockroaches are XX and males are XO, i.e., one X and no Y chromosome [13, 45]. 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 [46], reminiscent of element F, the ancestral brachyceran X chromosome. We tested the hypothesis that the German cockroach X chromosome is homologous to fly element F, which would suggest that a cockroach and most flies share an X chromosome despite ~ 400 My divergence.



Results

Decreased sequencing coverage of element F homologs 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 [18]. We used available whole-genome sequencing data [47] to calculate the relative coverage of male (M) and female (F) 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 (Fig. 2a), as expected, because we recalibrated the $\log_2 \frac{M}{F}$ values to have a median of 0 (see the “Methods” section). 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 (Fig. 2a). In total, 3499 of the 28,141 annotated genes have female-biased coverage ($\log_2 \frac{M}{F} \leq -1$), whereas only 1363 genes have male-biased coverage ($\log_2 \frac{M}{F} \geq 1$), consistent with a heavy shoulder of X-linked genes. Assuming the 1363 male-biased genes represent the false-positive rate, we expect 2136/3499 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 [46], which means that fewer than 2345 genes (28,141/12) should be X-linked.

To test the hypothesis that the German cockroach X chromosome is homologous to the ancestral brachyceran fly X (i.e., Muller element F), we evaluated if cockroach genes with *D. melanogaster* homologs on element F have lower $\log_2 \frac{M}{F}$ than genes with homologs on the other 5 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 (Fig. 2b). In contrast, the 51 cockroach element F homologs have a median $\log_2 \frac{M}{F} < 0$, and the average $\log_2 \frac{M}{F}$ for element F homologs is significantly less than the other genes ($P=10^{-10}$ using a Mann-Whitney U test comparing element F homologs with elements A–E). If all element F homologs 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 homologs have a median $\log_2 \frac{M}{F} > -1$. Therefore, we hypothesize that a disproportionate amount of, but not all, element F homologs are X-linked in German cockroach.

We next estimated the frequency of element F homologs that are X-linked in the German cockroach. First, we used the mclust package in R to fit a mixture of normal distributions to the $\log_2 \frac{M}{F}$ values of element F homologs [48]. The best fitting mixture consists of 3 distributions, with 1 centered at a mean of -1.02 (Table 1), close to the expectation of $\log_2 \frac{M}{F} = -1$ for X-linked genes. This suspected X-linked distribution contains $\sim 41\%$ of the 51 element F homologs, and it has very little overlap with the other 2 distributions (Fig. 2b). One of the other 2 distributions is centered very close to 0 (the expectation for autosomal genes), and it has very low variance. The third distribution has a mean $\log_2 \frac{M}{F} = -0.23$ and a large variance. We suspect that the 2 distributions with $\log_2 \frac{M}{F} > -1$ correspond to element F homologs that are autosomal in *B. germanica*. These 2 distributions may be the result of fitting normal distributions to a single non-normal distribution with a mode at $\log_2 \frac{M}{F} = 0$ and a long tail extending into $\log_2 \frac{M}{F} < 0$. Consistent with this hypothesis, when we fit a mixture of 2 normal distributions to the $\log_2 \frac{M}{F}$ values of element F homologs, we obtain 1 distribution with a mean

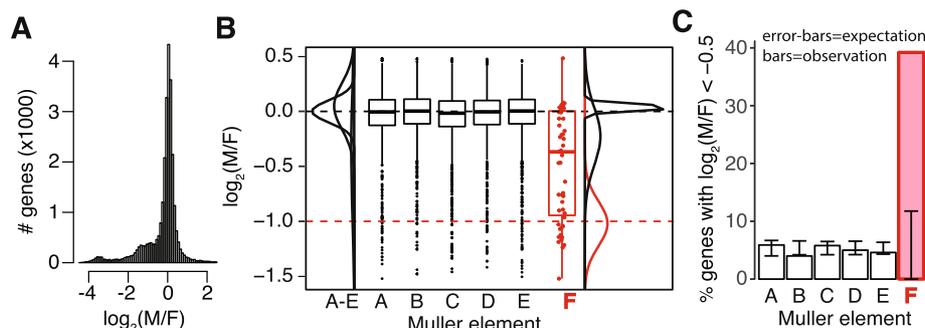


Fig. 2. Reduced male-to-female sequence coverage of Muller element F homologs. **a** The distribution of $\log_2 \frac{M}{F}$ for all annotated genes in the *B. germanica* genome is shown, truncated to not include extreme values. **b** Boxplots show the distributions of $\log_2 \frac{M}{F}$ for *B. germanica* 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 homolog is shown as a red dot on top of the box plot. The normal distributions from the mixture models for element A–E and element F homologs are shown next to the boxplots. **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 homologs on an element with moderately female-biased coverage

Table 1 Counts and proportions of genes assigned to each normal distribution in a mixture model of $\log_2 \frac{M}{F}$ values

Distribution	Proportion	Mean	Variance
Element F genes (51 total)			
1	0.41	-1.020	0.041
2	0.36	-0.225	0.069
3	0.23	0.038	0.001
Element A–E genes (5,602 total)			
1	0.01	-0.888	5.600
2	0.41	0.001	0.010
3	0.41	0.049	0.045
4	0.17	-0.348	0.329

Genes have homologs in *D. melanogaster* on either element F or elements A–E

$\log_2 \frac{M}{F} = -1.00$ that has 43% of element F homologs and a second distribution with a mean $\log_2 \frac{M}{F} = -0.09$ that has 57% of element F homologs (Additional file 2). Moreover, with a mixture of 4 normal distributions, we recover 2 distributions centered near $\log_2 \frac{M}{F} = -1$ that together have 40% of element F homologs. Therefore, regardless of the number of distributions in our mixture model, we recover at least 40% of cockroach element F homologs that fall within a distribution consistent with X-linkage.

In contrast to element F, the $\log_2 \frac{M}{F}$ values for cockroach genes with *D. melanogaster* homologs on elements A–E can be best explained by a mixture of 4 distributions (Table 1). The distribution within this mixture model that is most consistent with X-linkage has a mean of -0.89, a large variance of 5.6, and contains only 37 of the 5602 element A–E homologs. Most element A–E homologs (4957) are assigned to 2 distributions with means of 0.0015 and 0.049, which are both consistent with autosomes (Fig. 2b). Together, our analysis of mixture models suggest that a large fraction of element F homologs are X-linked in German cockroach, whereas the vast majority of element A–E homologs are autosomal.

