Genomic origins of insect sex chromosomes
Heath Blackmon and Jeffery P Demuth

Recent efforts to catalog the diversity of sex chromosome systems coupled with genome sequencing projects are adding a new level of resolution to our understanding of insect sex chromosome origins. Y-chromosome degeneration makes sequencing difficult and may erase homology so rapidly that their origins will often remain enigmatic. X-chromosome origins are better understood, but thus far prove to be remarkably labile, often lacking homology even among close relatives. Furthermore, evidence now suggests that differentiated X or Y-chromosomes may both revert to autosomal inheritance. Data for ZW systems is scarcer, but W and Y-chromosomes seem to share many characteristics. Limited evidence suggests that Z-chromosome homology is more conserved than X counterparts, but broader sampling of both sex chromosome systems is needed.

Addresses
Department of Biology, University of Texas at Arlington, Box 19498, Arlington, TX 76019, USA

Corresponding author: Demuth, Jeffery P (jodemuth@uta.edu)

Insects have evolved myriad ways of producing males and females, but chromosomal sex determination is the most common (Box 1). Sex chromosomes are thought to originate from a pair of ancestral autosomes that evolve a sex determining locus [1]. The sex-limited allele defines a proto Y or W chromosome in male or female heterogametic respectively. As sexually antagonistic mutations (i.e. those that benefit one sex at the expense of the other) arise, selection will favor decreased recombination between the male or female sex determining allele and the matching male or female beneficial allele. Once recombination becomes suppressed, the sex-limited region of the proto Y or W chromosome is subject to a variety of population genetic forces that are expected to result in mutational decay and gene loss [2]. The phylogenetically widespread observation of XO (ZO) species [3] indicates that decay of the Y(W) often results in its complete loss [4**]. However, decay is not inevitable.

Some lineages retain homomorphic sex chromosomes over long evolutionary time scales (e.g. Isoptera). Additionally, Y and W chromosomes often reemerge in clades that are ancestrally XO or ZO (e.g. Lepidoptera), a strong indication - even in the absence of sequence data - that not all sex-limited chromosomes are homologous [5*,6]. Following we review the genomic origins and turnover of differentiated sex chromosome in insects with particular emphasis on recent genome sequencing and comparative phylogenetic efforts (for more comprehensive reviews of sex chromosomes and their evolution see [7,8]).

Y-chromosome evolution
Male heterogamety (XY, XO) is the most abundant form of sex determination in insects (Box 1), having been documented in 24 of 28 orders encompassing 77% of sexually reproducing species investigated. However, the sex specific portions of Y-chromosomes are some of the least well-understood regions of all genomes. Since they do not recombine, they cannot be studied with traditional genetic methods and the decay process leaves them highly repetitive and heterochromatic, making them very difficult to sequence and assemble (but see [9]). Indeed, the most universal observation about sex specific chromosomes (Y or W) is that they rapidly accumulate transposable elements and become heterochromatic. While genomes from both sexes have been included in many whole-genome sequencing projects, efforts to systematically mine this potential source of Y chromosome information are relatively recent (e.g. [10,11]) and in most species Y-chromosome reads remain relegated to the large bin of small, unmapped scaffolds. Consequently, the most detailed information we have about Y-chromosome origins, structure, and gene content is based on concerted efforts to sequence a few species of mammals [9,12,13,14,15*,16,17] and Drosophila [18–23].

The D. melanogaster Y, which is thought to represent the ancestral Y of the genus, has only 13 protein-coding genes, all of which derive from autosomal duplications. One of them, Suppressor of Stellate, is multicopy and has convergently acquired and amplified the number of X-linked copies. Intriguingly, Soh and colleagues [9] recently discovered a similar convergent amplification of genes on the mouse X and Y. In both mouse and D. melanogaster the dually acquired X and Y amplicons may be the relic of a co-evolved meiotic driver/Suppressor system [24–26]. The lack of orthologous genes between the ancestral Drosophila X and Y has caused some to speculate that the Y arose from a supernumerary B chromosome [20,21,27,28]. However, it is also possible that homology has been erased by gene losses in the ancestor of...
Drosophilid flies. The latter explanation finds indirect support from three lines of evidence in concert. First, there is considerable turnover of Y-linked genes across the genus. Only 3 of the 12 single copy Y-linked autosomal protein-coding genes are conserved across all members of the genus studied so far, except for the *D. pseudoobscura* lineage, which shares none [21]. Second, the *D. pseudoobscura* neo-Y is highly degenerated and heterochromatic, like the *D. melanogaster* Y, yet is not homologous. In the *D. pseudoobscura* lineage, the ancestral *Drosophila* X fused to Muller element D (Figure 1) and the ancestral Y reverted to autosomal inheritance by fusing to the dot chromosome [29,30]. Despite being only 13–15 million years old (compared to >60 for the ancestral Y) the neo-Y of *D. pseudoobscura* retains only ~15 genes (or pseudo-genes) [18], suggesting that the ancestral Y of *D. melanogaster* has had ample time to lose all of its X-linked homologs. Finally, investigations of the neo-sex chromosomes in *D. miranda* and *D. albomicans* suggest that Y-chromosomes begin the decay process very early by down regulating gene expression of Y-linked copies, and then rapidly begin to lose genes [22,31,32]. In *D. miranda* the neo-Y, which formed by a Y to autosome fusion approximately 1–2 million years ago, has already accumulated a large number of transposable elements and 40% of the ancestral autosomal genes have been pseudogenized or lost [23]. The data from *Drosophila* are consistent with theoretical expectations and from Y-chromosome observations in therian mammals where the Y rapidly lost genes early in its evolution but has remained stable and gained several genes that are exclusively expressed in the male germline [15*].

