Evolution, 2023, **77(11)**, 2504–2511
https://doi.org/10.1093/evolut/qpad169
Advance access publication 20 September 2023
Brief Communication

The X chromosome of insects likely predates the origin of class Insecta

Melissa A. Toups[1] and Beatriz Vicoso[2]

[1]Department of Life and Environmental Sciences, Bournemouth University, Poole, United Kingdom
[2]Institute of Science and Technology Austria, Klosterneuburg, Austria

Corresponding author: Department of Life and Environmental Sciences, Bournemouth University, Poole BH12 5BB, United Kingdom.
Email: mtoups@bournemouth.ac.uk

Abstract

Sex chromosomes have evolved independently multiple times, but why some are conserved for more than 100 million years whereas others turnover rapidly remains an open question. Here, we examine the homology of sex chromosomes across nine orders of insects, plus the outgroup springtails. We find that the X chromosome is likely homologous across insects and springtails; the only exception is in the Lepidoptera, which has lost the X and now has a ZZ/ZW sex-chromosome system. These results suggest the ancestral insect X chromosome has persisted for more than 450 million years—the oldest known sex chromosome to date. Further, we propose that the shrinking of gene content the dipteran X chromosome has allowed for a burst of sex-chromosome turnover that is absent from other speciose insect orders.

Keywords: chromosomal evolution, evolutionary genomics, sex chromosomes

Introduction

Although sexual reproduction was already present in the ancestor of all animals, the specific mechanisms controlling the development of males and females differs widely between clades. Often, sex chromosomes such as the X and Y of mammals carry the master sex-determination genes (Bachtrog et al., 2014). Despite sharing broad characteristics (e.g., the Y is typically gene poor and heterochromatic), these chromosomes evolved from autosomes independently multiple times across animals, and the processes shaping their convergent differentiation have been studied extensively. While young and undifferentiated sex chromosomes often undergo turnover (i.e., they revert to autosomes, as another chromosome pair takes on the role of sex determination), fully-differentiated sex chromosomes are thought to be extremely stable (Pokorná & Kratochvíl, 2009). This view of differentiated sex chromosomes as an "evolutionary trap" was largely shaped by model organisms such as eutherian mammals (Pokorná & Kratochvíl, 2009; Vicoso, 2019), which have all shared the same sex chromosomes for at least 166 million years (Veyrunes et al., 2008). What drives some sex chromosomes to turnover, while others differentiate fully and are maintained over long periods is still unclear. How long sex chromosomes can be conserved for, and what eventually leads to their loss, are also still open questions.

Insects are the most speciose and evolutionary successful animals on the planet, and have colonized almost all terrestrial and freshwater ecosystems (Blackmon et al., 2017; Misof et al., 2014; Stork, 2018). The origins of insects is dated as far back as 441 MYA (Misof et al., 2014), either concurring with or shortly after the evolution of land plants (Misof et al., 2014; Morris et al., 2018). Insects use a great diversity of genetic sex-determining mechanisms, including the more evolutionary common systems of XY and ZW sex chromosomes, as well as more rare genetic sex-determining systems such as a paternal genome elimination (PGE) and haplodiploidy (Blackmon et al., 2017). Many orders also have a preponderance of X0 systems, which are relatively rare in other groups of organisms (Bachtrog et al., 2014), and may have been the ancestral state of the clade (Blackmon et al., 2017). Alternatively, these X0 systems may be the result of independent loss of the male-specific Y, which can occur in ancient sex chromosomes (Bachtrog et al., 2014; Blackmon & Demuth, 2014).

Whether this diversity in sex-determining systems reflects the independent gain and loss of sex-linked chromosomes, or instead the conservation (and modification) of an ancestral pair of chromosomes, similar to the mammalian system, has been a longstanding question. An early comparison of *Drosophila melanogaster* (Diptera), *Bombyx mori* (Lepidoptera), and *Tribolium castaneum* (Coleoptera) came to the conclusion that different chromosomal elements have been coopted in these different groups (Pease & Hahn, 2012). However, the systematic sequencing of species of various insect orders has shifted this view. First, studies that compared multiple species within orders have detected conservation of the X-linked gene content in Hemiptera (Mathers et al., 2021; Pal & Vicoso, 2015), Orthoptera (Li et al., 2022), and Coleoptera (Bracewell et al., 2023), despite the diversity of sex-determination systems found in these groups (including X0, XY, multiple X chromosomes, but also PGE and haplodiploidy (Blackmon et al., 2017)). The only currently known

Received April 24, 2023; revisions received September 4, 2023; accepted September 18, 2023
Associate Editor: Richard Meisel; Handling Editor: Jason Wolf

exception to the conservation of an X chromosome within an insect order is Diptera, where sex-chromosome turnover from the ancestral element F has been documented extensively involving many different chromosomal elements of the genome (Vicoso & Bachtrog, 2015).

Recent studies comparing the gene content of different orders have also suggested that conservation of the X chromosome may occur over longer periods of time. Homology with the ancestral X chromosome of Diptera (element F) has been detected in Blattodea (Meisel et al., 2019), Odonata (Chauhan et al., 2021), Coleoptera (Chauhan et al., 2021; Li et al., 2022), Hemiptera (Li et al., 2022), and Orthoptera (Li et al., 2022), raising the possibility that the X chromosome has been conserved since the split of Paleopterans and Neopterans 365 million years ago. However, given the sparse distribution of data across the insect phylogeny, and the small number of chromosomes in some clades, it has been unclear if this is simply recruitment of shared gene content in different orders, recruitment of the same chromosome as a sex chromosome multiple times, or a shared ancestral sex chromosome (Chauhan et al., 2021; Li et al., 2022). For instance, in the best studied insect order, Diptera, conservation of the X chromosome was originally assumed, when the same chromosomal element (called "Muller element A") was found to correspond to the X of the fruit fly *D. melanogaster* and the mosquito *A. gambiae*, the first two species to be sequenced (Zdobnov et al., 2002). The sampling of more clades showed that a different chromosomal element ("element F") was ancestrally the X, and that element A was secondarily acquired as the X independently by both fruit flies and mosquitoes (Vicoso & Bachtrog, 2015).

