Sex Chromosomes Evolved from Independent Ancestral Linkage Groups in Winged Insects

James B. Pease[1] and Matthew W. Hahn*[,1,2]

[1]Department of Biology, Indiana University

[2]School of Informatics and Computing, Indiana University

***Corresponding author:** E-mail: mwh@indiana.edu.

Associate editor: John Parsch

Abstract

The evolution of a pair of chromosomes that differ in appearance between males and females (heteromorphic sex chromosomes) has occurred repeatedly across plants and animals. Recent work has shown that the male heterogametic (XY) and female heterogametic (ZW) sex chromosomes evolved independently from different pairs of homomorphic autosomes in the common ancestor of birds and mammals but also that X and Z chromosomes share many convergent molecular features. However, little is known about how often heteromorphic sex chromosomes have either evolved convergently from different autosomes or in parallel from the same pair of autosomes and how universal patterns of molecular evolution on sex chromosomes really are. Among winged insects with sequenced genomes, there are male heterogametic species in both the Diptera (e.g., *Drosophila melanogaster*) and the Coleoptera (*Tribolium castaneum*), female heterogametic species in the Lepidoptera (*Bombyx mori*), and haplodiploid species in the Hymenoptera (e.g., *Nasonia vitripennis*). By determining orthologous relationships among genes on the X and Z chromosomes of insects with sequenced genomes, we are able to show that these chromosomes are not homologous to one another but are homologous to autosomes in each of the other species. These results strongly imply that heteromorphic sex chromosomes have evolved independently from different pairs of ancestral chromosomes in each of the insect orders studied. We also find that the convergently evolved X chromosomes of Diptera and Coleoptera share genomic features with each other and with vertebrate X chromosomes, including excess gene movement from the X to the autosomes. However, other patterns of molecular evolution—such as increased codon bias, decreased gene density, and the paucity of male-biased genes on the X—differ among the insect X and Z chromosomes. Our results provide evidence for both differences and nearly universal similarities in patterns of evolution among independently derived sex chromosomes.

Key words: evolution, X chromosome, Z chromosome.

Introduction

Animals and plants have evolved an array of different sex-chromosome karyotypes, including male heterogametic (XX/XY), female heterogametic (ZZ/ZW), and homomorphic systems (Bull 1983; Charlesworth and Mank 2010). The generally accepted model for the evolution of heteromorphic sex chromosomes proposes that these chromosomes evolved from a pair of autosomes containing a sex-determining locus (Rice 1996; Charlesworth et al. 2005; Bachtrog 2006). Natural selection is then thought to favor the tight linkage of sex-determining alleles with sexually antagonistic alleles (i.e., alleles that are favorable in one sex but harmful in the other), which further selects for inhibited local recombination between the proto-sex chromosomes (Fisher 1931; Bull 1983; Rice 1987). Though there is some disagreement about the mechanisms that drive the cessation of recombination (Ironside 2010), the end result of this process is two highly differentiated chromosomes.

Exactly how and why autosomes containing sex-determining loci begin a progression to heteromorphic sex chromosomes—and whether these will become XY or ZW pairs—is unclear, especially as there appear to be stable homomorphic systems in many taxa (Gilchrist and Haldane 1947). One further complication is that in some homomorphic species the sex-determining genes can move between undifferentiated autosomes (e.g., in salmonid fishes; Woram et al. 2003). This movement implies that the progression to heteromorphic sex chromosomes may involve different pairs of autosomes, even among closely related taxa that differ only in the location of the sex-determining locus (Ross et al. 2009). Indeed, a comparison of the Z chromosome of birds with the X chromosome of therian mammals (marsupials and placentals) demonstrated that the sex chromosomes are not homologous to one another but rather to different autosomes in each species (Bellott et al. 2010). This strongly suggests that these sex chromosomes independently evolved from different autosomal pairs in the amniote ancestor (Bellott et al. 2010).

While finding independent origins for the X and Z chromosomes in amniotes is striking, these sex chromosomes appear to have evolved only once in each clade; all therian mammals have homologous XY systems and all birds have homologous ZW systems (Graves 2008). Further study of the frequency of parallel or convergent evolution of sex chromosomes requires a clade with a number of different sex-chromosome systems represented in disparate taxa.

Fish are one such clade (Mank et al. 2006; Graves and Peichel 2010), though there are relatively few sequenced genomes in this group. Another opportunity is found in the winged insects, where sex-chromosome systems are highly variable even within individual orders and there are whole-genome sequences representing each system. For instance, the order Diptera contains species with XX/XY, XX/XO, ZZ/ZW, multiple-X, and homomorphic sex-chromosome karyotypes (Marin and Baker 1998; Rai and Black 1999; Davies and Roderick 2005; Kaiser and Bachtrog 2010). Based on whole-genome data, relationships between the sex chromosomes within the Diptera are fairly clear. The presence of a large number and proportion of shared orthologs on the X chromosomes of the fruit fly *Drosophila melanogaster* and the mosquito *Anopheles gambiae* strongly suggests that they are derived from a homologous ancestor (Zdobnov et al. 2002) but not necessarily one that was a differentiated X chromosome in the ancestral species. These X chromosomes are also homologous to one arm of chromosome 1—which contains the sex-determining locus—in the homomorphic mosquito, *Aedes aegypti* (Nene et al. 2007). However, the X chromosome in the stalk-eyed fly *Teleopsis dalmanni* is not homologous to the other sequenced dipteran X chromosomes, instead sharing greatest similarity with *D. melanogaster* chromosome 2L (Baker and Wilkinson 2010). The phylogenetic distribution of sex-chromosome systems (Marin and Baker 1998; Rai and Black 1999) and the patterns of X-chromosome evolution (Toups and Hahn 2010) suggest that both the common ancestor of all dipterans and the more recent common ancestor of mosquitoes had homomorphic sex chromosomes. This implies that the evolution of differentiated sex chromosomes has occurred both from the same pair of autosomes in *D. melanogaster* and *An. gambiae* (parallel evolution) and from a different pair of autosomes in *Te. dalmanni* (convergent evolution).

