Reversal of an ancient sex chromosome to an autosome in Drosophila
Beatriz Vicoso1 & Doris Bachtrog1

Although transitions of sex-determination mechanisms are frequent in species with homomorphic sex chromosomes1–3, heteromorphic sex chromosomes are thought to represent a terminal evolutionary stage owing to chromosome-specific adaptations such as dosage compensation or an accumulation of sex-specific mutations4–7. Here we show that an autosome of Drosophila, the dot chromosome, was ancestrally a differentiated X chromosome. We analyse the whole genome of true fruitflies (Tephritidae), flesh flies (Sarcophagidae) and soldier flies (Stratiomyidae) to show that genes located on the dot chromosome of Drosophila are X-linked in outgroup species, whereas Drosophila X-linked genes are autosomal. We date this chromosomal transition to early drosophilid evolution by sequencing the genome of other Drosophilidae. Our results reveal several puzzling aspects of Drosophila dot chromosome biology to be possible remnants of its former life as a sex chromosome, such as its minor feminizing role in sex determination4 or its targeting by a chromosome-specific regulatory mechanism13. We also show that patterns of biased gene expression of the dot chromosome during early embryogenesis, oogenesis and spermatogenesis resemble that of the current X chromosome. Thus, although sex chromosomes are not necessarily evolutionary end points and can revert back to an autosomal inheritance, the highly specialized genome architecture of this former X chromosome suggests that severe fitness costs must be overcome for such a turnover to occur.

Sex is an important and conserved feature, yet sex-determination mechanisms are labile in many taxa with non-differentiated, homomorphic sex chromosomes, and often vary among closely related species or among individuals within a species14. Highly differentiated, heteromorphic sex chromosomes (that is, a degenerate, gene-poor Y chromosome and an often dosage-compensated X), however, seem to represent an evolutionary end point and become a permanent fixture of the genome in many species groups1–3,4. Special adaptations on highly evolved sex chromosomes—such as dosage compensation or inactivation during male meiosis, an accumulation of sex-determining and sex-specific genes, or inviability of YY individuals—prevent the reversal of heteromorphic sex chromosomes back to autosomes1–4, and sex-chromosome differentiation is viewed as an evolutionary one-way street. Indeed, the heteromorphic sex chromosomes of both mammals and birds originated independently from an ancestor with homomorphic sex chromosomes over 100 million years ago, and the stable inheritance of the sex chromosomes in these two clades reflects their highly specialized genome architecture15. By contrast, homomorphic sex chromosomes in fish, amphibians and reptiles often show high rates of turnover between species16–18. Although observations in vertebrates support the notion that heteromorphic sex chromosomes present an ‘evolutionary trap’ inert to turnover19, little is known about such transitions in other taxa. Here, we uncover a sex-chromosome reversion within Diptera (flies) in the genetic model organism Drosophila.

In many higher Diptera (suborder Brachycera), including Drosophila, the basic karyotype (2n = 12) consists of 5 pairs of large euchromatic rods (named Muller elements A–E in Drosophila; each containing well over 2,000 genes) and a smaller heterochromatic dot chromosome8 (element F in Drosophila; containing only about 100 genes; see Fig. 1). The gene content of Muller elements is highly conserved across Diptera7. One of the large rods (element A) segregates as the X chromosome in Drosophila, and sex is determined by the dose of the X-linked gene Sex lethal (Sxl), with diploid XX embryos developing into females and haploid XY into males10. The identity of the sex chromosomes and the master sex-determination gene is not conserved across other Diptera families11; instead, a small and often heterochromatic chromosome pair segregates as the sex chromosomes in several other Brachycera12. No genes have yet been isolated from the sex chromosomes in outgroup species, and the relationship with Drosophila chromosomes, if any, is unclear.

We used whole-genome sequencing to identify sex-linked genes in species from several higher Diptera families, including the black soldier fly Hermetia illucens (Stratiomyidae, a basal Brachycera), the olive fruitfly Bactrocera oleae (Tephritidae), the grey fleshly Sarcophaga bullata (Sarcophagidae), the zoophilic fruitfly Phortica variegata (Stegagininae, a sister clade to Drosophila within Drosophilidae), as well as the basal Drosophila species Drosophila buscki (Fig. 1 and Supplementary Fig. 1). Paired-end Illumina genomic reads were obtained from single males and females of each species, assembled, and the resulting scaffolds were mapped to the different Muller elements according to

---

1Department of Integrative Biology, Center for Theoretical Evolutionary Genomics, University of California Berkeley, Berkeley, California 94720, USA.