Here, we analyze published genome assemblies from nine orders of insects, and the outgroup Collembola, to determine if the X chromosome originated once or multiple times independently. We find that the X chromosome is homologous across eight orders of insects and the outgroup, Collembola. The only exception to this pattern, the lepidopteran Z chromosome, is not homologous to the X chromosome shared across insects, and the transition to female heterogamety was accompanied by a sex-chromosome turnover. We find that the apparent homology of the X chromosomes across multiple insect orders is more likely the result of ancestry rather than independent recruitment of shared genes. Further, we discuss the evolution of shared gene content of the X chromosome in different orders of insect across the phylogeny, and propose that the shrinking of gene content in the dipteran X chromosome may have led to its turnover, giving rise to the tremendous diversity of XY systems in this group.

Methods

Acquisition of genome assemblies and annotations

All species and their sex-chromosome systems used in this study are listed in Table 1. We inferred the sex-chromosome systems for each species using the Tree of Sex database (The Tree of Sex Consortium et al., 2014).

Chromosome-level genome assemblies and annotations were downloaded from the National Center for Biotechnology Information (NCBI) website for seven orders of class Insecta, including the blue-tailed damselfly (Insecta: Odonata), *Ischnura elegans* (Price et al., 2022), the American grasshopper (Insecta: Orthoptera), *Schistocerca americana* (Childers et al., 2021), the pea aphid (Insecta: Hemiptera), *Acyrthosiphon pisum* (Li et al., 2019), the Australian sheep blowfly (Insecta: Diptera), *Lucilia cuprina* (https://www.ncbi.nlm.nih.gov/assembly/GCF_022045245.1), the silkworm (Insecta: Lepidoptera), *Bombyx mori* (https://www.ncbi.nlm.nih.gov/assembly/GCF_014905235.1), the seven-spotted ladybird (Insecta: Coleoptera), *Coccinella septempunctata* (Crowley et al., 2021b), and the common green lacewing (Insecta: Neuroptera), *Chrysoperla carnea* (Crowley et al., 2021a). Details of the downloaded chromosome-level assemblies are available in Supplementary Table S1. Furthermore, we downloaded X and autosome assignments of previously published transcripts for the Cristina's Timema (Insecta: Phasmatodea), *Timema cristinae* (Parker et al., 2022).

Additionally, we downloaded chromosome-level assemblies of the common yellow sally (Insecta: Plecoptera), *Isoperla grammatica* (McSwan et al., 2023), and two springtails (Entognatha: Collembola), *Sinella curviseta* (Zhang et al., 2019) and *Allacma fusca* (Jaron et al., 2023), all of which lack accompanying gene annotations. We therefore downloaded RNA-sequencing reads from the SRA (Sequence Read Archive) on NCBI (Supplementary Table S1) in order to construct gene annotations for these species. Reads were first trimmed using Trimmomatic v0.39 (Bolger et al., 2014), aligned with HISAT2 (Kim et al., 2019), and resulting samfiles were converted to sorted bamfiles using Samtools v.1.16 (Li et al., 2009). GTF

Table 1. Species analyzed in this study and their sex-chromosome system. The male karyotype is shown as: total chromosome number (haploid number of autosomes + male sex-chromosome complement).

Order	Species	Common name	Male karyotype
Plecoptera	*Isoperla grammatica*	Common yellow sally	26(12A + X_1X_2)
Orthoptera	*Schistocerca americana*	American grasshopper	23(11A + X)
Phasmatodea	*Timema cristinae*	Cristina's Timema	25(12A + X)
Lepidoptera	*Bombyx mori*	silkworm	56(27A + ZZ)
Diptera	*Lucilia cuprina*	Australian sheep blowfly	12(5A + XY)
Coleoptera	*Coccinella septempunctata*	Seven-spotted ladybird	20(9A + XY)
Neuroptera	*Chrysoperla carnea*	Common green lacewing	12(5A + XY)
Hemiptera	*Acyrthosiphon pisum*	Pea aphid	7(3A + X)
Odonata	*Ischnura elegans*	Blue-tailed damselfly	27(13A + X)
Collembola	*Sinella curviseta*	Slender springtail	11(5A + X)
Collembola	*Allacma fusca*	Globular springtail	10(4A + X_1X_2)

files were constructed by StringTie v.2.2.1 (Pertea et al., 2015). Coding sequences were extracted and longest ORFs were identified using Transdecoder v.5.5.0 (Haas, BJ. https://github.com/TransDecoder/TransDecoder). We then extracted the longest isoforms using a custom perl script.

While two of these species, *Allacma fusca* and *Isoperla grammatica*, had identified X chromosomes in their assembly, *Sinella curviseta* did not. To identify the X chromosome in this species, we downloaded paired-end Illumina WGS reads from the SRA from a pool of individuals including both males and females. Reads were aligned using default parameters in Bowtie2 v.2.4.4 (Langmead & Salzberg, 2012), and uniquely mapping reads were extracted. We then used soap.coverage (version 2.7.7, https://github.com/gigascience/bgi-soap2/tree/master/tools/soap.coverage/2.7.7) to estimate coverage. The scaffold CM023202.2 had significantly lower coverage than the other five, consistent with the prediction for the X chromosome in a mixed-sex pooled sample (Supplementary Figure S1). We therefore refer to scaffold CM023202.2 as the X chromosome throughout the manuscript, and rename the other scaffolds as listed in Supplementary Table S2.