Among all winged insects with fully sequenced genomes, several orders and sex-chromosome karyotypes are represented. These include XX/XY in the Diptera (e.g., *D. melanogaster* and *An. gambiae*) and Coleoptera (*Tribolium castaneum*), ZZ/ZW in the Lepidoptera (*Bombyx mori*), homomorphic systems in the Diptera (*Ae. aegypti*), and haplodiploidy in the Hymenoptera (*Nasonia vitripennis*). Although phylogenetic evidence points to incompletely heteromorphic or homomorphic sex chromosomes in the ancestor of winged insects, nothing outside of Diptera is known about the relationships among the sex chromosomes. One general hypothesis is that there may have been factors that predisposed one or more ancestral autosomes to later transform into the sex chromosomes, possibly because of the presence of conserved sex-determining genes (Graves and Peichel 2010). Considering just the Dipteran X, the Coleopteran X, and the Lepidopteran Z, there are three specific models that could explain the evolution of sex chromosomes given the currently established phylogeny (Huerta-Cepas et al. 2010). In a model of complete independence, all three lineages would have evolved sex chromosomes from distinct ancestral autosomes (fig. 1*A*). In a model of partial independence, the two X chromosomes may have evolved from a "proto-X" autosome, whereas the Z originated independently (fig. 1*B*). Finally, in a single-origin model, all three sets of sex chromosomes may have evolved from a common "proto-sex" autosome, which may have had properties predisposing its transformation into the various heteromorphic sex chromosomes (fig. 1*C*). Distinguishing among these models will provide evidence on the propensity for parallel versus convergent evolution of sex chromosomes.

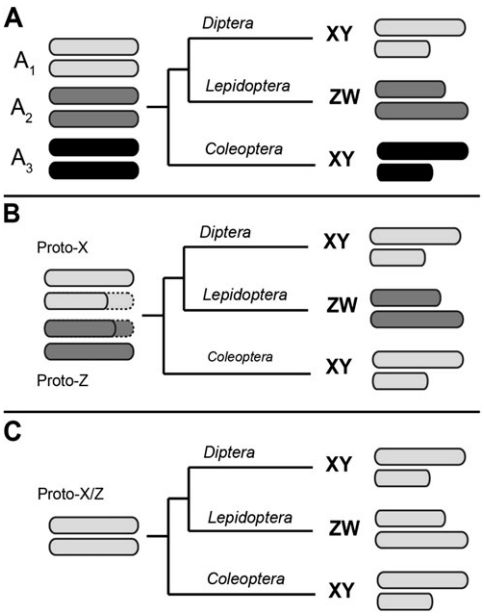

FIG. 1. Three possible models of sex-chromosome evolution in three orders of insects: (*A*) complete independence, where all sex chromosomes originate independently, (*B*) partial independence, where the X and Z originate independently, (*C*) single origin, where all sex chromosomes evolved from a proto-sex chromosome. ("A" = ancestral autosome.)

Comparative genomics has also revealed many molecular features of sex chromosomes that appear to have evolved independently in multiple lineages. In those cases where the gene content of the sex chromosomes is known, the heterogametic chromosome (Y or W) has repeatedly lost many genes and genetic functions while gaining a large number of transposable elements and other repetitive sequences (Lahn and Page 1997; Skaletsky et al. 2003; Abe et al. 2008; Bachtrog et al. 2008). The homogametic sex chromosomes (X or Z) of several species have also been shown to have distinct properties of gene composition and gene evolution when compared with autosomes (Vicoso and Charlesworth 2006; Ellegren and Parsch 2007; Wilson and Makova 2009; Charlesworth and Mank 2010; Kaiser and Bachtrog 2010; Ellegren 2011). Among insect X chromosomes (particularly the *Drosophila* X), previous analyses have found unique patterns of evolution in terms of: excess gene movement off the X (Betran et al. 2002; Dai et al. 2006; Meisel et al. 2009; Vibranovski et al. 2009), increased codon bias on the X (Hambuch and Parsch 2005; Singh et al. 2005), and a decreased proportion

Table 1. Summary of Insect Genomes.

	Anopheles gambiae	*Bombyx mori*	*Drosophila melanogaster*	*Nasonia vitripennis*	*Tribolium castaneum*
Sex karyotypes	XX/XY	ZZ/ZW	XX/XY	Haplodiploid	XX/XY
Chromosomes/arms	5	28	5	5	10
Total genes	11,603	14,623	13,833	18,822	16,564
Unmapped genes	491 (4.2%)	834 (5.7%)	70 (0.5%)	5,838 (31%)	2,431 (14.7%)
Autosomal gene density (genes/Mb)	53.2	27.2	119.3	69.5	78.4
X/Z gene density (genes/Mb)	43.8	29.2	97.3	N/A	73.1

of male-biased genes on the X (Parisi et al. 2003; Sturgill et al. 2007). In addition, a comparison of the mammal X and bird Z revealed convergent duplication of many testis-expressed genes and concomitant expansion of intergenic regions (Mueller et al. 2008; Bellott et al. 2010); however, there is no such increase in testis-biased genes on the *Drosophila* X chromosome (Parisi et al. 2003; Sturgill et al. 2007), though to our knowledge no analysis of gene density on the X relative to autosomes has been reported. The observation of both shared and uniquely evolved features of unrelated homogametic sex chromosomes suggests that there may be lineage-specific forces driving their evolution. Therefore, we also examined many of these patterns of molecular evolution among winged insects.