---

Figure 1 | Sex chromosomes in higher Diptera revealed by genome analysis. Evolutionary relationship inferred from 185 conserved protein-coding genes (93,134 amino acids) using PhyML (with bootstrap values indicated at the nodes), and male-to-female coverage ratio across chromosome elements (Muller elements A–F) in the Dipterans studied. X chromosomes (red) have only half the read coverage in males versus females. Boxes extend from the first to the third quartile and whiskers to the most extreme data point within 1.5 times the interquartile range.

DOI: 10.1038/nature12235

©2013 Macmillan Publishers Limited. All rights reserved
their homology to *Drosophila melanogaster* genes (Supplementary Table 1). Male and female reads were mapped back to the scaffolds, and male-to-female coverage was used to detect sex-linked sequences, with X-linked scaffolds only showing half the coverage in males compared to females. Consistent with Muller–A being the ancestral sex chromosome in all Drosophilidae, male-to-female coverage is only half that relative to autosomes in *D. busckii* and *P. variegata* (Fig. 1, Supplementary Table 2 and Supplementary Figs 2–4). In the more distantly related outgroups *B. oleae*, *S. bullata* and *H. illucens*, male-to-female coverage shows that element A is an autosome; instead, element F, the ‘dot chromosome’ of *Drosophila*, has significantly reduced male-to-female coverage in outgroup species (Fig. 1, Supplementary Table 2 and Supplementary Figs 5–7). This indicates that element F is a heteromorphic sex-chromosome pair in several Brachycera, with a completely degenerate Y. In fact, the phylogenetic pattern suggests that the dot was a sex chromosome in an ancestor of *Drosophila*, and only reverted to an autosome in the lineage leading to Drosophilidae.

Certain features of their genome architecture distinguish X chromosomes from autosomes: (1) they contain sex-determining factors (*Sex-complex element* in *Drosophila*)<sup>10</sup>; (2) they show a non-random gene content (a deficiency of male- and excess of female-biased genes in *Drosophila*)<sup>11</sup>; (3) they are transcriptionally inactivated during spermatogenesis<sup>12</sup>; (4) chromosome haploidy is not lethal; and (5) chromosome-wide regulatory mechanisms ensure balanced levels of gene expression in the heterogametic sex (the male-specific lethal (MSL) complex in *Drosophila*). Interestingly, the dot chromosome of *D. melanogaster* harbours several peculiar characteristics that have puzzled drosophilists for decades, but can now be understood as possible remnants of its history as an X chromosome. Specifically, the dot has a minor role in sex determination and contains feminizing factors<sup>13</sup>, with increased dosage of the dot chromosome shifting 2X:3A (A denotes autosome) intersex individuals towards female development<sup>14</sup>. Consistent with its feminizing effect, genes located on the dot chromosome generally show higher expression in female compared to male embryos during early development, similar to X-linked genes (Fig. 2a). There is also a deficit of testes and an excess of ovary expression on the dot, resembling sex-biased expression profiles of the current X chromosome (Fig. 2b). In addition, genes on the dot chromosome are downregulated during male meiosis (Fig. 2c), mimicking the phenomenon of male germ-line X inactivation. Finally, flies with only one copy of the dot chromosome are viable and fertile, and the dot is targeted by the chromosome-specific protein POF, which is involved in transcriptional regulation of genes located on the dot chromosome<sup>6</sup>. This resembles the MSL complex, the only other known protein complex that specifically targets a chromosome, and POF may be part of a putative ancestral mechanisms of dosage compensation<sup>17,18</sup> (Fig. 2d). In some *Drosophila* species, POF shows male-specific binding to the X chromosome in addition to binding to the dot, further supporting its involvement in dosage compensation<sup>17</sup>. Some components of the MSL complex are necessary for normal expression of genes located on the dot<sup>15</sup>, consistent with interactions of the regulatory network for the current and former sex chromosome of *Drosophila* (Fig. 2e). Thus, many features of the dot chromosome in *Drosophila* resemble unique characteristics of the current X chromosome that distinguish it from autosomes, and can be interpreted as signatures of its former life as a differentiated X chromosome. Female-biased expression during early embryogenesis, an excess of ovary expression and a deficiency of testis expression on Muller F (but not Muller A) are all observed in Diptera species in which the dot segregates as the X (Fig. 3). This confirms that these peculiarities of the dot were present in the X-linked ancestor, and also shows that Muller A only acquired them once it became sex-linked in the lineage leading to *Drosophila*.