The distributions of $\log_2 \frac{M}{F}$ seem to describe 2 classes of element F homologs: autosomal genes with $\log_2 \frac{M}{F} > -0.5$ and X-linked genes with $\log_2 \frac{M}{F} < -0.5$ (Fig. 2b). If there is an excess of element F homologs on the cockroach X, we expect a higher frequency of element F homologs to have $\log_2 \frac{M}{F} < -0.5$ than genes on the other 5 elements. We therefore counted the number of genes with $\log_2 \frac{M}{F} < -0.5$ on each of the 6 Muller elements (Table 2). To determine a null distribution of those genes on each element, we randomly assigned the total number of genes with $\log_2 \frac{M}{F} < -0.5$ to the 6 elements based on the size of each Muller element (measured as the total number of cockroach genes on the element) in 1000 bootstrap replicates of the data. A significant excess of cockroach

element F homologs have $\log_2 \frac{M}{F} < -0.5$ relative to our null expectation (Fig. 2c). This provides further evidence that an excess of element F homologs is X-linked in German cockroach.

Reduced heterozygosity of element F homologs 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 polymorphic genetic variants in X-linked genes, whereas males must be hemizygous (only one allele per gene). If element F homologs are X-linked in cockroach, we expect to observe an excess of element F homologs without heterozygous variants in an individual male when compared to element A–E homologs and also when compared to female heterozygosity in element F homologs. To test this prediction, we used the available cockroach genome sequence data to identify heterozygous sequence variants in cockroach genes (Additional file 1).

The German cockroach genome project generated sequence data from a single male and single female of an inbred laboratory strain [47]. We therefore expect to observe no heterozygous variants in the male for X-linked genes, but the female could have heterozygous X-linked variants. However, there are also likely to be errors in variant calling and genotyping that could produce false-positive heterozygous calls. Because of these false positives, we may observe heterozygous variants in element F homologs in males even if the genes are X-linked. To address this limitation, we tested for reduced heterozygosity in element F homologs in males, rather than an absence of heterozygous variants.

We first compared the heterozygosity of cockroach genes in males and females across Muller elements (Fig. 3). In females, there is no significant difference in the 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, male element F homologs have significantly fewer heterozygous variants than genes on elements A–E ($P = 0.017$ in a Mann-Whitney U test). This reduced male heterozygosity in element F homologs is consistent with an excess of element F homologs 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 calculated, for each gene, a ratio of the number male heterozygous variants to the total number of heterozygous variants in the male and female samples. This value ranges from 0 (if a gene only has heterozygous variants in females) to 1 (if a gene only has heterozygous variants in males). Equal heterozygosity in both sexes has

Table 2 Genes with female-biased sequencing coverage and a *D. melanogaster* homolog on each Muller element

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

a value of 0.5. Of the 40 element F homologs 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 (Fig. 3c). This is significantly greater than the 2.5% of element A–E homologs 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 homologs on the German cockroach X chromosome.

Validation of candidate X-linked element F homologs

We selected two element F homologs that we hypothesize are X-linked (BGER000638 and BGER000663) to validate using quantitative PCR (qPCR). Both genes have $\log_2 \frac{M}{F} < -1$, and one gene (BGER000638) has three times as many heterozygous variants in the female compared to the male (Additional file 1). The other gene has no heterozygous variants in either sex. We found that 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 3) [49]. This is the expected result if

both genes are X-linked. Therefore, male:female sequencing coverage, heterozygosity, and qPCR provide consistent evidence that element F homologs are X-linked in German cockroach.

The cockroach X chromosome may be dosage compensated in males

We next tested if the haploid dosage of element F homologs affects their expression in male cockroach. The ideal data to test for the effects of a haploid X are expression measurements from males and females from the same tissue and developmental stage [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 the 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 cockroach genes with homologs

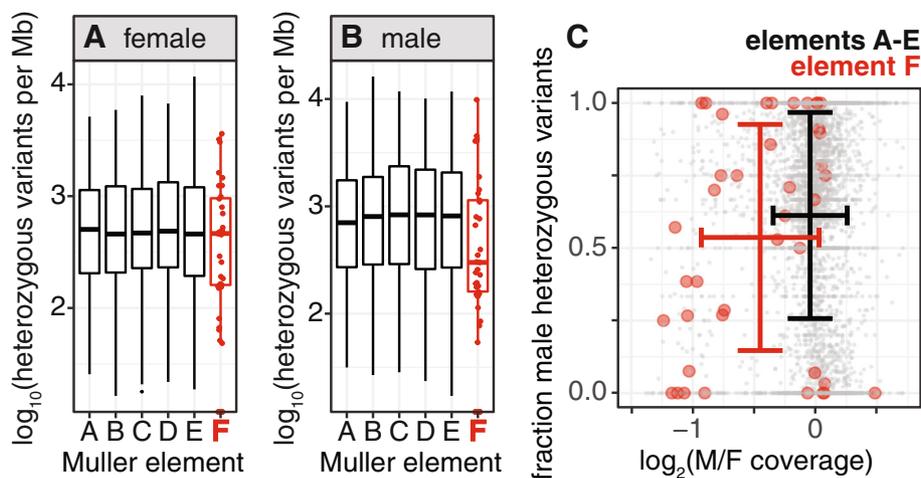


Fig. 3. Reduced male heterozygosity in Muller element F homologs. **a, b** Boxplots show the distributions of heterozygous variants per Mb in males and females within genes assigned to each Muller element on a \log_{10} scale. Each element F homolog is shown as a red dot on top of the box plot. **c** 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 homologs in gray and element F homologs 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 homologs in black and element F homologs in red

on element F and elements A–E. Indeed, there is no significant difference in the median \log_2 FC between element F homologs and element A–E homologs ($P = 0.15$ 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; Fig. 4a, b).