The increasing number of complete genomes in winged insects provides a new opportunity to study the evolution of sex chromosomes. In the following, we examine the genomes of five species that both encompass a variety of sex-specific karyotypes and have a majority of their sequenced genomes assigned to physical chromosomes: *An. gambiae* (XX/XY), *B. mori* (ZZ/ZW), *D. melanogaster* (XX/XY), *Tr. castaneum* (XX/XY), and *N. vitripennis* (haplodiploid). The comparison of these different systems will allow us to distinguish among models for the evolution of winged insect sex chromosomes (fig. 1), while also revealing any patterns of shared and unique characteristics of sex chromosomes.

Materials and Methods

A complete set of peptide sequences were obtained for *An. gambiae* (VectorBase v3.6; Lawson et al. 2009), *B. mori* (KAIKObase v2.0; Shimomura et al. 2009), *D. melanogaster* (FlyBase v5.32; Tweedie et al. 2009), *N. vitripennis* (NasoniaBase v1.2; Munoz-Torres et al. 2011), and *Tr. castaneum* (BeetleBase v3.0; Kim et al. 2010). A summary of the genome data is shown in table 1. Peptide sequence local alignment scores for the full set of five genomes were calculated using mpiBLAST v1.5.0 (http://www.mpiblast.org). BlastP bit scores < 200 were excluded, and the remaining genes were clustered using MCL (v10.201; van Dongen 2000). After clustering, we removed all proteins unmapped to chromosomes in their respective genomes. Of the 14,082 clusters, 1,514 contained exactly one peptide from each of the five species. These likely represent one-to-one orthologs across species and are used for many of the analyses included here; we refer to them as "exclusive ortholog sets."

In order to quantify the degree to which chromosomes are homologous, we first counted the number of exclusive ortholog sets on each chromosome for all possible pairings among the five species. Because there are different numbers of chromosomes in each species and different sizes of chromosomes within each species, simply counting the number of exclusive orthologous sets shared between chromosomes is not a good measure of homology. For instance, a large chromosome in one species would mistakenly be inferred as homologous to many smaller chromosomes in each of the other species. Therefore, in order to accurately assess the probability that two chromosomes share a large number of orthologous genes because of shared history and not convergent acquisition of a small number of transposed genes, we generated a randomized distribution of orthologous pairings. For each pair of species, we first distributed 1,514 simulated genes among chromosomes proportional to the true number of total genes on each chromosome for each of the two species and then randomly paired genes in order to simulate exclusive ortholog sets. (We treated chromosome arms as whole chromosomes in *An. gambiae* and *D. melanogaster*.) The observed number of shared orthologs for each pair of chromosomes was then compared against 100,000 randomly generated distributions. The resulting *P* values represent the proportion of simulated ortholog pairs that were greater than that observed in the data. Cases where the *P* value is less than 10^{-5} indicate that the number of exclusive ortholog sets shared between two chromosomes exceeded all values in the simulated data.

In order to measure codon usage bias, we computed the effective number of codons (ENC; Wright 1990) for all coding sequences on each chromosome in the four heteromorphic species. ENC measures the number of codons used per gene, without the need to specify optimal codons; it is therefore a useful measure for comparisons across species. Differences in codon usage among chromosomes were tested using Wilcoxon's rank test.

To investigate patterns of gene movement to and from X chromosomes, we took advantage of the fact that the movement of genes duplicated via retrotransposition can be easily polarized. Retrotransposition occurs when a parental gene is transcribed, the resulting mRNA is reverse transcribed into cDNA, and it is then inserted into the genome at a new location (Hollis et al. 1982; Karin and Richards 1982; Ueda et al. 1982). The daughter gene therefore lacks all introns—they are removed during processing of the mRNA—but retains coding exons and untranslated regions. One can therefore polarize the direction of duplication from parental genes (with introns) to daughter genes (without introns). To find retrotransposed

duplicates, we separately calculated Blast scores and clustered with MCL for all peptides in a single species. Within-species clusters that include only one gene sequence with no introns and one gene sequence with >1 intron were considered putative daughter retrogenes and parental genes, respectively. The location of each gene on a chromosome was recorded, discarding any pair that did not have both genes mapped to a chromosome as well as any pairs with both genes on the same chromosome. We retained only retrogene-parent pairs with sequence similarity >50% and where the retrogene was at least 70% the length of the parent gene; sequence identity was calculated using the Needleman–Wunsch algorithm (Python module nwalign v0.3.1). In order to determine whether there was biased movement of genes between chromosomes, we compared the observed counts of movements from the X to autosomes, from the autosomes to the X, and between autosomes to the values expected based on individual chromosome lengths and the number of genes on each chromosome (Betran et al. 2002). Observed and expected values were compared with a χ^2 goodness-of-fit test (degrees of freedom = 2).

Results and Discussion

Independent Origins of Sex Chromosomes

We used the presence of shared orthologous genes to assess the homology of individual chromosomes among the winged insects. Our randomization method (see Materials and Methods) allowed us to control for the possibility that two chromosomes share genes by chance alone, which is equivalent to a scenario in which there has been so much gene movement between chromosomes that there is no signal of shared history. Low *P* values for the number of exclusive ortholog sets shared between two chromosomes implies that there is a low probability that the observed number of genes would occur at random. The most parsimonious inference is therefore that the chromosomes have a shared history. Conversely, because multiple linked genes can be moved between chromosomes via single mutational events, significant pairings do not necessarily indicate wholesale homology between chromosomes; instead, they indicate that there are significant portions of the two chromosomes that have a common ancestor. Many previous studies of chromosomal homology have taken the spatial relationship of genes into account (i.e., the synteny of contiguous gene blocks). Because very high rates of rearrangement within single insect chromosomes can occur even within a genus (e.g., *Drosophila*; Bhutkar et al. 2008), resolving these movements among species in distantly diverged orders is likely a near impossibility. Considering these high rates and the limitations of current genome assemblies, we did not consider the spatial relationship of genes along chromosomes here.