Detecting orthology to the blue-tailed damselfly and the slender springtail

We selected the blue-tailed damselfly, *I. elegans*, as our reference species because it is the sister lineage to all other insects considered. We therefore refer to *I. elegans* as our reference species, and all other species as our focal species in analyses. However, we also performed all the analyses using the springtail *S. curviseta* as reference, and our result remains largely unchanged.

As a first step, for the seven species with published gene annotations, we processed the translated.cds files and gff files from NCBI (Supplementary Table S1) through the R package GENESPACE (Lovell et al., 2022), which produces simplified gff files and peptide sequences. For the remaining species with chromosome assemblies but no gene annotation—*A. fusca*, *S. curviseta*, and *I. grammatica*—we used the longest isoform peptide sequence and modified gtf output from Stringtie, and processed these through GENESPACE as well. For *T. cristinae*, we processed the gff file using the script getAnnoFastaFromJoingenes.py from the Augustus software package (Stanke et al., 2008), and then selected the longest isoform using a custom perl script.

We then used a reciprocal blast approach to define 1-to-1 orthologs between each of our focal species and *I. elegans*, using the processed peptide files from GENESPACE as input. To combine fragmented blast results into a single hit, we used the script blast_outfmt6_group_segments.pl from the Trinity software package (Grabherr et al., 2011). The number of 1:1 orthologs (reciprocal best hits) for comparisons with both *I. elegans* as a reference and *S. curviseta* as a reference are listed in Table 2 and Supplementary Table S3, respectively. We then counted the number of genes observed as on the X (or Z) in both the focal species and the reference species. To compute the expected number of shared genes for the X chromosome in the reference species and each chromosome in the focal species, we first computed the proportion of 1:1 orthologs on each chromosome in the focal species. This proportion was then multiplied by the number of X-linked 1:1 orthologs in *I. elegans*. Significance for homology between the X chromosome of the reference species, *I. elegans*, and all chromosomes of the focal species was initially computed using a binomial test, with a Bonferroni correction to adjust for multiple tests within a species. Further, we simulated enrichment of shared X-linked gene content between two species harboring an ancestral X chromosome, allowing for varying proportions of genes to move across the genome (Supplementary Figure S2a; Supplemental Methods). We did not detect a twofold enrichment of shared X-linked genes under the null model of free gene movement when simulating the same number of genes as found on the X chromosome and autosomes of *I. elegans* (Supplementary Figure S2b; Supplemental Methods), and therefore used this as an additional threshold to denote homology between chromosomes. Similarly, we performed the same analyses using *S. curviseta* as a reference species, using 1:1 orthologs between *S. curviseta* and the focal species. The only modification to this analysis was a reduced enrichment threshold of 1.5fold, as springtails are more distantly related to other taxa. Importantly, this also exceeds the threshold of enrichment of our simulations using parameters from *I. elegans*. As *S. curviseta* has a larger X chromosome than *I. elegans* (and therefore our simulations), it is expected that the 1.5fold threshold is conservative.

Additionally, we assessed statistical significance between X chromosomes of different species using a Monte Carlo approach. Specifically, we randomized chromosome assignments in each comparison for each 1:1 ortholog 100,000 times. The

Table 2. Excess of shared gene content between *I. elegans* and the X/Z of the focal species.

Order	Focal species	1:1 Orthologs	Chr	Expected shared X	Observed shared X	Exp/obs	*p*-value
Plectoptera	*I. grammatica*	3458	X_1	9	56	6.2	$<10^{-5}$
Plectoptera	*I. grammatica*	3458	X_2	10	27	2.7	$<10^{-5}$
Orthoptera	*S. americana*	5995	X	47	186	4.0	$<10^{-5}$
Phasmatodea	*T. cristinae*	2166	X	7	33	4.7	$<10^{-5}$
Lepidoptera	*B. mori*	4625	Z	13	3	0.2	.9999
Diptera	*L. cuprina*	4076	X	1	7	7.0	3×10^{-5}
Coleoptera	*C. septempunctata*	4904	X	23	100	4.3	$<10^{-5}$
Neuroptera	*C. carnea*	5147	X	23	94	4.1	$<10^{-5}$
Hemiptera	*A. pisum*	4320	X	33	86	2.6	$<10^{-5}$
Collembola	*S. curviseta*	3570	X	46	116	2.5	$<10^{-5}$
Collembola	*A. fusca*	3318	X_1	56	109	1.9	$<10^{-5}$
Collembola	*A. fusca*	3318	X_2	36	35	1.0	.624

resulting *p*-values are the proportion of randomizations in which the number of shared orthologs in the randomized dataset are greater or equal to the observed number of shared X-linked (Z-linked) genes. Where $p < 10^{-5}$, all simulated datasets have fewer shared orthologs than the observed data.

Detecting fusions and fissions to sex chromosomes across the insect phylogeny

Fusions and fissions are common in sex-chromosome evolution (Bachtrog et al., 2014; Pennell et al., 2015), and may have played an important role in insects. In order to try to detect these events across the phylogeny, we then identified the regions of the genome in our reference species (*I. elegans* and *S. curviseta*) which were homologous to the X chromosome of the focal species. Specifically, we performed the reciprocal analysis as described above, where we conducted binomial tests to identify all chromosomes in the reference species that were homologous to the X chromosome of the focal species, and adjusted for multiple tests using a Bonferroni correction. We again defined excess as twofold using *I. elegans*, and 1.5fold using *S. curviseta*.

Synteny detection in insect genomes

Finally, we selected four insect species to examine synteny across the phylogeny: *S. americana*, *C. septempunctata*, *C. carnea*, and *I. elegans*, spanning ~400 million years of evolution. We also examined pairwise synteny for all focal species vs *I. elegans* as a reference. Synteny maps were constructed using GENESPACE v0.94 (Lovell et al., 2022).