Only a subset of element F homologs is expected to be X-linked in cockroach based on $\log_2 \frac{M}{F}$ sequencing coverage (Fig. 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 homologs with evidence of X-linkage ($\log_2 \frac{M}{F} < -0.5$) and element F homologs that appear to be autosomal ($\log_2 \frac{M}{F} \geq -0.5$). Indeed, there is no significant difference in \log_2 FC between the two subsets of element F homologs ($P=0.84$ for male head vs mixed sex head, $P = 0.30$ for male head vs whole adult females, with both P values from Mann-Whitney U tests; Fig. 4c, d). The same is true for element A–E homologs: there is no significant difference in \log_2 FC of male head vs mixed sex head between low and high coverage element A–E homologs ($P = 0.054$ in a Mann-Whitney U test) nor is there a significant difference

in \log_2 FC of male head vs whole adult female between low and high coverage element A–E homologs ($P = 0.65$ in a Mann-Whitney U test). The comparison of \log_2 FC in male vs mixed sex head for element A–E homologs has the lowest P value. If this low P value was evidence for a lack of dosage compensation, we would expect genes with low male sequencing coverage ($\log_2 \frac{M}{F} < -0.5$) to have lower male expression than genes with higher male sequencing coverage ($\log_2 \frac{M}{F} \geq -0.5$). However, genes with low male sequencing coverage have higher male expression (median \log_2 FC = 0.0039) than genes with higher male sequencing coverage (median \log_2 FC = -0.15). Therefore, the limited RNA-seq data that are available suggest that the German cockroach X chromosome may be 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 [18]. Males of the blow fly *L. cuprina* are haploid for such an X chromosome, and their

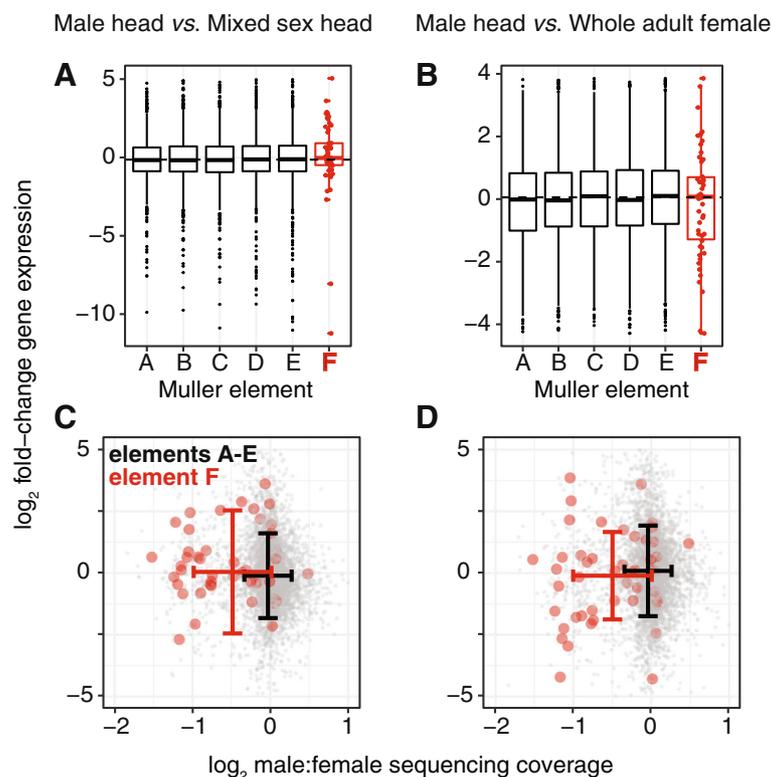


Fig. 4. No reduced male expression of element F homologs. **a, b** Boxplots show the distributions of \log_2 FC of the expression between either male and mixed sex heads or male heads and female whole adults for genes with *D. melanogaster* homologs on each Muller element. Each element F homolog is shown 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 FC of the expression for genes assigned to Muller elements, with element A–E homologs in gray and element F homologs in red. The standard deviations of $\log_2 \frac{M}{F}$ coverage and \log_2 FC expression are shown for element A–E homologs in black and element F homologs in red

X-linked genes are upregulated by an RNA-binding protein encoded by a homolog of *Drosophila Pof* [42, 44]. POF localizes nearly exclusively to element F gene bodies in *D. melanogaster* [41, 50–52]. There is a *Pof* homolog in the cockroach genome (BGER016147), which we aligned to the *D. melanogaster* protein sequence. The most conserved region of *D. melanogaster Pof* overlaps with a predicted RNA-binding domain within the cockroach protein sequence (Fig. 5a, b). Therefore, a key component of the molecular machinery which regulates the dosage compensation on the X-linked fly element F 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 [53–57]. 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* [58]. These predicted functional domains overlap with the portions of the cockroach proteins that are most conserved relative to *D. melanogaster* *Egg* (Fig. 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* [59]. There is one predicted homolog of *wde* in the cockroach genome annotation (BGER025676), but an independently sequenced