Because patterns of chromosomal homology are already established for *D. melanogaster* and *An. gambiae* (Zdobnov et al. 2002), we were able to use this comparison as a positive control for our ortholog clustering method. Consistent with previous work, we found highly significant pairings between *D. melanogaster*:*An. gambiae* chromosomes X:X, 2L:3R, 2R:2L, 3L:2L, and 3R:2R (supplementary table S1*b*, Supplementary Material online). These results suggest that our method correctly identifies homologous relationships between chromosomes, at least at this level of divergence (200–250 million years; Gaunt and Miles 2002; Wiegmann et al. 2009). This comparison also confirms the homology of the Dipteran X chromosomes as chromosomes, even if these were not differentiated in their most recent common ancestor.

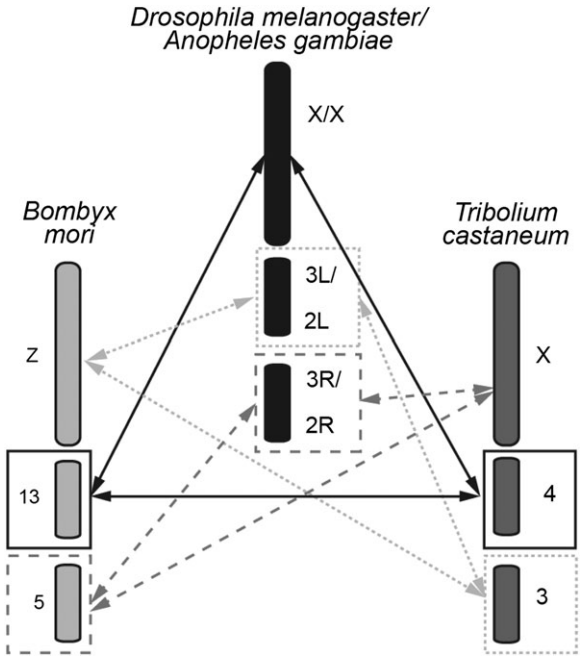

FIG. 2. Most likely relationships between sex chromosomes in each species. The sex chromosomes of each species show homology to autosomes in the other species but not to sex chromosomes. Only the most significant relationship ($P < 0.00001$) for each X or Z and their respective top hits is shown (for full results, see supplementary table S1, Supplementary Material online).

All other sex chromosomes showed no significant similarity with each other, though each did show significant similarity to autosomes in the other species considered. Figure 2 summarizes the relationships between individual sex chromosomes and the autosome sharing the greatest numbers of genes with them ($P < 10^{-5}$) in the Diptera, Coleoptera, and Lepidoptera (see also supplementary table S1, Supplementary Material online). For instance, the *B. mori* Z chromosome shares a significant number of orthologs with *D. melanogaster* chromosome 3L but not with the X chromosome of this species. Among X chromosomes, the X of *D. melanogaster*/*An. gambiae* shares only three exclusive orthologs with the X of *Tr. castaneum*. This is less than ~0.3% of all genes on the Dipteran X and is comparable to the number shared with the Z chromosome of *B. mori* (4, or ~0.4% of all genes). Although the X of *Tr. castaneum* is homologous to *D. melanogaster* 3R and *An. gambiae* 3R, each of which also shares homology with the *D. melanogaster*/*An. gambiae* X, it is not the same genes that are orthologous in these

comparisons (supplementary fig. S1, Supplementary Material online).

Taken together, our data support the independent evolution of sex chromosomes from separate ancestral linkage groups in all three insect orders (fig. 1A). The independent evolution of the XX/XY karyotypes in both the Coleopteran and Dipteran lineages also suggests that the sex chromosomes had not fully differentiated in the shared ancestor of Lepidoptera and Diptera, which is further supported by analyses strongly suggesting that the Dipteran common ancestor did not have differentiated sex chromosomes (see Introduction). However, we cannot be sure that the ancestor of all winged insects was homomorphic given the apparent frequency with which different sex chromosome systems have evolved. Furthermore, the ancestral chromosomes were likely different for each sex chromosome (consistent with the model presented in fig. 1A), indicating that there were no predisposing features of the ancestral chromosomes that caused them to evolve in parallel toward the various forms of heterogamety. While these insect orders last shared a common ancestor ~300 Ma (Wiegmann et al. 2009), and the sequenced genomes may not be representative of every member of their respective orders, our results do strongly support at least three independent origins of heterogametic sex chromosomes. Relationships between the X and Z chromosomes in the heteromorphic species and the chromosomes of *N. vitripennis* were mixed (supplementary table S1 and also see table S2, Supplementary Material online), suggesting that there may not have been a large autosome present in the common ancestor of all winged insects that was the progenitor of all X chromosomes (cf. fig. 1B). However, it is also true that the *N. vitripennis* genome has the largest proportion of unmapped genes of all species considered here (table 1); continued efforts to map genes onto chromosomes in this species may either overturn or further support a single ancestral chromosome.

We considered several aspects of our analyses that could have impacted our results. It was possible that we would not have enough power to detect significant relationships among the sex chromosomes and, therefore, would not be able to detect shared histories even if they existed. However, the fact that we do observe highly significant signals of shared histories between the sex chromosomes and individual autosomes obviates this concern. Assignments of orthology can often be highly dependent on initial clustering parameters, so we reclustered genes with bit-score cutoffs of 35, 70, 100, and 200, while also varying the clustering granularity (MCL parameter "I") from 4 to 7. We found that all homologous relationships were qualitatively robust to these parameter changes (supplementary table S3, Supplementary Material online). We also used orthologous assignments that allowed for multiple paralogs within any species, as long as there were only genes from a single chromosome for each species (inclusive ortholog sets). This allowed clusters on the same chromosome to be counted as a single orthologous comparison point but avoided ambiguous clusters where peptides were present on different chromosomes in the same species. The change in methodology increased the number of orthologous sets to approximately 2,800–3,800 for each pair of species but again caused no change to the homologous relationships observed among the sex chromosomes (supplementary table S4, Supplementary Material online). The robustness of the results to varying parameters and methods of analysis suggests that our conclusions are also robust.