Unlike evolutionary theory and observations surrounding the decay of the Y, we have a relatively limited understanding of the factors that govern rates of Y chromosome gain and loss (i.e. XO->XY or XY->XO transitions). However, the mechanics of meiosis might play a role [5*]. Analysis of all available Coleoptera sex chromosome karyotype data (4724 karyotypes analyzed over a phylogeny with 1126 operational taxonomic units) shows that species requiring chiasmatic meiosis gain and lose Y-chromosomes much faster than species that evolve alternative segregation mechanisms (e.g. achiasmatic meiosis or asynaptic sex chromosomes). Based on this observation Blackmon and Demuth proposed the fragile Y hypothesis, suggesting that as recurrent selection to
reduce recombination between the X and Y-chromosome continually shrinks the pairing region, it increases the probability of aneuploid gamete production (i.e. Y-chromosome loss) in taxa that require XY chiasmata for proper segregation during meiosis. The fragile Y’s predicted association between meiotic pairing mechanisms and Y-chromosome loss also appears to hold in mammals, but with much more limited data. Testing the fragile Y hypothesis in other insects remains a promising area for research, but awaits more complete phylogenetic sampling and information about meiotic machinery in additional groups.

X-chromosome evolution

Because X-chromosomes recombine in females, they do not suffer the same mutational decay as Y-chromosomes and are consequently better represented and assembled in whole-genome sequencing projects. While they are more like autosomes than Y-chromosomes, Xs also experience evolutionary pressures that shape their content uniquely, such as the inequitable distribution between males and females, and the need to compensate gene expression levels for the loss of Y homologs (i.e. dosage compensation, \[^{[33]}\]). Perhaps the most striking discoveries over the last decade are that, 1) X chromosomes of even closely related lineages may have little or no homology, 2) transposable element (TE) invasion may facilitate the evolution of dosage compensation, and 3) even old X-chromosomes can return to autosomal inheritance.

Within the genus Drosophila there are at least four neo-X chromosome systems, each with unique origins (Figure 1). The D. pseudoobscura X-autosome fusion noted above is the oldest and has evolved full dosage compensation using the same molecular mechanisms found on
the ancestral X (Figure 1.1). The more recent neo-X chromosomes of *D. miranda* (1–2 mya) arose by a fusion of the Y to Muller element C leaving the X to segregate as two pieces (ancestral X + D as in *D. pseudoobscura*, and the neo X = ancestral C; Figure 1.2). Interestingly, the variety of ages among the three X-chromosome segments in *D. miranda* indicates that early TE inversion provides important regulatory sequences that are refined over time to efficiently recruit the dosage compensation machinery [34*]. Two other neo sex chromosome systems arose from yet other fusions. In *D. albomicans* the X and Y both fused to a single autosome containing Muller C + D elements (~0.1 mya) [32] (Figure 1.3) and in *D. americana* there is an X to Muller B fusion that is still segregating in natural populations [35] (Figure 1.4). Looking more broadly, it was recently shown that the ancestral X of *Drosophila* is not even the ancestral X of Diptera. Rather, the *Drosophila* dot chromosome, or Muller element F, was the ancestral X but reverted to autosomal inheritance in *Drosophila*, thus explaining many of the peculiar sex chromosome-like features of the dot [36*].

Given the revolving door of X-chromosome evolution in Diptera, and the overall turnover of sex chromosome systems suggested in Box 1, perhaps it is not surprising that the shared region of the *D. melanogaster* and *Anopheles gambiae* X is not homologous to the X of the beetle *Tribolium castaneum* or the Z of the silk moth *Bombyx mori*, but in each case the X (or Z) is homologous to a unique autosome in the other species [37**] (Figure 1). Note that the lack of homology should not be interpreted as evidence that chromosomal sex determination arose de novo in Diptera, Coleoptera, and Lepidoptera; rather it is presumably a product of the kinds of sex chromosome to autosome fusions we observe in *Drosophila* playing out over deeper time. The rarity of either hermaphroditism or environmental sex determination in hexapods suggests that the ancestor of all insects had genetic sex determination. Furthermore, the presence of XY systems in non-insect hexapods and most early branching insect orders, suggests that XY is the most likely ancestral sex determination mode in insects.