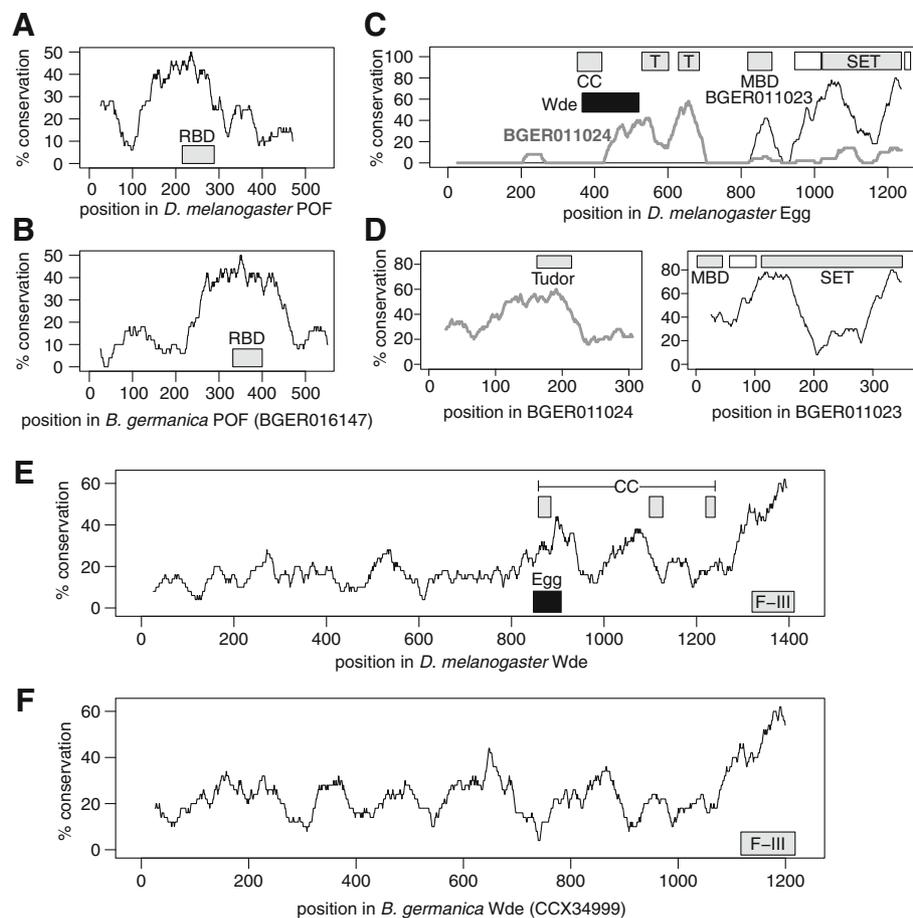


Fig. 5. Three key regulators of element F transcription and chromatin are conserved in cockroaches. Lines show the percent amino acid sequence conservation. The coordinates of the following predicted functional domains are shown as gray boxes in each graph. **a, b** RNA-binding domain (RBD). **c, d** 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** The region of *Wde* that interacts with *Egg*

cockroach *wde* gene (CCX34999) is longer than the *wde* homolog predicted by the automated annotation [60]. 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 [58]. The C-terminal end of CCX34999 is also the most conserved part of the protein relative to *D. melanogaster* Wde (Fig. 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, is among the most conserved regions of the *D. melanogaster* proteins when compared to the cockroach homologs (Fig. 5c, e). Therefore, homologs of *Pof* and its two key interactors are present in the German cockroach genome, showing it is possible 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 Muller element F, which is X-linked in most flies. First, there is a reduced sequencing coverage of nearly half of the Muller element F homologs in male cockroach, consistent with a haploid dose of the X chromosome in males (Fig. 2). Second, there is a decreased heterozygosity of element F homologs in male cockroach, including those with reduced male sequencing coverage (Fig. 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 homologs 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, as opposed to convergent recruitment of the same genes onto the fly and cockroach X. 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 [46], similar to the brachyceran fly X-linked element F [61]. X chromosomes tend to be euchromatic in males [35–37], 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 the testes [62], where sex chromosome-specific heterochromatinization could differ from the normal behavior in somatic cells [63]. 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 homologs with evidence for X-linkage in cockroach greatly exceeds the expectation if the X chromosomes of flies and cockroaches were independently derived (Fig. 2c). Fourth, the fraction of element F homologs that appear to be X-linked in cockroaches (> 40%) 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 the rates of gene relocation between Muller elements in *Drosophila*. If element F was the ancestral X chromosome of the MRCA of flies and cockroaches, 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 [64]. Based on the frequency of gene relocation between Muller elements in *Drosophila* [65] and the sizes of the elements in *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 the “Methods” section for calculations). There are up to 30 (60% of 51) *D. melanogaster* element F homologs that do not have evidence for X-linkage in cockroach (Fig. 2b). Gene movement alone can thus explain 7–8 of these apparently autosomal element F homologs.

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 [44]. Assuming a linear relationship between divergence time [38, 66] and conservation of element F gene content, we would expect only 11.1% of cockroach genes with element F homologs to be X-linked:

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

Our estimate of the fraction of element F homologs that are X-linked in *B. germanica* (> 40%) is in between the estimates predicted based on the 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 movement 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 (Fig. 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 *Drosophila* ovary [67, 68]. 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 may be dosage compensated in males—we find no evidence for reduced expression of element F homologs in male cockroaches, regardless of whether the genes appear to be haploid in males (Fig. 4). Previous work found evidence that the cockroach *tra* homolog may regulate dosage compensation because knockdown of *tra* in cockroach females results in female-specific lethality of their progeny [69]. Here, we found that homologs of genes involved in regulating the expression of element F genes in flies are present in the cockroach genome, with their functional domains conserved (Fig. 5). This is consistent with cockroaches and flies sharing a mechanism of X chromosome dosage compensation that has been conserved since their MRCA. Future work should further investigate if the regulators of sex determination and dosage compensation in flies (e.g., *tra* and *Pof*) have similar roles in cockroach. An important limitation of our analysis is that we did not compare the same tissues between males and females [10, 11]. Our inference of dosage compensation may be confounded by, for example, differences in cell types between tissues [70]. Further work is therefore necessary to more rigorously test for dosage compensation of the cockroach X chromosome with appropriate gene expression comparisons between males and females.

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 fly sex determination system, which relies on a dominant male determiner located on the Y chromosome (Fig. 1). It is unlikely that the same X-linked dosage-sensitive factors determine sex in cockroaches 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 brachyceran flies, which likely used a Y-linked male determiner (Fig. 1). Moreover, the sexually dimorphic splicing of the sex determination pathway gene *tra* differs between German cockroaches and flies [69]. Therefore, we hypothesize that *B. germanica* has a homologous X chromosome with the MRCA of brachyceran flies, 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 [71].

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 would be 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

Drosophila and other fly genomes are organized into six chromosomes (or chromosome arms) known as Muller elements [27, 39, 72, 73]. Muller element F is the ancestral X chromosome of brachyceran flies, and elements A–E are autosomal in flies with this ancestral karyotype [18]. We assigned each *B. germanica* gene with a single *D. melanogaster* homolog to the Muller element of its homolog. We retrieved the *D. melanogaster* homologs of *B. germanica* genes from the Baylor College of Medicine i5k Maker annotation, version 0.5.3 [47]. This annotation pipeline was performed as part of the *B. germanica* genome project [47]. We only assigned *B. germanica* genes to Muller elements if they have a single *D. melanogaster* homolog in the annotation (i.e., we did not include genes with multiple predicted *D. melanogaster* homologs or without any predicted homologs).

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 [18]. 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; [47]), using BWA-MEM with default parameters [74]. 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 assigned to Muller elements is 0.

We used the *mclust* package to fit a mixture of multiple normal distributions to the $\log_2 \frac{M}{F}$ values [48]. We did this separately for element F homologs and genes assigned to elements A–E. The *Mclust()* function uses an expectation-maximization algorithm to obtain maximum likelihood estimators of the mean, variance, and number of genes in each normal distribution. It fits two different models for mixtures of 1 through 9 normal distributions: (1) mixture models where each normal distribution has the same variance (i.e., mixture of univariate normal distributions) and (2) mixture models where the normal distributions have unequal variances. We then compared Bayesian information criteria (BIC) across the nested models to determine the number of normal distributions that fit data the best (Additional file 2). We also compared BIC values to test if the best fitting distributions are univariate or have unequal variances.