To independently confirm our results using pairwise orthology assignments, we downloaded ortholog relationships between all pairs of the species considered here from the InParanoid database (Ostlund et al. 2010). The number of orthologs increased to 4,300–5,300 for the various pairwise comparisons, but all homologous relationships remained the same as in our original analysis (supplementary tables S2 and S4, Supplementary Material online). Comparison of the exclusive ortholog sets, inclusive ortholog sets, and InParanoid ortholog sets showed that all pairwise chromosome comparisons between species had correlations >0.95 for the number of shared genes (supplementary table S5, Supplementary Material online).

Convergent Features of Sex Chromosomes

Previous analyses of the mammalian X and bird Z chromosome identified a number of convergent molecular features, including low gene density compared to autosomes and an increased number of testis-specific genes (Bellott et al. 2010). We were therefore interested in whether there were convergent features of insect homogametic sex chromosomes. In particular, we examined four molecular features of the sex chromosomes considered here: 1) gene movement off the X and Z, 2) codon bias on the X and Z relative to autosomes, 3) gene density on the X and Z relative to autosomes, and finally, we discuss previous results on 4) sex-biased gene expression on the X and Z.

One striking feature of the *Drosophila* X chromosome is the pattern of gene movement involving this chromosome. In all *Drosophila* species examined, there is an excess of genes that are duplicated via retrotransposition that move from the X chromosome to the autosomes (Betran et al. 2002; Dai et al. 2006; Meisel et al. 2009; Vibranovski et al. 2009). This pattern has also previously been found on the X chromosome of *An. gambiae* (Toups and Hahn 2010) and on the X chromosomes of multiple mammalian species (Emerson et al. 2004; Vinckenbosch et al. 2006; Potrzebowski et al. 2008), though it is only detected in two of the four *Teleopsis* species examined (Baker and Wilkinson 2010).

As an excess of movement off the X has already been observed in *D. melanogaster* and *An. gambiae*, we examined the movement of retrogenes in the *Tr. castaneum* genome by comparing the location of parent and daughter genes (see Materials and Methods). We found a highly significant pattern of excess retrotransposition from the X chromosome to the autosomes in *Tr. castaneum* ($P < 10^{-29}$; table 2), consistent with previous results from XY systems. This excess movement in an independently evolved X chromosome further suggests that gene traffic off the X may be a nearly

Table 2. Pattern of Retrotransposition in *Tribolium castaneum*.

	Observed	Expected
X→A	15	1.4
A→X	0	1.1
A→A	9	21.5

universal feature of this chromosome, especially as it occurs even in newly evolved X chromosomes (i.e., neo-Xs; Meisel et al. 2009). Interestingly, similar analyses of the Z chromosome in *B. mori* and the chicken genome (also a ZW system) revealed no significant patterns of movement, indicating that X-biased retrotransposition appears to be a specific property of species with XY chromosome systems (Hillier et al. 2004; Toups et al. 2011). Again, no significant deviations in the results were found when Blast, clustering, or sequence identity parameters were changed.

Previous studies have shown more codon bias in *D. melanogaster* for genes that are X-linked, regardless of whether they are sex biased in expression (Hambuch and Parsch 2005; Singh et al. 2005). The fact that more codon bias for X-linked genes was also found in *D. pseudoobscura* and *C. elegans* suggested that this might be a universal property of X chromosomes (Singh et al. 2005). Two hypotheses based on population genetic models have been put forward to explain this pattern: either the strength of selection on translational efficiency is greater on the X due to dosage balance constraints or the effective number of males is much smaller than the effective number of females, making selection for optimal codons on the X more efficient (synonymous changes are assumed to be codominant, precluding an advantage for hemizygous selection against recessive mutations; Singh et al. 2005). If the former hypothesis is correct, we would not expect to see more codon bias on either the *Tr. castaneum* X or the *B. mori* Z chromosome, as both lack global dosage compensation (Zha et al. 2009; Prince et al. 2010). (Note that hypertranscription of the X, as is observed in *Tr. castaneum*, may be an intermediate step in the evolution of dosage compensation, but that this does not affect the predictions of the model.) If the latter hypothesis is correct, we would only expect to see more codon bias on the *Tr. castaneum* X, assuming the effective number of males is lower in all species considered. For both hypotheses, we expect *An. gambiae* to closely resemble *D. melanogaster*.

We compared codon usage among all chromosomes in each of the four species with heteromorphic sex chromosomes considered here, using the ENC statistic (Wright 1990). Smaller values of ENC mean that fewer synonymous codons are being used, which indicates more codon bias, whereas the opposite is true for larger values of ENC. As expected, the X chromosomes of *An. gambiae* and *D. melanogaster* show a lower average ENC per gene than all autosomes ($P < 0.0001$; fig. 3). However, the X chromosome