Results and discussion

The X chromosome is conserved across insects

We first systematically tested whether the same sex chromosome was used throughout insects. To do so, we compared the gene content of the X of clades for which homology to the dipteran X had been detected in previous studies (Odonata, Coleoptera, Orthoptera, Hemiptera, and Diptera itself; Chauhan et al., 2021; Li et al., 2022) as well as four additional orders (Phasmatodea, Neuroptera, Plecoptera, and the ZW Lepidoptera). All but one genome assembly were chromosome-level and had annotated sex chromosomes. For the phasmatodean *T. cristinae*, only scaffolds were available, but these were previously assigned as X-linked or autosomal based on their male and female genomic coverage (Parker et al., 2022). We used the damselfly *I. elegans* (Odonata) as the reference species, as it is sister to all other insect lineages considered. In all seven insect orders with heterogametic XY or XO sex chromosomes, we detected homology with the X chromosome of the damselfly, *I. elegans* ($p < .0001$ in every case, Monte Carlo randomization, Figure 1; Table 2; Supplementary Figure S3). Further, in the order Plecoptera, *I. grammatica* has two X chromosomes (X_1 and X_2), both of which are homologous to the *I. elegans* X chromosome. Importantly, the lepidopteran Z chromosome shows no homology to the *I. elegans* X chromosome (Figure 1; Table 2; Supplementary Figure S3), which is consistent with previous results (Chauhan et al., 2021), and suggests that the switch in heterogamety that occurred in this lineage was associated with a turnover in sex chromosome. Finally, in *I. grammatica*, *S. americana*, *C. septempunctata*, and *C. carnea*, an autosome was also homologous to the X chromosome of *I. elegans*, though in all cases it had a smaller excess of shared gene content than the X chromosome (Figure 1; Supplementary Table S4).

We also implemented a synteny analysis using GENESPACE. We were able to detect conservation of the X chromosome across the phylogeny when using four species representing ~400 million years of evolution: *I. elegans*, *S. americana*, *C. carnea*, and *C septempunctata* (Supplementary Figure S5). Importantly, odonates have the lowest known rates of fissions and fusions among insects (Alfieri et al., 2023). Coleopterans and orthopterans also have relatively few fissions and fusions, compared to other insects orders, such as dipterans and lepidopterans (Alfieri et al., 2023). In this light, it perhaps unsurprising we detected synteny only when using these data. We are unable to detect synteny when all species are used, as few syntenic blocks are shared by all species, perhaps due to increased rates of rearrangements in some insect orders. Therefore, we also analyzed synteny in a pairwise fashion between all focal species and *I. elegans*. With the exception of the dipteran X chromosome, we detected synteny between the X chromosome of all focal species and the *I. elegans* X chromosome (Supplementary Figure S6).

The ancestral insect X chromosome is shared with springtails (Entognatha: Collembola)

We then investigated whether the origin of the X may have predated the appearance of insects altogether. To do so, we compared the *I. elegans* genome to the closest outgroup of insects, the springtails (Collembola). We selected two species with genome assemblies, *Sinella curviseta* and *Allacma fusca*, which are at least 193 million years diverged from each other (Kumar et al., 2017; Yu et al., 2021). We then downloaded RNA-seq reads for both species, as neither genome had gene annotations, and constructed transcriptome assemblies to use in downstream analyses. The *A. fusca* annotation has two X chromosomes, X_1 and X_2 (Jaron et al., 2022, 2023). In contrast, no information is available as to what chromosome is the X in *S. curviseta*. We therefore downloaded pooled male and female genomic DNA from the SRA, and performed a coverage analysis to identify the X chromosome. We detected homology between the X chromosome of *I. elegans* and the inferred X of *S. curviseta*, as well as the X_1 chromosome of *A. fusca* (Table 1; Figure 1), supporting a more ancient origin of the X than of insects.

Potential fusion of *I. elegans* chromosome 2 to the X in distant lineages

It has been suggested that it may be selectively beneficial to recruit specific chromosomes or genomic regions as sex chromosomes, if these regions harbor genes involved in sex determination or a preexisting excess of sexually antagonistic gene content (Anderson et al., 2020; Charlesworth & Charlesworth, 1980; Jeffries et al., 2018; Marshall Graves & Peichel, 2010; Toups et al., 2019). To investigate this, we performed the reciprocal homology analysis to that described above, that is, we asked which autosomes of *I. elegans* became X-linked in each of the other lineages. We then tested if one or more autosomes were recruited onto the X more often than expected by chance. We find that chromosome 2 in *I. elegans* is sex-linked in the phylogenetically distantly related taxa of *I. grammatica, T. cristinae, L. cuprina*, and *C. carnea* (Supplementary Table S4). If chromosome 2 became sex-linked independently through fusion events in these lineages, then it did more than expected by chance within insects