It is of interest to consider how this transition could have happened. Sex in most outgroup species of *Drosophila* is determined by the presence of a factor on the Y chromosome (M factor) causing maleness<sup>20</sup>. Such a dominant-Y system with element F as the sex chromosome could evolve into the *Drosophila* system (dose-dependent sex determination with element A as the sex chromosome) through various intermediate steps<sup>1</sup>, and we outline three possible paths that involve mutational events that have been observed in Diptera species. This sex-chromosome transition could be initiated by a single epistatic mutation on element A (M*) that makes individuals male regardless of their sex-chromosome karyotype, or through a translocation of the existing male-determining factor (M) onto element A (Fig. 4). Novel sex-determining genes occurring on different chromosomes or translocations of M factors onto autosomes have been observed in houseflies<sup>21</sup> or humpbacked

**Figure 2 | Properties of the dot chromosome in *Drosophila melanogaster* that resemble that of an X chromosome.** Panel a. Zygotic transcription, before the onset of MSL-dosage compensation, is female-biased for genes located on the dot (data from ref. 28). b. Genes located on the dot chromosome show increased transcription (in fragments per kilobase of transcript per million mapped reads, FPKM) in the ovary and decreased transcription in testis (data from ref. 29). c. An excess of genes located on the dot is downregulated during male meiosis (data from ref. 30). d. Chromosome-specific transcriptional regulation of genes located on the dot chromosome by the protein POF. e. Misregulation of genes located on the dot chromosome in mutants for some components of the MSL dosage compensation complex (roX1/roX2 mutant in grey, MSL-2 mutant in green; data from ref. 19). Levels of significance (based on resampling for panels b and e and using Chi-squared tests for panels a and c) are represented by *P < 0.05 and **P < 0.01. Boxes (panels b and e) as in Fig. 1. WT, wild type.
flies. The fixation of the new male-determining gene on element A would lead to reversal of the ancestral X (Muller F) to an autosomal, and the ancestral Y would be completely lost. Loss of the Y is only possible if the ancestral Y chromosome carried no essential male-fertility genes, as seems to be the case in several Diptera species, or if those male-fertility genes moved to another chromosome. The emergence of a male-determining gene on element A would cause male-limited transmission of this chromosome and—as higher Diptera males generally lack recombination—set in motion genome-wide degeneration of the non-recombining proto-Y. Eventually, Sxl was recruited as a dose-dependent sex-determination gene in Drosophila, and MSL-mediated dosage compensation evolved on element A. The current karyotype of Drosophila could also have evolved through a chromosomal fusion between the ancestral Y chromosome and element A (Fig. 4). This would create a male-limited neo-Y (the fused element A) that initially is identical to the neo-X (the unfused element A). The non-recombining neo-Y would undergo chromosome-wide degeneration, and Y-autosome fusions coupled with neo-Y degeneration have happened repeatedly in several Drosophila species and other Diptera. This fused neo-Y chromosome would form the current Y of Drosophila, and male-fertility genes, if present on the ancestral Y, could be preserved on the chimaeric Y. Both element F and element A would simultaneously segregate as X chromosomes, and eventually Sxl would take over the sex-determining function on element A, and a non-disjunction event could restore diplodiploidy for element F in both sexes. Thus, under both the translocation of an existing or the emergence of a new M-factor scenario, the current Y chromosome of Drosophila shares no homology to the Y of its ancestor, whereas under the chromosomal-fusion model, the current Drosophila Y is a chimera of the ancestral Y and the degenerated element A and may thus harbour some genes that are also Y-linked in outgroup species. We detect no Y-linked protein-coding genes that are shared between Drosophila and Diptera where Muller F is the sex chromosome (Supplementary Tables 3–5 and Supplementary Figs 8–10), providing some support against a fusion between the new and ancestral Y. However, it should be noted that the gene content of the Y is generally poorly conserved, even within Drosophila, and the high turnover of Y-linked genes may obscure any homology between the Y chromosome of drosophilids and that of their outgroups.