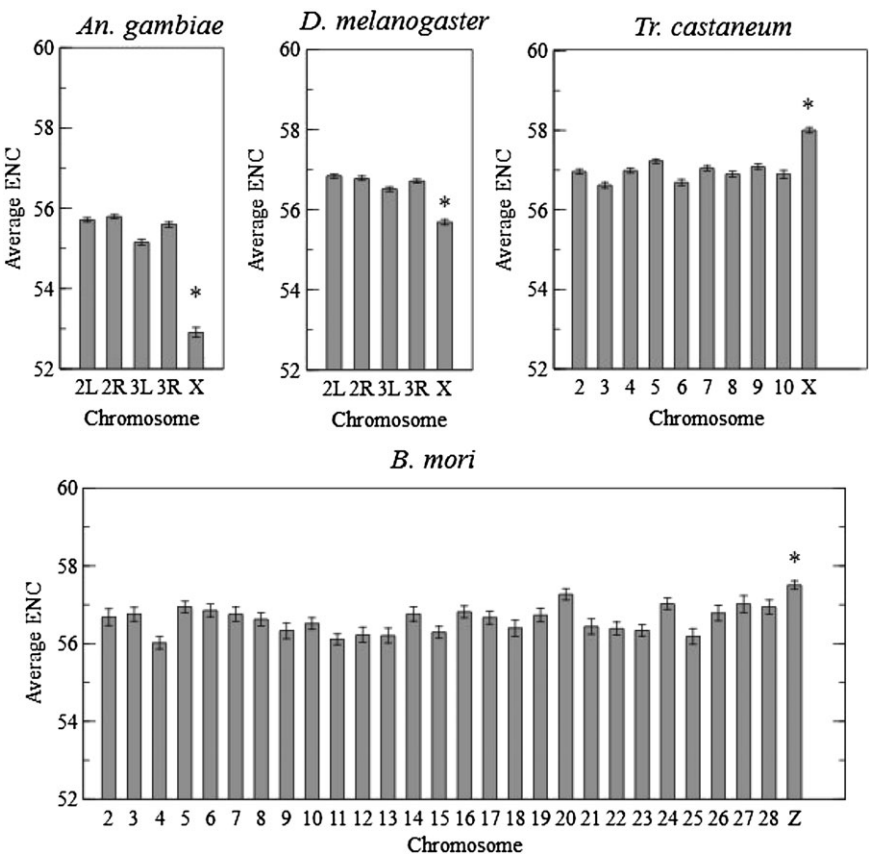

Fig. 3. Average ENC for all genes on a chromosome (standard error bars are shown). The *Drosophila melanogaster* and *Anopheles gambiae* X chromosomes show lower mean ENC compared to autosomes, while the *Tribolium castaneum* X and *Bombyx mori* Z have higher mean ENC (*indicates $P < 0.001$; Wilcoxon's rank test).

in *Tr. castaneum* and the Z chromosome in *B. mori* have a significantly higher ENC than their respective autosomes, indicating less codon bias ($P < 0.0001$; fig. 3). It is clear from the data shown in figure 3 that the average codon bias on autosomes is approximately the same for all four species, with only *An. gambiae* autosomes showing slightly lower ENC than the other taxa. Such comparisons indicate that it is changes in codon bias on the sex chromosomes that differentiate these species and not a genome-wide change in codon usage bias. As outlined above, these results are consistent with a role for dosage compensation on the dipteran X chromosomes driving the increased codon bias, as there does not appear to be dosage compensation on either the *Tr. castaneum* X or the *B. mori* Z chromosome (but see Walters and Hardcastle 2011). Overall, our results show that increased codon bias is not a general pattern of X chromosome (or homogametic sex chromosome) evolution.

As discussed above, analyses of vertebrate genomes have found lower gene densities on the X and Z chromosomes as compared to autosomes in the same species (Ross et al. 2005; Bellott et al. 2010). We reasoned that this pattern could be a universal feature of homogametic sex chromosomes and therefore asked whether there were consistent differences in the gene densities of insect X and Z chromosomes as compared to autosomes. We found that only *D. melanogaster* and *An. gambiae* had significantly lower gene densities on the sex chromosome ($P < 0.001$; table 1). In *B. mori*, gene density was actually higher on the Z chromosome and in *Tr. castaneum* the density was lower but was not significant ($P = 0.97$ and $P = 0.88$, respectively; no such comparison can be made for the haplodiploid *N. vitripennis*). One complication of this analysis is that we do not know whether lower gene density was present in the common ancestor of *D. melanogaster* and *An. gambiae* or whether it evolved along both lineages as the sex chromosomes became differentiated. Knowing the gene density on *Ae. aegypti* chromosome 1—which is homologous to the X chromosomes of the other Dipterans—or on the convergently evolved X chromosome of *Te. dalmanni* would help to answer this question; unfortunately, neither species has a genome assembly of sufficient quality to allow us to calculate gene density. Regardless, results from the species with relatively high-quality assemblies reveal a heterogeneous pattern of convergent evolution; some lineages have evolved lower gene densities and some have not.

Finally, a deficit of male-biased (and especially testis-biased) genes exists on the X chromosomes of all *Drosophila* species examined (Parisi et al. 2003; Sturgill et al. 2007). Previous studies have found that this pattern does not hold on the X chromosomes of *Te. dalmanni* (Baker and Wilkinson 2010), most mammals (Rinn and Snyder 2005), the *B. mori* Z (Arunkumar et al. 2009), or the *An. gambiae* X (Hahn and Lanzaro 2005), although a more recent report has found a deficit in *An. gambiae* (Baker et al. 2011). The X chromosome of *Tr. castaneum* has both a deficit of male-biased genes and an excess of female-biased genes, a pattern not seen in *Drosophila* (Prince et al. 2010). In fact, as mentioned earlier, the X chromosome of mammals has an excess of testis-biased genes (Bellott et al. 2010), as do the Z chromosomes of birds (Storchova and Divina 2006; Bellott et al. 2010) and Lepidoptera (Arunkumar et al. 2009). When considered together with the convergent patterns of gene movement across X chromosomes (including *Tr. castaneum*), it is clear that the movement of male- and testis-biased genes off X chromosomes is not sufficient to drive the observed underrepresentation of male-biased genes on X chromosomes. This result should not be surprising given that the number of male-biased genes "missing" from the *Drosophila* X is an order-of-magnitude larger than the excess number of genes moving off the X (Betran et al. 2002; Parisi et al. 2003; Sturgill et al. 2007; Meisel et al. 2009; Vibranovski et al. 2009). Although we do not know why some aspects of X chromosomes appear to be universally present—even when they must have evolved convergently in multiple lineages—it is clear that not all distinctive features of X chromosomes are shared across taxa.