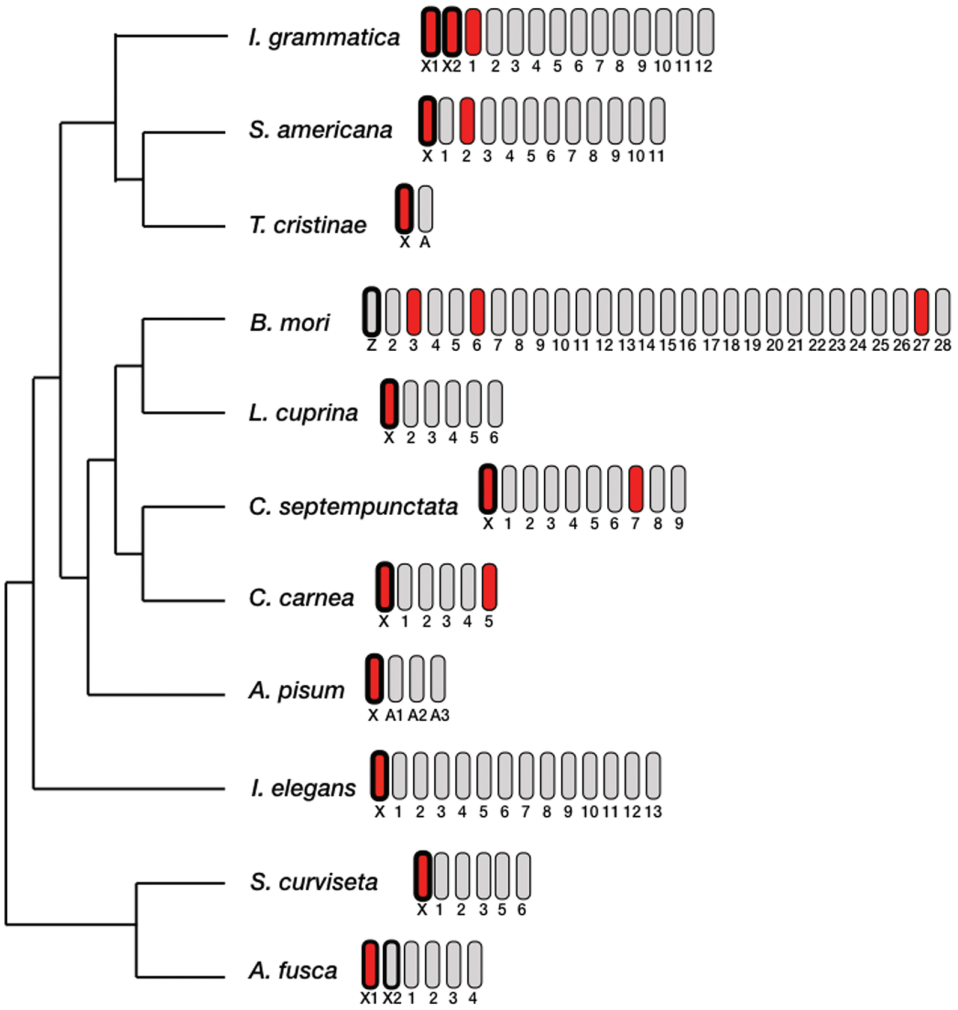

Figure 1. Homology of X chromosomes across insects. Phylogeny is adapted from Blackmon et al. (2017). Idiograms for each species are shown, with chromosomes homologous to the *I. elegans* X chromosome colored red, and those not homologous colored gray. Sex chromosomes for each species are outlined in black.

(binomial test, $p < .05$). However, if we extend our analysis to include Collembola, it is no longer statistically significant (binomial test, $p = .07282$). While it is tempting to speculate that the sex-linkage of (a part of) chromosome 2 may be advantageous, it is also possible the region of the genome corresponding to *I. elegans* chromosome 2 was incorporated into the X chromosome of insects and subsequently lost in *C. septempunctata*, *A. pisum*, and *S. americana*. Finally, rearrangements such as fissions and fusions occur in Coleoptera (Alfieri et al., 2023; Bracewell et al., 2023) and Hemiptera (Alfieri et al., 2023; Mathers et al., 2021), and the above result may simply occur because of the taxa chosen within these orders. More extensive sampling would be needed to distinguish between these three hypotheses. Interestingly, the X chromosome of the slender springtail, *S. curviseta*, and the X_1 chromosome of the globular springtail, *A. fusca*, is homologous to chromosome 7 in *I. elegans*, in addition to the X chromosome. While we cannot in this case distinguish between a fusion of the X and autosome 7 in Collembola or a loss of part of the ancestral hexapod X (now corresponding to chromosome 7) in the common ancestor of insects, this further emphasizes the long-term conservation of gene content within Collembola, and arthropods in general.

Few genes shared among all Xs

Figure 1 suggests that while the ancestral X chromosome has remained sex-linked throughout insect history, substantial chromosomal reshuffling has occurred, as multiple chromosomes correspond to the *I. elegans* X in the other species. Even if the same chromosome is used as the X in all insect orders, extensive gene movement within and between chromosomes, combined with rare translocations, could lead to different gene sets remaining X-linked throughout the insect phylogeny. To investigate this, we compared the gene content of the X of the different orders. We identified a total of 2,253 1:1 orthologs between the Plecoptera, Orthoptera, Diptera, Coleoptera, Neuroptera, Hemiptera, Odonata, and Collembola. Of these, only 1 was shared across the entire phylogeny on the X chromosome (Figure 2). This is largely because the X chromosome in Diptera is considerably smaller than the other orders, with only nine 1:1 orthologs present on the X (out of 207 total X genes). When Diptera is removed from the analysis, 14 genes are shared between the X chromosomes of insects and springtails. Importantly, the X chromosome of all orders are homologous with the Collembola X chromosome (Supplementary Figure S4; Supplementary Tables S3 and S5). Despite these constraints, it is clear that even when the same chromosome