Cytological and genetic studies have shown that the dot chromosome of D. busckii is fused to both the X and Y chromosome. Interestingly, this basal species in the Drosophila genus also seems to lack MSL-mediated dosage compensation; instead, POF binds to the fused elements A and F in males only, and it had been suggested that the A–F fusion represents the ancestral state of Drosophila. Our data, however, clearly indicate that D. busckii harbours a recent, secondary fusion: P. variegata, an outgroup species, only has element A as its sex chromosome, and the reduction in male-to-female coverage for element F in D. busckii is much less pronounced than for element A. This suggests that reads from the Y-fused homologue of element F are still mated by aligning reads back to scaffolds using the Burrows-Wheeler Aligner (BWA) and running SOAPCoverage on the alignments. Mapping of D. melanogaster genes to genomic scaffolds using the BLAST-like alignment tool (BLAT) with a translated query and database allowed assignment of scaffolds to Muller elements. BLAT queries uncovered genes conserved in all species, and their corresponding

Figure 3 | Gene expression in early embryos and adult gonads in outgroup Diptera species. a. Zygotic transcription is female-biased for genes located on the X chromosome (element F) of B. oleae, but not the autosomal element A. FPKM values were estimated by RNA-sequencing of male and female stage 5 embryos. b. Overexpression in ovary and underexpression in testis on the X chromosome (element F) but not the autosomal element A in B. oleae (top) and S. bullata (bottom). FPKM values were estimated through RNA-sequencing of dissected testis and ovaries. Levels of significance (based on Chi-square tests for panel a and using one-tailed Wilcoxon tests for panel b) are represented by *P < 0.05, **P < 0.01 and ***P < 0.001. Boxes (panel b) as in Fig. 1.

Figure 4 | Turnover of sex chromosomes in Drosophila. Hypothetical transition from an ancestral karyotype with element F segregating as a sex chromosome, to the karyotype observed in Drosophila in which element A is the sex chromosome. In outgroup species, maleness is determined by the presence of a factor on the Y chromosome (M factor). Element A could either have acquired a new, epistatic M* factor (scenario 1), or the existing M factor could have transposed to element A (scenario 2). This transition could also have been initiated by a fusion of the ancestral Y to element A (scenario 3). Degeneration of the male-limited, non-recombining element A, followed by recruitment of Sxl for sex determination, would create the ancestral karyotype of Drosophila.
protein sequence was extracted using GeneWise and aligned using MUSCLE. After removal of low-quality alignments by Gblocks, the resulting alignments were run through PhyML to obtain the phylogenetic tree and corresponding bootstrap values. Transcriptome data for *D. melanogaster* were obtained from the following sources: male and female early embryo transcriptome data (female-to-male slope expression values for zygotic genes) are taken from supplementary table 1 of ref. 28, ovary and testis expression data were obtained from ref. 29, expression profiles during spermatogenesis are taken from supplementary table 1 of ref. 30, and expression data for MSL mutants and wild type are taken from ref. 19 and were downloaded from the NCBI Gene Expression Omnibus (GEO) depository (GSE3990 and GSE12054). The polytene chromosome image (immunostained for POF and MSL) is a courtesy by J. Larsson. Gene expression in *B. oleae* from early male and female embryos was assayed through RNA-sequencing of stage 5 embryos, following ref. 28, and gene expression profiles for dissected testes and ovaries were obtained from *B. oleae* and *S. bullata*, through RNA-sequencing.

**Full Methods** and any associated references are available in the online version of the paper.

Received 14 June 2012; accepted 30 April 2013.

Published online 23 June 2013.