Quantitative 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 3). 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 3.

Differential heterozygosity in males and females

We tested for genes with reduced heterozygosity in males (including 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 [75–77]. Because there is no reference variant set for cockroaches, we used the following steps to extract high confidence variants [71]. First, we used Picard Tools version 1.133 to identify and remove duplicate reads, and we realigned indels with GATK. Then, we performed naive variant calling using the GATK HaplotypeCaller with a phred-scaled confidence threshold of 20. 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 also selected the highest confidence insertions and deletions (indels) from the first pass (QD < 2.0, FS > 200, SOR > 10, ReadPosRankSum < -20). 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). All three male sequencing libraries were treated as a single sample in this analysis because they came from the same individual male [47]. We used hard cutoff values because we did not have sufficient data to train a probabilistic variant filter. We then extracted variants that mapped to *B. germanica* genes (from the v. 0.5.3 i5k annotation). Variants were considered to be within a gene if they fell within the beginning and end coordinates of an annotated gene, including within exons or introns.

We identified heterozygous variants as those with two different alleles at that site in either the male or female sample. The two alleles could be either be one reference allele and one alternate, or they could be two alternate alleles. To calculate heterozygous variants per Mb within each gene, we used the differences of the beginning and end coordinates of each annotated gene in the genome assembly as a measure of gene length. To calculate the fraction of heterozygous variants in the male, we counted the number of heterozygous variants in the male (H_m) and female (H_f) samples separately for each gene. We then divided the number of heterozygous variants in the male sample by the sum of the number of heterozygous variants in the

male and female samples for each gene ($H_m/[H_m + H_f]$).

Differential gene expression using RNA-seq data

We compared the expression of genes in adult male heads (NCBI SRA accessions SRX3189901 and SRX3189902) with expression in a mixed sex adult head sample (SRX682022) using available RNA-seq data [78, 79]. We also compared male head expression with expression in whole adult females (SRX2746607 and SRX2746608) [47]. We aligned the RNA-seq reads from each library to *B. germanica* transcripts (from the version 0.5.3 i5k annotation) using kallisto [80]. 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 the expression for each gene between male heads and mixed sex heads, as well as between male heads and whole adult females [81]. All reads from a given accession were treated as belonging to a single 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 (POE, Eggless, and Winder) with their *B. germanica* homologs using MUSCLE [82]. 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 [58] or retrieved from UniProt [83].

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) [65], 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* [84], the divergence time between *D. melanogaster* and *D. pseudoobscura* (54.9 My) [85], and the divergence time between flies and cockroaches

(386.9 My) [17]. 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) [65]. 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:

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

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12915-019-0721-x>.

Additional file 1: Table with supporting data showing read mapping coverage, heterozygosity, and gene expression analysis for each cockroach gene.

Additional file 2: Tables with supporting data showing properties of mixture models fit to $\log_2 \frac{M}{F}$ values for cockroach genes with *D. melanogaster* homologs on element F and elements A–E.

Additional file 3: Methods and results of qPCR validation of 2 candidate X-linked cockroach genes. Supporting Data, Table S1, Figure S1. Table S1: PCR primers used to test for X-linkage in qPCR assay. Figure S1: Relative concentrations of BGER000638 and BGER000663 in male and female *B. germanica*.

Acknowledgements

We are grateful to the constructive comments provided by two anonymous reviewers, which improved the quality of this manuscript. Computational analyses were performed on the Maxwell cluster provided by the Research Computing Data Core at the University of Houston.

Authors' contributions

RPM and JRW conceived the project. RPM analyzed the genomic data. JRW collected the biological materials, extracted the DNA, and designed the 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.

Funding

This work was supported the National Science Foundation grant DEB-1845686 to RPM, by the National Institutes of Health grant R35GM122592 to Artyom Kopp and the National Institute of Food and Agriculture funds from the University of California to Artyom Kopp.

Availability of data and materials

The data analyzed in the current study are available from NCBI under the following accessions. Cockroach whole genome sequencing libraries are available in SRX693111, SRX693112, SRX693113, and SRX693110. The reference *B. germanica* genome assembly is available from JPZV00000000.1. RNA-seq data are available from accessions SRX3189901, SRX3189902, SRX2746607, SRX2746608, and SRX682022.

Ethics approval and consent to participate

No ethics approval or consent was required for this work.

Competing interests

The authors declare that they have no competing interests.