Supplementary Materials

Supplementary figure S1 and tables S1–S5 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Melissa Toups, Lynda Delph, John Parsch, and three anonymous reviewers for comments on the manuscript and Claudio Casola and Dan Schrider for advice on data analysis. This work was supported by National Science Foundation grant DBI-0845494 (to M.W.H.).

References

Abe H, Fujii T, Tanaka N, et al. (10 co-authors). 2008. Identification of the female-determining region of the W chromosome in *Bombyx mori*. *Genetica* 133:269–282.

Arunkumar KP, Mita K, Nagaraju J. 2009. The silkworm Z chromosome is enriched in testis-specific genes. *Genetics* 182:493–501.

Bachtrog D. 2006. A dynamic view of sex chromosome evolution. *Curr Opin Genet Dev.* 16:578–585.

Bachtrog D, Hom E, Wong KM, Maside X, de Jong P. 2008. Genomic degradation of a young Y chromosome in *Drosophila miranda*. *Genome Biol.* 9:R30.

Baker DA, Nolan T, Fischer B, Pinder A, Crisanti A, Russell S. 2011. A comprehensive gene expression atlas of sex- and tissue-specificity in the malaria vector, *Anopheles gambiae*. *BMC Genomics* 12:296.

Baker RH, Wilkinson GS. 2010. Comparative Genomic Hybridization (CGH) reveals a neo-X chromosome and biased gene movement in stalk-eyed flies (genus *Teleopsis*). *PLoS Genet.* 6:e1001121.

Bellott DW, Skaletsky H, Pyntikova T, et al. (8 co-authors). 2010. Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature* 466:612–616.

Betran E, Thornton K, Long M. 2002. Retroposed new genes out of the X in *Drosophila*. *Genome Res.* 12:1854–1859.

Bhutkar A, Schaeffer SW, Russo SM, Xu M, Smith TF, Gelbart WM. 2008. Chromosomal rearrangement inferred from comparisons of 12 *Drosophila* genomes. *Genetics* 179:1657–1680.

Bull JJ. 1983. Evolution of sex determining mechanisms. Menlo Park (CA): Benjamin/Cummings Publishing Company.
Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* 95:118–128.
Charlesworth D, Mank JE. 2010. The birds and the bees and the flowers and the trees: lessons from genetic mapping of sex determination in plants and animals. *Genetics* 186:9–31.
Dai H, Yoshimatsu TF, Long M. 2006. Retrogene movement within- and between-chromosomes in the evolution of *Drosophila* genomes. *Gene* 385:96–102.
Davies N, Roderick GK. 2005. Dipteran sex chromosomes in evolutionary developmental biology. In: Yeates DK, Wiegmann BM, editors. The evolutionary biology of flies. New York: Columbia University Press. p. 196–213.
Ellegren H. 2011. Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. *Nat Rev Genet.* 12:157–166.
Ellegren H, Parsch J. 2007. The evolution of sex-biased genes and sex-biased gene expression. *Nat Rev Genet.* 8:689–698.
Emerson JJ, Kaessmann H, Betran E, Long M. 2004. Extensive gene traffic on the mammalian X chromosome. *Science* 303:537–540.
Fisher RA. 1931. The evolution of dominance. *Biol Rev.* 6:345–368.
Gaunt MW, Miles MA. 2002. An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. *Mol Biol Evol.* 19:748–761.
Gilchrist BM, Haldane JBS. 1947. Sex linkage and sex determination in a mosquito, *Culex Molestus*. *Hereditas* 33:175–190.
Graves JAM. 2008. Weird animal genomes and the evolution of vertebrate sex and sex chromosomes. *Annu Rev Genet.* 42:565–586.
Graves JAM, Peichel CL. 2010. Are homologies in vertebrate sex determination due to shared ancestry or to limited options? *Genome Biol.* 11:205.
Hahn MW, Lanzaro GC. 2005. Female-biased gene expression in the malaria mosquito *Anopheles gambiae*. *Curr Biol.* 15:192–193.
Hambuch TM, Parsch J. 2005. Patterns of synonymous codon usage in *Drosophila melanogaster* genes with sex-biased expression. *Genetics* 170:1691–1700.
Hillier LW, Miller W, Birney E, et al. (172 co-authors). 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695–716.
Hollis GF, Hieter PA, McBride OW, Swan D, Leder P. 1982. Processed genes: a dispersed human immunoglobulin gene bearing evidence of RNA-type processing. *Nature* 296:321–325.
Huerta-Cepas J, Marcet-Houben M, Pignatelli M, Moya A, Gabaldon T. 2010. The pea aphid phylome: a complete catalogue of evolutionary histories and arthropod orthology and paralogy relationships for *Acyrthosiphon pisum* genes. *Insect Mol Biol.* 19:13–21.
Ironside JE. 2010. No amicable divorce? Challenging the notion that sexual antagonism drives sex chromosome evolution. *BioEssays* 32:718–726.
Kaiser VB, Bachtrog D. 2010. Evolution of sex chromosomes in insects. *Annu Rev Genet.* 44:91–112.
Karin M, Richards RI. 1982. Human metallothionein genes: primary structure of the metallothionein-II gene and a related processed gene. *Nature* 299:797–802.
Kim HS, Murphy T, Xia J, Caragea D, Park Y, Beeman RW, Lorenzen MD, Butcher S, Manak JR, Brown SJ. 2010. BeetleBase in 2010: revisions to provide comprehensive genomic information for *Tribolium castaneum*. *Nucleic Acids Res.* 38:D437–D442.
Lahn BT, Page DC. 1997. Functional coherence of the human Y chromosome. *Science* 278:675–680.
Lawson D, Arensburger P, Atkinson P, et al. (24 co-authors). 2009. VectorBase: a data resource for invertebrate vector genomics. *Nucleic Acids Res.* 37:D583–D587.
Mank JE, Promislow DEL, Avise JC. 2006. Evolution of alternative sex-determining mechanisms in teleost fishes. *Biol J Linn Soc.* 87:83–93.
Marin I, Baker BS. 1998. The evolutionary dynamics of sex determination. *Science* 281:1990–1994.
Meisel RP, Han MV, Hahn MW. 2009. A complex suite of forces drives gene traffic from *Drosophila* X chromosomes. *Genome Biol Evol.* 1:176–188.
Mueller JL, Mahadevaiah SK, Park PJ, Warburton PE, Page DC, Turner JMA. 2008. The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. *Nat Genet.* 40:794–799.
Munoz-Torres MC, Reese JT, Childers CP, Bennett AK, Sundaram JP, Childs KL, Anzola JM, Milshina N, Elsik CG. 2011. Hymenoptera Genome Database: integrated community resources for insect species of the order Hymenoptera. *Nucleic Acids Res.* 39:D658–D662.
Nene V, Wortman JR, Lawson D, et al. (92 co-authors). 2007. Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science* 316:1718–1723.
Ostlund G, Schmitt T, Forslund K, Kostler T, Messina DN, Roopra S, Frings O, Sonnhammer ELL. 2010. InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res.* 38:D196–D203.
Parisi M, Nuttall R, Naiman D, Bouffard G, Malley J, Andrews J, Eastman S, Oliver B. 2003. Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. *Science* 299:697–700.
Potrzebowski L, Vinckenbosch N, Marques AC, Chalmel F, Jegou B, Kaessmann H. 2008. Chromosomal gene movements reflect the recent origin and biology of therian sex chromosomes. *PLoS Biol.* 6:e80.
Prince EG, Kirkland D, Demuth JP. 2010. Hyperexpression of the X chromosome in both sexes results in extensive female bias of X-linked genes in the flour beetle. *Genome Biol Evol.* 2:336–346.
Rai KS, Black WC. 1999. Mosquito genomes: structure, organization, and evolution. *Adv Genet.* 41:1–33.
Rice WR. 1987. The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. *Evolution* 41:911–914.
Rice WR. 1996. Evolution of the Y sex chromosome in animals. *Bioscience* 46:331–343.
Rinn JL, Snyder M. 2005. Sexual dimorphism in mammalian gene expression. *Trends Genet.* 21:298–305.
Ross JA, Urton JR, Boland J, Shapiro MD, Peichel CL. 2009. Turnover of sex chromosomes in the stickleback fishes (*Gasterosteidae*). *PLoS Genet.* 5:e1000391.
Ross MT, Grafham DV, Coffey AJ, et al. (279 co-authors). 2005. The DNA sequence of the human X chromosome. *Nature* 434:325–337.
Shimomura M, Minami H, Suetsugu Y, et al. (12 co-authors). 2009. KAIKObase: an integrated silkworm genome database and data mining tool. *BMC Genomics* 10:486.
Singh ND, Davis JC, Petrov DA. 2005. X-linked genes evolve higher codon bias in *Drosophila* and Caenorhabditis. *Genetics* 171:145–115.
Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, et al. (37 co-authors). 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423:825–837.
Storchova R, Divina P. 2006. Nonrandom representation of sex-biased genes on chicken Z chromosome. *J Mol Evol.* 63:676–681.
Sturgill D, Zhang Y, Parisi M, Oliver B. 2007. Demasculinization of X chromosomes in the *Drosophila* genus. *Nature* 450:238–241.
Toups MA, Hahn MW. 2010. Retrogenes reveal the direction of sex-chromosome evolution in mosquitoes. *Genetics* 186:763–766.