1. Bull, J. J. Evolution of Sex Determining Mechanisms (The Benjamin/Cummings Publishing Company, 1983).
2. Sarre, S. D., Ezaz, T. & Georges, A. Transitions between sex-determining systems in reptiles and amphibians, *Ann. Rev. Genomics Hum. Genet.* **12**, 391–406 (2011).
3. Pokorná, M. & Kratovil, L. Phylogeny of sex-determining mechanisms in squamate reptiles: are sex chromosomes an evolutionary trap? *Zool. J. Linn. Soc.* **156**, 168–183 (2009).
4. Ohta, S. Sex Chromosomes and Sex-Linked Genes (Springer Verlag, 1967).
5. Bridges, C. B. Sex in relation to chromosomes and genes. *Am. Nat.* **59**, 127–137 (1925).
6. Larsson, J., Chen, J. D., Rasheva, V., Resmussen-LeStander, A. & Pirrotta, V. Painting of fourth, a chromosome-specific protein in *Drosophila*. *Proc. Natl Acad. Sci. USA* **98**, 6273–6278 (2001).
7. Bellott, D. W. & Page, D. C. Reconstructing the evolution of vertebrate sex chromosomes. *Cold Spring Harb. Symp. Quant. Biol.* **72**, 168–183 (2007).
8. Muller, J. H. in The New Systematics (ed. Huxley, J.) 185–268 (Clarendon Press, 1940).
9. Holt, R. A. et al. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298**, 129–149 (2002).
10. Cline, T. W. *The Drosophila sex determination signal: how do flies count to two?* Trends Genet. **9**, 385–390 (1993).
11. Zacharopolou, A. Cytogenetic analysis of mitotic and salivary gland chromosomes in the Medfly *Ceratitis capitata*. *Genome* **29**, 67–71 (1987).
12. White, M. J. Cytological evidence on the phylogeny and classification of the Diptera. *Evolution* **3**, 252–261 (1949).
13. Vicoso, B. & Bachtrog, D. Lack of global dosage compensation in *Schistosoma mansoni*, a female-heterogametic parasite. *Genome Biol. Evol.* **3**, 230–235 (2011).
14. Paris, M. et al. Faculty of genes on the *Drosophila X chromosome* showing male-biased expression. *Science* **299**, 697–700 (2003).
15. Lifschytz, E. & Lindsley, D. L. The role of *X*-chromosome inactivation during spermatogenesis. *Proc. Natl Acad. Sci. USA* **69**, 182–186 (1972).
METHODS
Sample collection. *D. busckii* individuals from an inbred isofemale line originally caught in Tallinn, Estonia in the year 2000 were provided by J. Larsson. Wild-caught male and female *P. variegata* were collected in Bari, Italy by S. Kunhi Purayil and I. Kadow, shipped in ethanol and kept at −20 °C until DNA extraction. Papue of *S. bullata* were obtained from Carolina Biological and incubated at 20°C until adults emerged. These were immediately sexed and used for DNA extraction. *B. oleae* males and females were provided by P. Mavragani from a stock maintained at the Aristotle University of Thessaloniki, shipped in ethanol and kept at −20 °C until DNA extraction. *H. illucens* were provided by U. Schmidt-Ott, shipped in ethanol and kept at −20 °C until DNA extraction.

Library preparation and sequencing. DNA was extracted from a single male and a single female individual from each species using the Puregene Core Kit A (Qiagen). Library preparation and paired-end sequencing of *B. oleae*, *P. variegata*, *S. bullata* and *H. illucens* were performed at the Beijing Genomics Institute (http://en.genomics.cn/). The *D. busckii* libraries were prepared in the Backtroh laboratory following the standard Illumina protocol, and sequenced at the Berkeley sequencing facility. For each species, libraries and sequencing were performed for a single male and a single female separately. The raw sequenced data consists of 90-base pair (bp) paired-end reads, with an insert size of 500 bp for all libraries. All the reads generated for this analysis have been deposited at the NCBI Short Reads Archive under bioproject accession SRP021047.

Genome assembly. We obtained 90-bp paired-end reads for one male and one female of each species, of which the first 10 bp were trimmed. For each species, the male and female genomic reads were assembled using SOAPdenovo (http://soap-genomics.org.cn) with default parameters and a K-mer value of 31. Gaps in the assembly were reduced further using GapCloser (http://soap.genomics.org.cn/). Only scaffolds longer than 1,000 bp were kept for further analysis. Statistics for the assemblies are presented in Supplementary Table 1. In the case of *D. busckii*, pooling male and female reads for the assembly led to reduced coverage of element F in both male and females (see below and Supplementary Fig. 2). This is probably due to the hybrid assembly of the neo-X and neo-Y (element F fused to the X, and element F fused to the Y, respectively), as the neo-Y is not yet fully degenerated and some neo-Y-derived reads are included in the assembly. Because this hybrid assembly differs from the true sequence of the neo-X, female reads will not fully map to it, leading to decreased female coverage. This bias disappeared when only female reads were used for the assembly (Supplementary Fig. 3). This female-only assembly was therefore used for all further analyses. It should be noted that this does not affect our conclusions, as the male-to-female coverage ratio is reduced for elements A and F independently of the assembly used in the analysis.