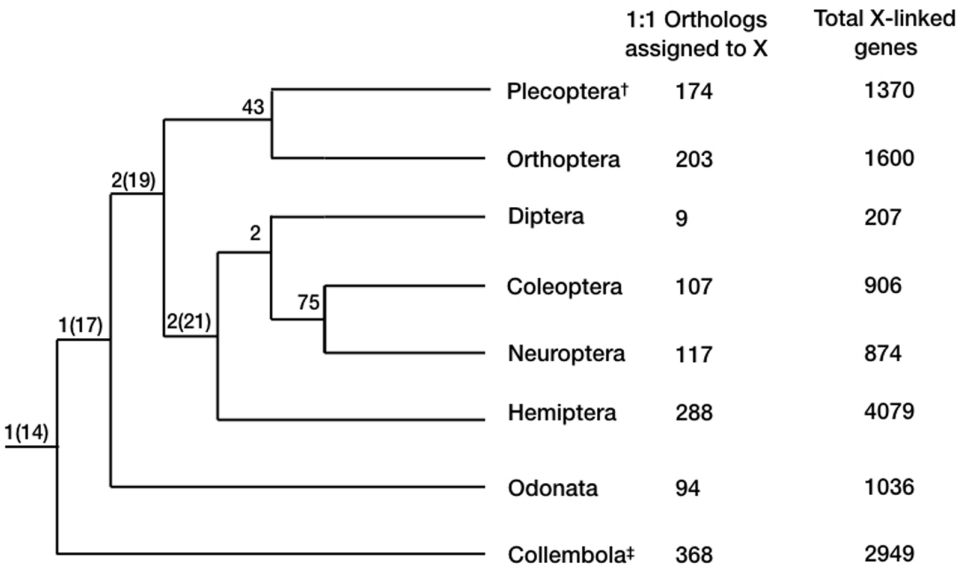

Figure 2. Shared gene content across the insect and springtail phylogeny. Values in parentheses are computed excluding the dipteran X chromosome. †Values for Plecoptera are combined for X_1 and X_2, as both are homologous to the *I. elegans* X chromosome. ‡Values for Collembola only consider *S. curviseta*.

is used as the X over long periods of time, the majority of X-linked genes are not necessarily shared.

Finally, it is clear from Figure 2 that the small number of genes found on the dipteran X chromosome is a derived feature, which may help explain a paradoxical observation: how does an X chromosome that has been conserved since the origin of insects suddenly start undergoing turnover repeatedly in flies and mosquitoes? Such turnover is thought to be prevented by the damaging consequences of putting a highly specialized sex chromosome in an autosomal context (for instance, this could lead to dosage compensation being active despite no longer being needed). The resulting fitness effect is likely proportional to the number of genes affected, and the reduction in gene content of the dipteran X may have been a prerequisite for its reversal to an autosome.

Conservation of the X chromosome, or convergent recruitment of X-linked genes?

An alternative hypothesis to explain our result is the convergent recruitment of genes to nonhomologous X chromosomes across insect orders. While we cannot fully exclude this possibility, it is worth considering what parameter space could lead to the extent of shared X-linked genes that we observe. To explore this, we performed simulations of X chromosome conservation and convergence (Supplementary Methods and Supplementary Figure S7). In both, we started with 13 autosomes and a single X chromosome, as is found in *I. elegans*. Under conservation of the same X, our observed/expected ratio of two–sixfold was recovered when between 0.4 and 0.7 of the genes had moved in each lineage. Under convergent recruitment of genes to nonhomologous X chromosomes, such a range was only recovered when a large proportion of the genome was under selection to move to the X chromosome (at least 10%, i.e., over 1,000 genes), and a substantial proportion of these selected genes actually moved to the X in each lineage (i.e., selection was strong). Such consistently strong selection of a large proportion of the genome over a large timeframe (~450 million years) seems unlikely, and, the independent recruitment of the same genes to different X chromosomes less parsimonious than a single origin of the X. We therefore argue that conservation of the X chromosome, with gene movement, is the more likely explanation.

Conclusions

We present evidence that the X chromosome is shared among at least eight insect orders and the outgroup Collembola, and originated prior to the evolution of Class Insecta itself. This is an addition to the German cockroach, *Blatella germanica* (Insecta: Blattodea), whose X is homologous to element F in dipterans (Meisel et al., 2019), which we omitted from this analysis because there are currently no chromosome-level genome assemblies. This suggests that the insect X chromosome is at least 450 MY old (CI: 425.4–478.1), which is the most ancient sex chromosome identified to date. Further, we suggest that the shrinking X chromosome content of the dipteran X chromosome allowed it to escape the "evolutionary trap" of highly differentiated sex chromosomes, which led to high rates of sex-chromosome turnover among fly lineages (Vicoso & Bachtrog, 2015).

Supplementary material

Supplementary material is available online at *Evolution*.

Data availability

We analyzed only publicly available data, and information about where the data are archived is in Supplementary Table S1. Additional datafiles and pipeline are available on Dryad (https://doi.org/10.5061/dryad.hx3ffbgkt) and custom scripts for data analysis are available on Zenodo (https://doi.org/10.5281/zenodo.8138705). Code for simulations are available at (https://git.ista.ac.at/bvicoso/veryoldx).

Author contributions

M.A.T. and B.V. conceived and designed the study. M.A.T. performed the analyses and drafted the manuscript. B.V.

designed the simulations. M.A.T. and B.V. revised the manuscript. Editorial processing of the manuscript was done independently of B.V., who is an associate editor of *Evolution*.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

All computational analyses were performed on the server at Institute of Science and Technology Austria. We thank Marwan Elkrewi and Vincent Bett for analytical advice, and Tanja Schwander and Vincent Merel for useful discussions. We also thank Matthew Hahn for comments on an earlier version of the manuscript.

References

Alfieri, J. M., Jonika, M. M., Dulin, J. N., & Blackmon, H. (2023). Tempo and mode of genome structure evolution in insects. *Genes*, 14(2), 336. https://doi.org/10.3390/genes14020336

Anderson, N. W., Hjelmen, C. E., & Blackmon, H. (2020). The probability of fusions joining sex chromosomes and autosomes. *Biology Letters*, 16(11), 20200648. https://doi.org/10.1098/rsbl.2020.0648

Ashman, T-L., Bachtrog, D., Blackmon, H., Goldberg, E. E., Hahn, M. W., Kirkpatrick, M., Kitano, J., Mank, J. E., Mayrose, I., Ming, R., Otto, S. P., Peichel, C. L., Pennell, M. W., Perrin, N., Ross, L., Valenzuela, N., & Vamosi, J. C.; The Tree of Sex Consortium (2014). Tree of sex: A database of sexual systems. *Scientific Data*, 1, 140015. https://doi.org/10.1038/sdata.2014.15

Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M, Otto, S. P., Ashman, T. -L., Hahn, M. W., Kitano, J., Mayrose, I., Ming, R., Perrin, N., Ross, L., Valenzuela, N., & Vamosi, J. C.; The Tree of Sex Consortium. (2014). Sex determination: Why so many ways of doing it? *PLoS Biology*, 12(7), e1001899. https://doi.org/10.1371/journal.pbio.1001899.