Toups MA, Pease JB, Hahn MW. 2011. No excess gene movement is detected off the avian or lepidopteran Z chromosome. *Genome Biol Evol.* 3:1381–1390.

Tweedie S, Ashburner M, Falls K, et al. (8 co-authors). 2009. FlyBase: enhancing *Drosophila* Gene Ontology annotations. *Nucleic Acids Res.* 37:D555–D559.

Ueda S, Nakai S, Nishida Y, Hisajima H, Honjo T. 1982. Long terminal repeat-like elements flank a human immunoglobulin epsilon pseudogene that lacks introns. *EMBO J.* 1:1539–1544.

van Dongen S. 2000. A cluster algorithm for graphs: Centrum Wiskunde & Informatica.

Vibranovski MD, Zhang Y, Long M. 2009. General gene movement off the X chromosome in the *Drosophila* genus. *Genome Res.* 19:897–903.

Vicoso B, Charlesworth B. 2006. Evolution on the X chromosome: unusual patterns and processes. *Nat Rev Genet.* 7:645–653.

Vinckenbosch N, Dupanloup I, Kaessmann H. 2006. Evolutionary fate of retroposed gene copies in the human genome. *Proc Natl Acad Sci U S A.* 103:3220–3225.

Walters JR, Hardcastle TJ. 2011. Getting a full dose? Reconsidering sex chromosome dosage compensation in the silkworm, *Bombyx mori*. *Genome Biol Evol.* doi:10.1093/gbe/evr1036

Wiegmann BM, Trautwein MD, Kim J-W, Cassel BK, Bertone MA, Winterton SL, Yeates DK. 2009. Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. *BMC Biol.* 7:34.

Wilson MA, Makova KD. 2009. Genomic analyses of sex chromosome evolution. *Annu Rev Genomics Hum Genet.* 10:333–354.

Woram RA, Gharbi K, Sakamoto T, et al. (13 co-authors). 2003. Comparative genome analysis of the primary sex-determining locus in salmonid fishes. *Genome Res.* 13:272–280.

Wright F. 1990. The 'effective number of codons' used in a gene. *Gene* 87:23–29.

Zdobnov EM, von Mering C, Letunic I, et al. (33 co-authors). 2002. Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science* 298:149–159.

Zha X, Xia Q, Duan J, Wang C, He N, Xiang Z. 2009. Dosage analysis of Z chromosome genes using microarray in silkworm, *Bombyx mori*. *Insect Biochem Mol Biol.* 39:315–321.