Mapping of the genomic scaffolds to Muller elements. We downloaded all *D. melanogaster* coding sequences from Flybase (http://flybase.org/) and, for each gene, kept only the longest coding sequence. The resulting *D. melanogaster* coding sequences were mapped to the genome assembly of each outgroup using the BLAST-like alignment tool (BLAT)39 with both a translated query and a translated database, and only the hit with the highest match score was kept for each gene. Hits with a match score below 50 were excluded. A Perl script was used to count the number of *Drosophila* genes per Muller element that mapped to each scaffold. The scaffold was assigned to the Muller element with the largest number of matching genes. When the same number of genes from different Muller elements mapped to a scaffold, a score was calculated for each Muller element by summing the match scores of all the hits for that element, and the scaffold was assigned to the element with the largest score. The rate of concordance of genes on scaffolds (the number of genes that map to the element their scaffold was assigned to versus total number of genes, calculated using scaffolds carrying at least three genes) is high: *D. busckii*, 96%; *P. variegata*, 92%; *S. bullata*, 80%; *B. oleae*, 88% and *H. illucens*, 71%.

Estimation of male and female genomic coverage. Male and female reads were mapped separately to the genomic scaffolds using the Burrows-Wheeler aligner (BWA)40 set to the default parameters for paired-end reads. The resulting sequence alignment/map (SAM) alignments were used to estimate the male and female coverage depth for each scaffold using SoapCoverage (http://soap.genomics.org.cn/). Supplementary Figs 2–7 show the male and female coverage distributions for each Muller element for all the species investigated. For each species, we used one-tailed Wilcoxon tests between each element and the rest of the sample to identify systematic reductions in male-to-female coverage. The results of the tests are presented in Supplementary Table 2.

Building the phylogenetic tree. From the BLAT results (*D. melanogaster* genes versus outgroup genomic sequence, see the section Mapping of the genomic scaffolds to Muller elements), we mapped back to the *D. melanogaster* genes that had an alignment score above 1,000 for all of the outgroups species. The genomic regions corresponding to this conserved set of genes were extracted from the genomic scaffolds using a Perl script.

GeneWise was used to infer the protein sequence of the genes in the different outgroups from the corresponding genomic region. To avoid potential biases that could arise by using the *D. melanogaster* protein sequences as input for GeneWise (as GeneWise may be more likely to insert errors in the more distant outgroups than in the closer ones), we used *Anopheles gambiense* protein sequences instead, as *A. gambiense* is an outgroup to all the species analysed. For each gene, the inferred protein sequences and the *D. melanogaster* and *A. gambiense* protein sequences were aligned using Muscle. We concatenated the Muscle output files, and ran the resulting concatenated alignment through Gblocks to remove gaps and regions of low alignment quality. The final alignment, consisting of 93,134 amino acids, was used as input for PhyML to obtain the phylogenetic tree and associated bootstrap values (Supplementary Fig. 1).

Resampling procedure. Differences between elements A and F versus autosomes (elements B–E) were tested by resampling n genes 1,000 times from the autosomal sample (where n is the number of genes for element A or F present in the sample). Levels of expression of elements A or F were considered to be significantly different from the autosomes if their observed median fell within the 5% one-sided tail of the resampled distribution.

Expression analysis of adult tissues. *B. oleae* individuals were obtained from olives collected on the University of California, Berkeley campus, and *S. bullata* adults were obtained from pupe purchased from Carolina Biological. Male and female adults were placed in separate vials immediately after emergence, and aged on standard *Drosophila* food for 4 days. Paired-end RNA-sequencing (RNA-seq) reads were obtained from testis and ovary. RNA was extracted from each tissue using a Qiagen RNeasy kit following the manufacturer’s protocol. RNA-seq libraries and sequencing were performed at the Beijing Genomics Institute. All the reads generated for this analysis have been deposited at the NCBI Short Reads Archive under bioprojects SRP021043 and SRP021044. For each species, the resulting paired-end 90-bp reads were trimmed, pooled and assembled using SOAPdenovo (http://soap.genomics.org.cn/soapdenovo.html) with a K-mer value of 31 and an insert size of 200 bp, and this assembly was further improved using GapCloser (http://soap.genomics.org.cn/soapdenovo.html). Only scaffolds longer than 300 bp were kept for further analysis. Scaffolds were mapped to *D. melanogaster* coding sequences downloaded from Flybase (http://www.flybase.org) using BLAT; only the location with the best mapping score was kept for each scaffold. When more than one scaffold overlapped on the same *D. melanogaster* gene by more than 20 bp, only the scaffold with the highest mapping score was kept for further analysis. When several scaffolds mapped to different parts of the same gene, or had an overlap shorter than 20 bp, their sequences were concatenated. The resulting *B. oleae* and *S. bullata* gene sequences were assigned to Muller elements according to the genomic location of their *D. melanogaster* homologues. For each adult tissue, reads were mapped to the *B. oleae* or *S. bullata* gene sequences using Bowtie2 (ref. 33) with default parameters. FPKM values were then estimated from the resulting SAM alignments using Cufflinks with default parameters. Genes with FPKM < 10 in both testis and ovary were excluded.