Blackmon, H., & Demuth, J. P. (2014). Estimating tempo and mode of Y chromosome turnover: Explaining Y chromosome loss with the Fragile Y Hypothesis. *Genetics*, 197(2), 561–572. https://doi.org/10.1534/genetics.114.164269

Blackmon, H., Ross, L., & Bachtrog, D. (2017). Sex determination, sex chromosomes, and karyotype evolution in insects. *The Journal of Heredity*, 108(1), 78–93. https://doi.org/10.1093/jhered/esw047

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bracewell, R., Tran, A., Chatla, K., & Bachtrog, D. (2023). Sex chromosome evolution in beetles. *bioRXiv*. https://doi.org/10.1101/2023.01.18.524646

Charlesworth, D., & Charlesworth, B. (1980). Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. *Genetical Research*, 35(2), 205–214. https://doi.org/10.1017/s0016672300014051

Chauhan, P., Swaegers, J., Sánchez-Guillén, R. A., Svensson, E. I., Wellenreuther, M., & Hansson, B. (2021). Genome assembly, sex-biased gene expression and dosage compensation in the damselfly *Ischnura elegans*. *Genomics*, 113(4), 1828–1837. https://doi.org/10.1016/j.ygeno.2021.04.003

Childers, A. K., Geib, S. M., Sim, S. B., Poelchau, M. F., Coates, B. S., Simmonds, T. J., Scully, E. D., Smith, T. P. L., Childers, C. P., Corpuz, R. L., Hackett, K., & Scheffler, B. (2021). The USDA-ARS Ag100Pest Initiative: High-quality genome assemblies for agricultural pest arthropod research. *Insects*, 12(7), 626. https://doi.org/10.3390/insects12070626

Crowley, L.; University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, and Darwin Tree of Life Consortium. (2021b). The genome sequence of the seven-spotted ladybird, *Coccinella septempunctata* Linnaeus, 1758. *Wellcome Open Research*, 6, 319. https://doi.org/10.12688/wellcomeopenres.17346.1

Crowley, L. M.; University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, and Darwin Tree of Life Consortium. (2021a). The genome sequence of the common green lacewing, *Chrysoperla carnea* (Stephens, 1836). *Wellcome Open Research*, 6, 334. https://doi.org/10.12688/wellcomeopenres.17455.1

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), 644–652. https://doi.org/10.1038/nbt.1883

Jaron, K. S., Berg, M. P., Ellers, J., Hodson, C. N., & Ross, L.; University of Oxford and Wytham Woods Genome Acquisition Lab. (2023). The genome sequence of the springtail Allacma fusca (Linnaeus, 1758). *Wellcome Open Research*, 8, 319. https://doi.org/10.12688/wellcomeopenres.19690.1

Jaron, K. S., Hodson, C. N., Ellers, J., Baird, S. J. E., & Ross, L. (2022). Genomic evidence of paternal genome elimination in the globular springtail *Allacma fusca*. *Genetics*, 222(3), iyac117. https://doi.org/10.1093/genetics/iyac117

Jeffries, D. L., Lavanchy, G., Sermier, R., Sredl, M. J., Miura, I., Borzée, A., Barrow, L. N., Canestrelli, D., Crochet, P. -A., Dufresnes, C., Fu, J., Ma, W. -J., Garcia, C. M., Ghali, K., Nicieza, A. G., O'Donnell, R. P., Rodrigues, N., Romano, A., Martínez-Solano, I., … Perrin, N. (2018). A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. *Nature Communications*, 9(1), 4088. https://doi.org/10.1038/s41467-018-06517-2

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, 37(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kumar, S., Stecher, G., Suleski, M., & Hedges, S. B. (2017). TimeTree: A resource for timelines, timetrees, and divergence times. *Molecular Biology and Evolution*, 34(7), 1812–1819. https://doi.org/10.1093/molbev/msx116

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. https://doi.org/10.1038/nmeth.1923

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R.; 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, X., Mank, J. E., & Ban, L. (2022). Grasshopper genome reveals long-term conservation of the X chromosome and temporal variation in X chromosome evolution. *bioRXiv*. https://doi.org/10.1101/2022.09.08.507201

Li, Y., Park, H., Smith, T. E., & Moran, N. A. (2019). Gene family evolution in the pea aphid based on chromosome-level genome assembly. *Molecular Biology and Evolution*, 36(10), 2143–2156. https://doi.org/10.1093/molbev/msz138

Lovell, J. T., Sreedasyam, A., Schranz, M. E., Wilson, M., Carlson, J. W., Harkess, A., Emms, D., Goodstein, D. M., & Schmutz, J. (2022). GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *eLife*, 11, e78526. https://doi.org/10.7554/eLife.78526

Marshall Graves, J. A., & Peichel, C. L. (2010). Are homologies in vertebrate sex determination due to shared ancestry or to limited options?. *Genome Biology*, 11(4), 205. https://doi.org/10.1186/gb-2010-11-4-205

Mathers, T. C., Wouters, R. H. M., Mugford, S. T., Swarbreck, D., van Oosterhout, C., & Hogenhout, S. A. (2021). Chromosome-scale

genome assemblies of aphids reveal extensively rearranged autosomes and long-term conservation of the X chromosome. *Molecular Biology and Evolution*, 38(3), 856–875. https://doi.org/10.1093/molbev/msaa246