Author details

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

Received: 12 August 2019 Accepted: 11 November 2019

Published online: 05 December 2019

References

- Bull JJ. Evolution of sex determining mechanisms. Menlo Park: Benjamin/Cummings; 1983.
- Beukeboom L, Perrin N. The evolution of sex determination. New York: Oxford University Press; 2014.
- Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman T-L, Hahn MW, Kitano J, Mayrose I, Ming R, Perrin N, Ross L, Valenzuela N, Vamosi JC. Sex determination: why so many ways of doing it?. *PLoS Biol.* 2014;12(7):1001899. <https://doi.org/10.1371/journal.pbio.1001899>.
- Goodfellow PN, Lovell-Badge R. *SRY* and sex determination in mammals. *Ann Rev Genet.* 1993;27(1):71–92. <https://doi.org/10.1146/annurev.ge.27.120193.000443>.
- Baker BS, Belote JM. Sex determination and dosage compensation in *Drosophila melanogaster*. *Ann Rev Genet.* 1983;17:345–93.
- Cline TW. The *Drosophila* sex determination signal: how do flies count to two?. *Trends Genet.* 1993;9(11):385–90. [https://doi.org/10.1016/0168-9525\(93\)90138-8](https://doi.org/10.1016/0168-9525(93)90138-8).
- Erickson JW, Quintero JJ. Indirect effects of ploidy suggest X chromosome dose, not the X:A ratio, signals sex in *Drosophila*. *PLoS Biol.* 2007;5:332.
- Ohno S. Sex chromosomes and sex-linked genes: Springer; 1967.
- Mank JE. Sex chromosome dosage compensation: definitely not for everyone. *Trends Genet.* 2013;29(12):677–83. <https://doi.org/10.1016/j.tig.2013.07.005>.
- Chandler CH. When and why does sex chromosome dosage compensation evolve?. *Ann N Y Acad Sci.* 2017;1389(1):37–51. <https://doi.org/10.1111/nyas.13307>.
- Gu L, Walters JR. Evolution of sex chromosome dosage compensation in animals: a beautiful theory, undermined by facts and bedeviled by details. *Genome Biol Evol.* 2017;9(9):2461–76. <https://doi.org/10.1093/gbe/evx154>.
- van Doorn GS. Patterns and mechanisms of evolutionary transitions between genetic sex-determining systems. *Cold Spring Harb Perspect Biol.* 2014;6(8):017681. <https://doi.org/10.1101/cshperspect.a017681>.
- Kaiser VB, Bachtrog D. Evolution of sex chromosomes in insects. *Ann Rev Genet.* 2010;44:91–112. <https://doi.org/10.1146/annurev-genet-102209-163600>.
- Gempe T, Beyre M. Function and evolution of sex determination mechanisms, genes and pathways in insects. *Bioessays.* 2011;33(1):52–60. <https://doi.org/10.1002/bies.201000043>.
- Bopp D, Saccone G, Beyre M. Sex determination in insects: variations on a common theme. *Sex Dev.* 2014;8(1-3):20–8.
- Blackmon H, Ross L, Bachtrog D. Sex determination, sex chromosomes, and karyotype evolution in insects. *J Hered.* 2017;108(1):78–93. <https://doi.org/10.1093/jhered/esw047>.
- Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, Niehuis O, Petersen M, Izquierdo-Carrasco F, Wappler T, Rust J, Aberer AJ, Aspöck U, Aspöck H, Bartel D, Blanke A, Berger S, Böhm A, Buckley TR, Calcott B, Chen J, Friedrich F, Fukui M, Fujita M, Greve C, Grobe P, Gu S, Huang Y, Jermiin LS, Kawahara AY, Krogmann L, Kubiak M, Lanfear R, Letsch H, Li Y, Li Z, Li J, Lu H, Machida R, Mashimo Y, Kapli P, McKenna DD, Meng G, Nakagaki Y, Navarrete-Heredia JL, Ott M, Ou Y, Pass G, Podsiadlowski L, Pohl H, von Reumont BM, Schütte K, Sekiya K, Shimizu S, Slipinski A, Stamatakis A, Song W, Su X, Szucsich NU, Tan M, Tan X, Tang M, Tang J, Timelthaler G, Tomizuka S, Trautwein M, Tong X, Uchifune T, Walz MG, Wiegmann BM, Wilbrandt J, Wipfler B, Wong TKF, Wu Q, Wu G, Xie Y, Yang S, Yang Q, Yeates DK, 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 KM, Zhou X. Phylogenomics resolves the timing and pattern of insect evolution. *Science.* 2014;346(6210):763–7. <https://doi.org/10.1126/science.1257570>.
- Vicoso B, Bachtrog D. Reversal of an ancient sex chromosome to an autosome in *Drosophila*. *Nature.* 2013;499(7458):332–5.
- Sievert V, Kuhn S, Traut W. Expression of the sex determining cascade genes *Sex-lethal* and doublesex in the phorid fly *Megaselia scalaris*. *Genome.* 1997;40(2):211–4.
- 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.* 1998;125(8):1487–94. <http://arxiv.org/abs/http://dev.biologists.org/content/125/8/1487.full.pdf+html>.
- Saccone G, Peluso I, Artiaco D, Giordano E, Bopp D, Polito LC. The *Ceratitis capitata* homologue of the *Drosophila* sex-determining gene *sex-lethal* is structurally conserved, but not sex-specifically regulated. *Development.* 1998;125(8):1495–500.
- 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.* 2002;129(15):3715–25.
- Hall AB, Basu S, Jiang X, Qi Y, Timoshevskiy VA, Biedler JK, Sharakhova MV, Elahi R, Anderson MAE, Chen X-G, Sharakhov IV, Adelman ZN, Tu Z. A male-determining factor in the mosquito *Aedes aegypti*. *Science.* 2015;348(6240):1268–70. <https://doi.org/10.1126/science.aaa2850>.
- 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.* 2016;5:19281. <https://doi.org/10.7554/eLife.19281>.
- Krzywinski E, Dennison NJ, Lycett GJ, Krzywinski J. A maleness gene in the malaria mosquito *Anopheles gambiae*. *Science.* 2016;353(6294):67–9. <https://doi.org/10.1126/science.aaf5605>.
- Sharma A, Heinze SD, Wu Y, Kohlbrenner T, Morilla I, Brunner C, Wimmer EA, van de Zande L, Robinson MD, Beukeboom LW, Bopp D. Male sex in houseflies is determined by *Mdm*, a paralog of the generic splice factor gene *CWC22*. *Science.* 2017;356(6338):642–5. <https://doi.org/10.1126/science.aam5498>.
- Foster GG, Whitten MJ, Konovalov C, Arnold JTA, 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.* 1981;37:55–69. <https://doi.org/10.1017/S0016672300020012>.
- Baker RH, Wilkinson GS. Comparative genomic hybridization (CGH) reveals a neo-X chromosome and biased gene movement in stalk-eyed flies (genus *Teleopsis*). *PLoS Genet.* 2010;6(9):1001121.
- Kuroda MI, Hilfiker A, Lucchesi JC. Dosage compensation in *Drosophila*—a model for the coordinate regulation of transcription. *Genetics.* 2016;204(2):435–50. <https://doi.org/10.1534/genetics.115.185108>.
- Pane A, De Simone A, Saccone G, Polito C. Evolutionary conservation of *Ceratitis capitata transformer* gene function. *Genetics.* 2005;171(2):615–624. <https://doi.org/10.1534/genetics.105.041004>.
- Bopp D. About females and males: continuity and discontinuity in flies. *J Genet.* 2010;89(3):315–23.
- Scott MJ, Pimsler ML, Tarone AM. Sex determination mechanisms in the Calliphoridae (blow flies). *Sex Dev.* 2014;8(1-3):29–37. <https://doi.org/10.1159/000357132>.
- Meccariello A, Salvemini M, Primo P, Hall B, Koskinioti P, Dalíková M, Gravina A, Gucciardino MA, Forlenza F, Gregoriou M-E, Ippolito D, Monti SM, Petrella V, Perrotta MM, Schmeing S, Ruggiero A, Scolari F, Giordano E, Tsoumani KT, Marec F, Windbichler N, Arunkumar KP, Bourtzis K, Mathiopoulos KD, Ragoussis J, Vitagliano L, Tu Z, Papathanos