Expression analysis of early embryos. In order to assess early embryonic expression, adult individuals of *B. oleae* were placed in cages with fresh olives. Eggs were periodically collected from the olives and their chorions removed by immersing them in 50% bleach solution for 5 min. The de-chorionated embryos were staged by observation under a light microscope after immersion in halocarbon oil. Stage 5 embryos were selected, as in *D. melanogaster* they represent the earliest stage containing primarily zygote-derived RNAs (earlier stages contain mostly maternal RNAs48). RNA and DNA were extracted following the protocol described in ref. 28, and the DNA was used to sex the embryos using published sexing primers44. Libraries were made at the Backtroh laboratory following the protocol of ref. 28 and 100-bp single-end sequencing was performed at the Vincent J. Coates Genomics Sequencing Laboratory, Berkeley. The reads generated for this analysis have been deposited at the NCBI Short Reads Archive under bioproject SRP021044. The first 20 bp of the resulting reads were trimmed before further analysis. Reads were mapped to the *B. oleae* or *S. bullata* gene sequences (obtained in Expression analysis of adult tissues) using Bowtie2 (ref. 33) with default parameters. FPKM values were then estimated from the resulting SAM alignments using Cufflinks with default parameters. Genes with FPKM < 1 in both male and female embryos were excluded.

Searching for *Drosophila* Y-linked genes in *Drosophila* outgroups. The male DNA-seq reads available for *P. variegata*, *S. bullata*, *B. oleae* and *H. illucens* were assembled using SOAPdenovo (with a K-mer value of 31), and all scaffolds longer than 200 bp were kept for further analysis. One lane of forward female and male libraries were sequenced using Illumina HiSeq technology (2 × 100 bp paired-end reads). The reads were mapped back to these male scaffolds separately using BWA (with default parameters) and male and female coverage was estimated using SoapCoverage. We used single-end mapping of the forward read rather than paired-end mapping because the insert size of the paired-end fragments (500 bp) is longer than the minimum size of the fragments. Similarly, RNA-seq reads obtained
from male tissues were assembled using SOAPdenovo-Trans for *S. bullata* (whole male, male carcass, testis) and *B. oleae* (whole male, male carcass, testis), and all transcripts longer than 200 bp were kept for further analysis. One lane of forward female and male DNA-seq reads were mapped back separately to the resulting male transcripts using BWA (with default parameters) and male and female coverage was estimated using SoapCoverage.

The protein sequences of known ancestral Y-linked genes of *Drosophila* (kl-2, kl-3, Opr, PRY, Ppr-Y, but also Cy and ARY, which may be ancestral) were obtained from the NCBI (with identifier (GI) numbers 190608814, 219131049, 16519041, 217416310, 158529626, 190608812 and 281309229, respectively). Genomic scaffolds and transcripts containing homologues to these proteins were identified using tBLASTn (with an E value cutoff of 0.01). Sequences identified in our BLAST search that had no coverage in females, but had coverage higher than 0 in males, were classified as candidate homologues to the *Drosophila* Y-linked genes (Supplementary Tables 3 and 4). No such sequences were found in the male transcriptome or genome of *S. bullata*, *B. oleae* and *H. illicens*, the outgroups carrying the ancestral sex chromosome (Mueller element F). *P. variegata*, on the other hand, presents strong evidence of sharing both Cy and kl-2 with other drosophilids. Primers designed to amplify fragments of the candidate Y-linked *P. variegata* scaffolds 13109 and 15584, which share homology with kl-2 and Cy, respectively, amplified bands of the expected size in male but not female *P. variegata*, confirming their Y linkage (Supplementary Fig. 8).

In Diptera, there is a general tendency for the ribosomal DNA to reside on the sex chromosomes, but its location often differs between closely related species35. For example, both the X and the Y of *D. melanogaster* carry rDNA genes36, and the Y chromosome of *D. pseudoobscura*, which is non-homologous to the Y of *D. melanogaster*, has acquired rDNA loci after its formation37. On the other hand, rDNA loci in the *Drosophila* ananassae subgroup are located on autosomes, and *D. simulans*, the sister species of *D. melanogaster*, has lost its Y-linked rDNA cluster38. Thus, the absence or presence of rDNA on sex chromosomes contains limited information on whether the Y chromosomes of Diptera are homologous across families, and we therefore focused our analysis on protein-coding genes. Cytological studies have found that some (but not all) true fruitflies have rDNA loci on their Y (including *B. oleae*36), whereas house flies or flesh flies have rDNA clusters on their autosomes35-38.