McSwan, E., Clifford, C., Macadam, C. R., & Price, B. W.; Natural History Museum Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, C. R. Macadam, B. W. Price, and Darwin Tree of Life Consortium. (2023). The genome sequence of the Common Yellow Sally, *Isoperla grammatica* (Poda, 1761). *Wellcome Open Research*, 8, 107. https://doi.org/10.12688/wellcomeopenres.19066.1

Meisel, R. P., Delclos, P. J., & Wexler, J. R. (2019). The X chromosome of the German cockroach, *Blattella germanica*, is homologous to a fly X chromosome despite 400 million years divergence. *BMC Biology*, 17(1), 100. https://doi.org/10.1186/s12915-019-0721-x

Misof, B., Liu, S., Meusemann, K., Peters, R. S., Donath, A., Mayer, C., Frandsen, P. B., Ware, J., Flouri, T., Beutel, R. G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A. J., Aspöck, U., Aspöck, H., Bartel, D., … Zhou, X. (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science*, 346(6210), 763–767. https://doi.org/10.1126/science.1257570

Morris, J. L., Puttick, M. N., Clark, J. W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C. H., Yang, Z., Schneider, H., & Donoghue, P. C. J. (2018). The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 115(10), E2274–E2283. https://doi.org/10.1073/pnas.1719588115

Pal, A., & Vicoso, B. (2015). The X chromosome of hemipteran insects: Conservation, dosage compensation and sex-biased expression. *Genome Biology and Evolution*, 7(12), 3259–3268. https://doi.org/10.1093/gbe/evv215

Parker, D. J., Jaron, K. S., Dumas, Z., Robinson-Rechavi, M., & Schwander, T. (2022). X chromosomes show relaxed selection and complete somatic dosage compensation across *Timema* stick insect species. *Journal of Evolutionary Biology*, 35(12), 1734–1750. https://doi.org/10.1111/jeb.14075

Pease, J. B., & Hahn, M. W. (2012). Sex chromosomes evolved from independent ancestral linkage groups in winged insects. *Molecular Biology and Evolution*, 29(6), 1645–1653. https://doi.org/10.1093/molbev/mss010

Pennell, M. W., Kirkpatrick, M., Otto, S. P., Vamosi, J. C., Peichel, C. L., Valenzuela, N., & Kitano, J. (2015). Y fuse? Sex chromosome fusions in fishes and reptiles. *PLoS Genetics*, 11(5), e1005237. https://doi.org/10.1371/journal.pgen.1005237

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. -C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, 33(3), 290–295. https://doi.org/10.1038/nbt.3122

Pokorná, M., & Kratochvíl, L. (2009). Phylogeny of sex-determining mechanisms in squamate reptiles: Are sex chromosomes an evolutionary trap?. *Zoological Journal of the Linnean Society*, 156(1), 168–183. https://doi.org/10.1111/j.1096-3642.2008.00481.x

Price, B. W., Winter, M., & Brooks, S. J.; Natural History Museum Genome Acquisition Lab. (2022). The genome sequence of the blue-tailed damselfly, *Ischnura elegans* (Vander Linden, 1820). *Wellcome Open Research*, 7, 66. https://doi.org/10.12688/wellcomeopenres.17691.1

Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics*, 24(5), 637–644. https://doi.org/10.1093/bioinformatics/btn013

Stork, N. E. (2018). How many species of insects and other terrestrial arthropods are there on Earth?. *Annual Review of Entomology*, 63, 31–45. https://doi.org/10.1146/annurev-ento-020117-043348

Toups, M. A., Rodrigues, N., Perrin, N., & Kirkpatrick, M. (2019). A reciprocal translocation radically reshapes sex-linked inheritance in the common frog. *Molecular Ecology*, 28(8), 1877–1889. https://doi.org/10.1111/mec.14990

Veyrunes, F., Waters, P. D., Miethke, P., Rens, W., McMillan, D., Alsop, A. E., Grützner, F., Deakin, J. E., Whittington, C. M., Schatzkamer, K., Kremitzki, C. L., Graves, T., Ferguson-Smith, M. A., Warren, W., & Marshall Graves, J. A. (2008). Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. *Genome Research*, 18(6), 965–973. https://doi.org/10.1101/gr.7101908

Vicoso, B. (2019). Molecular and evolutionary dynamics of animal sex-chromosome turnover. *Nature Ecology & Evolution*, 3(12), 1632–1641. https://doi.org/10.1038/s41559-019-1050-8

Vicoso, B., & Bachtrog, D. (2015). Numerous transitions of sex chromosomes in Diptera. *PLoS Biology*, 13(4), e1002078. https://doi.org/10.1371/journal.pbio.1002078

Yu, D., Deharveng, L., Lukić, M., Wei, Y., Hu, F., & Liu, M. (2021). Molecular phylogeny and trait evolution in an ancient terrestrial arthropod lineage: Systematic revision and implications for ecological divergence (Collembola, Tomocerinae). *Molecular Phylogenetics and Evolution*, 154, 106995. https://doi.org/10.1016/j.ympev.2020.106995

Zdobnov, E. M., von Mering, C., Letunic, I., Torrents, D., Suyama, M., Copley, R. R., Christophides, G. K., Thomasova, D., Holt, R. A., Subramanian, G. M., Mueller, H. -M., Dimopoulos, G., Law, J. H., Wells, M. A., Birney, E., Charlab, R., Halpern, A. L., Kokoza, E., Kraft, C. L., … Bork, P. (2002). Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science*, 298(5591), 149–159. https://doi.org/10.1126/science.1077061

Zhang, F., Ding, Y., Zhou, Q. -S., Wu, J., Luo, A., & Zhu, C. -D. (2019). A high-quality draft genome assembly of *Sinella curviseta*: A soil model organism (Collembola). *Genome Biology and Evolution*, 11(2), 521–530. https://doi.org/10.1093/gbe/evz013