Despite the lack of homology among X chromosomes, we might expect that similar evolutionary pressures would produce convergent characteristics. However, among four molecular characteristics of *Drosophila* X-chromosomes: 1) gene movement, 2) codon bias, 3) gene density, and 4) sex-biased gene expression, only gene movement shows a consistent pattern outside Diptera [37**]. Biased gene movement from the X to autosomes is seen in all insect X-chromosomes surveyed to date. The pattern is particularly strong for duplication by retrotransposition, but also holds for DNA based duplications [37**,38,39,40,41] and extends even to most mammals surveyed (e.g. [42]). Autosomal duplicates tend to be more highly expressed in the male germline than their X-linked precursors in *Drosophila* spp., which fosters a hypothesis that genes move off of the X to avoid meiotic sex chromosome inactivation (MSCI) in the male germline [43]; however, the presence of MSCI in *Drosophila* is presently controversial [44–47] and is not known in other insects.

**ZW systems**

There are only two insect orders (Lepidoptera and Trichoptera) that have predominantly ZW sex chromosomes and data relevant to their sex chromosome origins is scarce. Trichoptera and the basal lineages of Lepidoptera are all ZO whereas possessing a W chromosome arose later and is characteristic of the ‘advanced’ Lepidoptera. While the W chromosome in some taxa arose by fusion of the ancestral Z to an autosome, in other taxa the origin of the W is less clear [48,49]. Most of what we know about W chromosome homology and evolution is based on fluorescent *in situ* hybridization (FISH) using whole-genome and or microdissected W-chromatin based probes. These studies demonstrate that W-chromosomes, like Y-chromosomes, are mostly heterochromatic and can lack homology even within families [50–52]; however, they are often asymmetric in that they may show lack of homology to the *Bombbyx* W, for example, but they do not identify the neo-W’s autosomal progenitor. We are aware of only one effort to directly sequence a Lepidoptera W-chromosome (*Ephistia kuehniella*) and they found many diverse transposable elements but no protein coding genes that would be useful for elucidating origins [53*]. Interestingly, recent RNA-seq experiments in silkworm discovered that the only transcripts generated by the W encode a dominant female determining factor, *Fem*, which is the precursor to a PIWI-interacting RNA [54].

Unlike the X, Y, and W chromosomes highlighted thus far, the ancestral Z chromosome of *Lepidoptera* seems to remain broadly conserved. However, this view relies on shared Z-linkage across seven species and the silkworm reference genome, which provides only limited evidence [48]. The paucity of data for ZW species merits caution in making strong conclusions. Indeed, a recent study incorporating FISH probes based on autosomal and Z linked genes in silkworm, discovered that a neo-Z chromosome arose by a fusion of the Z to *Bombbyx* chromosome 15 in the common ancestor of ~700 pest species [55].

**Conclusions**

Our understanding of insect sex chromosome origins and content is highly reliant on studies in Diptera; particularly *Drosophila* where we have the most detailed genomic data. Ironically, the rate of sex chromosome turnover we see in *Drosophila* suggests that we should not overly rely on them to make generalities about sex chromosomes in the rest of insects. Box 1 also reveals important biases in what we know. For instance, there is a paucity of data for female heterogametic taxa, which is partly an artifact of the technical difficulty collecting karyotype data in females and also because Lepidoptera have numerous
(2n = 60–62 is most common) holocentric chromosomes, making cytogenetic distinction of Z and W difficult [48]. Going forward, the reduced cost of whole-genome sequencing is likely to rapidly improve what we know about sex chromosomes. When genomes are sequenced from both sexes we now have multiple strategies for using read depth (e.g. [11,36]) and or kmer frequencies [10] to identify sex chromosome scaffolds. However, sex limited genomes of the Y or W are likely to remain recalcitrant until accurate, long-read sequencing is available. Creation of the Tree of Sex Database offers access to sex determination data for over 20,000 species across the tree of life [4**,56**]. Much of the data has never been analyzed in a comparative framework and could also offer new insights into the evolution of sex chromosomes (e.g. [57]).

References

34. Ellison CE, Bachtrog D: Dosage compensation via transposable element mediated rewiring of a regulatory network. Science 2013, 342:846-850. This study suggests that transposable elements may play a key role in rapid evolution of dosage compensation on neo-sex chromosomes.


36. Vicoso B, Bachtrog D: Reversal of an ancient sex chromosome to an autosome in Drosophila. Nature 2013, 490:332. The authors sequence males and females separately and use coverage depth differences to identify one of the first examples of a heteromorphic sex chromosome reverting to autosomal inheritance.

37. Pease JB, Hahn MW: Sex chromosomes evolved from independent ancestral linkage groups in winged insects. Mol Biol Evol 2012, 29:1645-1653. Using a bioinformatics approach the authors show that the X chromosomes in Diptera, Coleoptera and Lepidoptera are not homologous.


56. Tree of Sex Consortium: Sex determination: why so many ways of doing it? PLoS Biol 2014, 12:e1001899. Reviews a wide variety of sex determination methods and discusses the possible forces and pathways that may lead to transitions among them.