**Candidate Y-linked transcripts of *S. bullata* and *B. oleae* and their *D. melanogaster* homologues.** The trimmed whole-body male and female RNA-seq reads of *S. bullata* and *B. oleae*, as well as the male and female DNA-seq reads, were mapped back to the assembled male transcriptome of these species (see section Searching for *Drosophila* Y-linked genes in *Drosophila* outgroups above) using Bowtie2, keeping only reads that mapped fully in the alignment. After filtering out reads with mismatches, the DNA-seq and RNA-seq alignments were used as input for SoapCoverage and Cufflinks, respectively, to estimate the male and female DNA coverage and expression level (in FPKM). As Y-derived sequences are characterized by their lack of female coverage or expression, scaffolds that had at least 6 male reads mapping, had (female reads/male reads) < 0.1 and had (female expression/male expression) < 0.1 were classified as candidate Y-linked transcripts (39 transcripts in *B. oleae* and 87 in *S. bullata*). PCRs with primers designed for 10 candidate transcripts of *S. bullata* yielded male-specific bands for 2 transcripts (standard PCR with an annealing temperature of 58 °C, see Supplementary Fig. 9), confirming that the sample contains Y-derived sequences.

In order to check for Y-linkage of these candidate sequences in *D. melanogaster*, DNA was extracted (using a Qiagen DNeasy kit) from one male and one female *D. melanogaster* and DNA-seq libraries were made and sequenced at Beijing Genomics Institute (paired-end with 500-bp insert size, 2 GB of data per individual). We assembled the resulting trimmed genomic reads with SOAPdenovo (with a K-mer value of 31). The male and female coverage of each genomic scaffold was estimated by mapping the male and female DNA-seq reads against the assembled genome with Bowtie2 (keeping only fully mapped reads, and filtering out reads with mismatches) and processing the SAM alignment with SoapCoverage. The candidate Y-linked transcripts of *S. bullata* and *B. oleae* were then mapped against the *de novo* *D. melanogaster* genome assembly using tBLASTn (with an E value cutoff of 0.01, see Supplementary Table 5). As a control, 12 known Y-linked genes of *D. melanogaster* were also mapped to this assembly. All 12 *D. melanogaster* Y-linked genes mapped to scaffolds with male coverage >0 but no female coverage, confirming that most Y-derived coding sequence is represented in the *de novo* *D. melanogaster* assembly, and can be identified by its patterns of male and female coverage. Contrary to this, only 1 out of 39 of the *B. oleae* and none of the 87 *S. bullata* candidate Y-derived sequences mapped to *D. melanogaster* scaffolds with male-specific coverage, supporting a general lack of homology between the Y chromosome of drosophilids and that of their outgroups (Supplementary Table 5). Finally, the single *B. oleae* candidate that maps to a male-specific *D. melanogaster* scaffold was shown to yield bands of similar sizes in both sexes when a PCR was performed with primers designed for the scaffold in both *B. oleae* and *D. melanogaster*, showing that this fragment is not Y-linked in either species (Supplementary Fig. 10).

31. Kent, W. J. BLAT – the BLAST-like alignment tool. Genome Res. 12, 656–664 (2002).
32. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 26, 589–595 (2010).
33. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25 (2009).
34. Gabrieli, P. et al. Interchromosomal duplications on the Bactrocera oleae Y chromosome imply a distinct evolutionary origin of the sex chromosomes compared to Drosophila. PloS ONE 6, e17747 (2011).
35. Roy, V. et al. Evolution of the chromosomal location of rDNA genes in two *Drosophila* species subgroups: ananassae and melanogaster. Heredity 94, 388–395 (2005).
36. Drosopoulou, E. et al. Sex chromosomes and associated rDNA form a heterochromatic network in the polytene nuclei of Bactrocera oleae (Diptera: Tephritidae). Genetics 140, 169–180 (2012).
37. Parise-Maltempi, P. P. & Avancini, R. M. P. Cytogenetic study in Musciidae flies. Braz. J. Biol. 67, 945–950 (2007).