- PA, Robinson MD, Saccone G. *Maleness-on-the-Y (MoY)* orchestrates male sex determination in major agricultural fruit fly pests. *Science*. 2019;365(6460):1457–60. <https://doi.org/10.1126/science.aax1318>.
34. Riddle NC, Elgin SCR. The *Drosophila* dot chromosome: where genes flourish amidst repeats. *Genetics*. 2018;210(3):757–72. <https://doi.org/10.1534/genetics.118.301146>.
 35. Vicoso B, Charlesworth B. Evolution on the X chromosome: unusual patterns and processes. *Nat Rev Genet*. 2006;7:645–53.
 36. Gurbich TA, Bachtrog D. Gene content evolution on the X chromosome. *Curr Opin Genet Dev*. 2008;18:493–8.
 37. Wright AE, Dean R, Zimmer F, Mank JE. How to make a sex chromosome. *Nat Commun*. 2016;7:12087.
 38. Wiegmann BM, Trautwein MD, Winkler IS, Barr NB, Kim J-W, Lambkin C, Bertone MA, Cassel BK, Bayless KM, Heimberg AM, Wheeler BM, Peterson KJ, Pape T, Sinclair BJ, Skevington JH, Blagoderov V, Caravas J, Kutty SN, Schmidt-Ott U, Kampmeier GE, Thompson FC, Grimaldi DA, Beckenbach AT, Courtney GW, Friedrich M, Meier R, Yeates DK. Episodic radiations in the fly tree of life. *Proc Natl Acad Sci USA*. 2011;108(14):5690–5. <https://doi.org/10.1073/pnas.1012675108>.
 39. Muller HJ. Bearings of the 'Drosophila' work on systematics. In: Huxley J, editor. *The New Systematics*. Oxford: Clarendon Press; 1940. p. 185–268.
 40. Schaeffer SW, Bhutkar A, McAllister BF, Matsuda M, Matzkin LM, O'Grady PM, Rohde C, Valente VLS, Aguade M, Anderson WW, Edwards K, Garcia ACL, Goodman J, Hartigan J, Kataoka E, Lapoint RT, Lozovsky ER, Machado CA, Noor MAF, Papacait M, Reed LK, Richards S, Rieger TT, Russo SM, Sato H, Segarra C, Smith DR, Smith TF, Strelets V, Tobarí YN, Tomimura Y, Wasserman M, Watts T, Wilson R, Yoshida K, Markow TA, Gelbart WM, Kaufman TC. Polytene chromosomal maps of 11 *Drosophila* species: the order of genomic scaffolds inferred from genetic and physical maps. *Genetics*. 2008;179:1601–55. <https://doi.org/10.1534/genetics.107.086074>.
 41. 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*. 2012;32:2121–34.
 42. Linger RJ, Belikoff EJ, Scott MJ. Dosage compensation of X-linked Muller element F genes but not X-linked transgenes in the Australian sheep blowfly. *PLoS ONE*. 2015;10(10):0141544. <https://doi.org/10.1371/journal.pone.0141544>.
 43. Vicoso B, Bachtrog D. Numerous transitions of sex chromosomes in Diptera. *PLoS Biol*. 2015;13(4):1002078. <https://doi.org/10.1371/journal.pbio.1002078>.
 44. Davis RJ, Belikoff EJ, Scholl EH, Li F, Scott MJ. *no blokes* is essential for male viability and X chromosome gene expression in the Australian sheep blowfly. *Curr Biol*. 2018;28(12):1987–92. <https://doi.org/10.1016/j.cub.2018.05.005>.
 45. Ross MH, Cochran DG. Genetics of the German cockroach. *Comp Biochem Physiol A Comp Physiol*. 1989;94(4):551–4.
 46. Keil CB, Ross MH. C-banded meiotic karyotype of *Blattella germanica* from prophase I cells. *J Hered*. 1984;75(3):185–90. <https://doi.org/10.1093/oxfordjournals.jhered.a109909>.
 47. Harrison MC, Jongepier E, Robertson HM, Arning N, Bitård-Feidel T, Chao H, Childers CP, Dinh H, Doddapaneni H, Dugan S, Gowin J, Greiner C, Han Y, Hu H, Hughes DST, Huylmans A-K, Kemena C, Kremer LPM, Lee SL, Lopez-Ezquerria A, Mallet L, Monroy-Kuhn JM, Moser A, Murali SC, Muzny DM, Otani S, Piulachs M-D, Poelchau M, Qu J, Schaub F, Wada-Katsumata A, Worley KC, Xie Q, Ylla G, Poulsen M, Gibbs RA, Schal C, Richards S, Belles X, Korb J, Bornberg-Bauer E. Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nat Ecol Evol*. 2018;2:557–66. <https://doi.org/10.1038/s41559-017-0459-1>.
 48. Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *R Journal*. 2016;8(1):289–317. <https://doi.org/10.32614/RJ-2016-021>.
 49. Wiegmann BM, Trautwein MD, Kim J-W, Cassel BK, Bertone MA, Winterton SL, Yeates DK. Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. *BMC Biol*. 2009;7(1):34. <https://doi.org/10.1186/1741-7007-7-34>.
 50. Larsson J, Chen JD, Rasheva V, Rasmuson-Lestander AS, Pirrotta V. Painting of fourth, a chromosome-specific protein in *Drosophila*. *Proc Natl Acad Sci USA*. 2001;98:6273–8. <https://doi.org/10.1073/pnas.111581298>.
 51. Johansson A-M, Stenberg P, Bernhardtsson C, Larsson J. Painting of fourth and chromosome-wide regulation of the 4th chromosome in *Drosophila melanogaster*. *EMBO J*. 2007;26:2307–16.
 52. Johansson A-M, Stenberg P, Pettersson F, Larsson J. POF and HP1 bind expressed exons, suggesting a balancing mechanism for gene regulation. *PLoS Genet*. 2007;3:209.
 53. Seum C, Reo E, Peng H, Rauscher III FJ, Spierer P, Bontron S. *Drosophila* SETDB1 is required for chromosome 4 silencing. *PLoS Genet*. 2007;3(5):76. <https://doi.org/10.1371/journal.pgen.0030076>.
 54. Tzeng T-Y, Lee C-H, Chan L-W, Shen C-KJ. Epigenetic regulation of the *Drosophila* chromosome 4 by the histone H3K9 methyltransferase dSETDB1. *Proc Natl Acad Sci USA*. 2007;104(31):12691–6. <https://doi.org/10.1073/pnas.0705534104>.
 55. Brower-Toland B, Riddle NC, Jiang H, Huisinga KL, Elgin SCR. Multiple SET methyltransferases are required to maintain normal heterochromatin domains in the genome of *Drosophila melanogaster*. *Genetics*. 2009;181(4):1303–19. <https://doi.org/10.1534/genetics.108.100271>.
 56. Figueiredo MLA, Philip P, Stenberg P, Larsson J. HP1a recruitment to promoters is independent of H3K9 methylation in *Drosophila melanogaster*. *PLoS Genet*. 2012;8(11):1003061. <https://doi.org/10.1371/journal.pgen.1003061>.
 57. Lundberg LE, 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*. 2013;41(8):4481–94. <https://doi.org/10.1093/nar/gkt158>.
 58. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res*. 2017;45(D1):200–3. <https://doi.org/10.1093/nar/gkw1129>.
 59. Koch CM, 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*. 2009;5(9):1000644. <https://doi.org/10.1371/journal.pgen.1000644>.
 60. Herraiz A, Belles X, Piulachs M-D. Chorion formation in panoistic ovaries requires windei and trimethylation of histone 3 lysine 9. *Exp Cell Res*. 2014;320(1):46–53. <https://doi.org/10.1016/j.yexcr.2013.07.006>.
 61. Boyes JW, Van Brink JM. Chromosomes of calyptrate Diptera. *Can J Genet Cytol*. 1965;7(4):537–50. <https://doi.org/10.1139/g65-073>.
 62. White MJD. *Animal cytology and evolution*. London: Cambridge Univ Press; 1973.
 63. Lifshyzt E, Lindsley DL. The role of X-chromosome inactivation during spermatogenesis. *Proc Natl Acad Sci USA*. 1972;69:182–6. <https://doi.org/10.1073/pnas.69.1.182>.
 64. Bhutkar A, Russo SM, Smith TF, Gelbart WM. Genome-scale analysis of positionally relocated genes. *Genome Res*. 2007;17:1880–7. <https://doi.org/10.1101/gr.7062307>.
 65. Meisel RP, Han MV, Hahn MW. A complex suite of forces drives gene traffic from *Drosophila* X chromosomes. *Genome Biol Evol*. 2009;1:176–88.
 66. Ding S, Li X, Wang N, Cameron SL, Mao M, Wang Y, Xi Y, Yang D. The phylogeny and evolutionary timescale of Muscoidea (Diptera: Brachycera: Calypterae) inferred from mitochondrial genomes. *PLoS ONE*. 2015;10(7):0134170. <https://doi.org/10.1371/journal.pone.0134170>.
 67. Spradling AC, Mahowald AP. Amplification of genes for chorion proteins during oogenesis in *Drosophila melanogaster*. *Proc Natl Acad Sci USA*. 1980;77(2):1096–100. <https://doi.org/10.1073/pnas.77.2.1096>.
 68. Claycomb JM, Orr-Weaver TL. Developmental gene amplification: insights into DNA replication and gene expression. *Trends Genet*. 2005;21(3):149–62. <https://doi.org/10.1016/j.tig.2005.01.009>.
 69. Wexler J, Delaney EK, Belles X, Schal C, Wada-Katsumata A, Amicucci MJ, Kopp A. Hemimetabolous insects elucidate the origin of sexual development via alternative splicing. *eLife*. 2019;8:47490. <https://doi.org/10.7554/eLife.47490>.
 70. Montgomery SH, Mank JE. Inferring regulatory change from gene expression: the confounding effects of tissue scaling. *Mol. Ecol*. 2016;25(20):5114–28. <https://doi.org/10.1111/mec.13824>.
 71. Meisel RP, Gonzales CA, Luu H. The house fly Y Chromosome is young and minimally differentiated from its ancient X Chromosome partner. *Genome Res*. 2017;27(8):1417–26. <https://doi.org/10.1101/gr.215509.116>.
 72. Sturtevant AH, Novitski E. The homologies of the chromosome elements in the genus *Drosophila*. *Genetics*. 1941;26:517–41.
 73. Powell JR. *Progress and prospects in evolutionary biology: the Drosophila model*. New York: Oxford University Press; 1997.

74. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 2013;1303.3997v2. <https://arxiv.org/abs/1303.3997v2>.
75. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297–303. <https://doi.org/10.1101/gr.107524.110>.
76. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytzky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43(5):491–8.
77. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA. From fastq data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinformatics.* 2013;43:1–33. <https://doi.org/10.1002/0471250953.bi1110s43>.
78. Drinnenberg IA, deYoung D, Henikoff S, Malik HS. Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. *eLife.* 2014;3:03676. <https://doi.org/10.7554/eLife.03676>.
79. Robertson HM, Baits RL, Walden KKO, Wada-Katsumata A, Schal C. Enormous expansion of the chemosensory gene repertoire in the omnivorous german cockroach *Blattella germanica*. *J Exp Zool B.* 2018;330(5):265–78. <https://doi.org/10.1002/jez.b.22797>.
80. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotech.* 2016;34:525–7.
81. Love M, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550. <https://doi.org/10.1186/s13059-014-0550-8>.
82. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32:1792–7. <https://doi.org/10.1093/nar/gkh340>.
83. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 2017;45(D1):158–69. <https://doi.org/10.1093/nar/gkw1099>.
84. Gramates LS, Marygold SJ, Santos GD, Urbano J-M, Antonazzo G, Matthews BB, Rey AJ, Tabone CJ, Crosby MA, Emmert DB, Falls K, Goodman JL, Hu Y, Ponting L, Schroeder AJ, Strelets VB, Thurmond J, Zhou P, the FlyBase Consortium. FlyBase at 25: looking to the future. *Nucleic Acids Res.* 2017;45(D1):663–71. <https://doi.org/10.1093/nar/gkw1016>.
85. Tamura K, Subramanian S, Kumar S. Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Mol Biol Evol.* 2004;21:36–44.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

