Gene-rich X chromosomes implicate intragenomic conflict in the evolution of bizarre genetic systems

Noelle Anderson[1], Kamil S. Jaron[2], Christina N. Hodson[2], Matthew B. Couger[3], Jan Ševčík[4], Stacy Pirro[5], Laura Ross[2] and Scott William Roy[1,6]

[1]Department of Molecular and Cell Biology, University of California-Merced, Merced, CA 95343

[2]Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3FL

[3]Brigham and Women's Hospital, Department of Thoracic Surgery, Boston MA 02115

[4]Faculty of Science, Department of Biology and Ecology, University of Ostrava, Ostrava, Czech Republic

[5]Iridian Genomes, Inc., Bethesda, MD, 20817, USA

[6]Department of Biology, San Francisco State University, San Francisco, CA 94117

Abstract

Haplodiploidy and paternal genome elimination (HD/PGE) are common in animals, having evolved at least two dozen times. HD/PGE typically evolves from male heterogamety (i.e., systems with X chromosomes), however why X chromosomes are important for the evolution of HD/PGE remains debated. The Haploid Viability Hypothesis argues that X chromosomes promote the evolution of male haploidy by facilitating purging recessive deleterious mutations. The Intragenomic Conflict Hypothesis instead argues that X chromosomes promote the evolution of male haploidy due to conflicts with autosomes over sex ratios and transmission. To test these hypotheses, we studied lineages that combine germline PGE with XX/X0 sex determination (gPGE+X systems). Because the evolution of such systems involves changes in genetic transmission but not increases in male hemizygosity, a high degree of X linkage in these

systems is predicted by the Intragenomic Conflict Hypothesis but not the Haploid Viability Hypothesis. Through de novo genome sequence, we compared the genomes of 7 species with gPGE+X systems and 10 related species with typical XX/XY or XX/X0 genetic systems. We find highly increased X-linkage in modern and ancestral genomes of gPGE+X species, with an estimated 30 times more X-linked genes than in non-gPGE+X relatives. These results suggest a general role for intragenomic conflict in the origins of PGE/HD. These findings are among the first empirical results supporting a role for intragenomic conflict in the evolution of novel genetic systems.

Introduction

Many animal lineages have evolved genetic systems in which females are diploid but males are genetically haploid, with each male creating genetically identical sperm carrying the single haploid genome originally inherited from his mother (*1*). Such systems range from haplodiploidy (HD), in which males are produced from unfertilized eggs; to embryonic paternal genome elimination (ePGE), in which diploid male embryos somatically eliminate their paternal genome; to forms of germline-specific PGE (gPGE), where the paternal genome is somatically expressed but excluded during spermatogenesis (Figure 1a).

HD/PGE is widespread, seen in ~12% of arthropods and having evolved roughly two dozen times (*1*). This recurrent evolution perhaps reflects the various advantages of HD/PGE, particularly to mothers, who can increase the transmission of their genes over paternally-inherited genes, control the sex ratio, ensure reproductive success without a mate (in HD), and, under monogamy, reduce conflict between gregarious offspring (*2-6*).

Given these general benefits, why does HD/PGE evolve in some lineages and not in others? An important hint comes from the finding that HD/PGE tends to evolve from ancestral male heterogamety (XX/XY or XX/X0) (*7, 8*). There are two competing explanations for this association. According to the Haploid Viability hypothesis, hemizygosity of X-linked genes

facilitates purging of recessive deleterious mutations, increasing the fitness of newly-evolved haploid males (*5, 8, 9, 10*). According to the Intragenomic Conflict hypothesis, the importance in X linkage lies in inherent conflicts between X-linked and autosomal genes. For instance, X-linked genes could promote X chromosome drive, in which the X chromosome is transmitted through >50% of sperm, leading to female-biased population sex ratios. Such sex ratio skew is expected to drive counterstrategies to rebalance the sex ratio, possibly leading to new sex determination mechanisms (*11, 12*). HD/PGE in particular could evolve under the Intragenomic Conflict hypothesis through the exploitation of drive by maternal autosomes that increase their transmission by becoming effectively X-linked (*13*).

These two hypotheses differ in whether they predict an association between X linkage and the origins of gPGE, in which paternal chromosomes are only lost from the germline, but are expressed in the soma. This may be seen by noting that the origins of gPGE systems entails a turnover of sex determination and transmission, the aspects emphasized by the Intragenomic Conflict Hypothesis, but does not entail an increase in hemizygosity, the aspect emphasized by the Haploid Viability hypothesis (Fig. 1a). Specifically, if the association between X-linkage and the origins of novel sex determination and transmission systems is explained by novel systems being the outcomes of conflict between X-linked and autosomal genes (the Intragenomic Conflict Hypothesis), this association might equally be expected for gPGE, since the origins of gPGE involves a turnover of sex determination and transmission. On the other hand, if the association between X-linkage and the origins of novel systems is explained by X-linkage decreasing the costs of increased hemizygosity of genes in males, then this association is not expected for gPGE, since the origins of gPGE are not expected to involve an increase in hemizygosity of genes in males. Thus, an association between X-linkage and the origins of gPGE is predicted by the Intragenomic Conflict Hypothesis but not by the Haploid Viability hypothesis.

To our knowledge, this differential prediction has not been noted or tested. The gPGE genetic systems of flies in Sciaridae and Cecidomyiidae (two families in the diverse dipteran superfamily Sciaroidea) and of springtails in the order Symphypleona, offer a powerful opportunity. These three groups have independently evolved a variant of gPGE, in which males are produced through somatic elimination of paternal X chromosomes, while the remainder of the paternal genome is retained until its elimination during spermatogenesis (Fig. 1) (7,*14-17*).

To test these two hypotheses for the origins of HD/PGE, we performed genome sequencing and comparative analysis of 17 species of gPGE and related species. We find clear evidence for ancestral gene-rich X chromosome coincident with three independent origins of gPGE. These results provide the first empirical evidence for a role for intragenomic conflict in the origins of atypical genetic systems.

Results and Discussion

Increased numbers of X-linked genes in gPGE species relative to related species

To test whether the evolution of gPGE is associated with gene-rich X chromosomes, we determined genome-wide patterns of X chromosomal linkage for 17 species of flies and two species of springtail. For the flies, we sampled across seven families spanning the root of Sciaroidea including two families with gPGE and two outgroup species. For these springtails, we performed genomic sequencing of males from one species from the gPGE order Symphypleona, *Allacma fusca* (Fig. 1d), and males and females of *Orchesella cincta*, from Entomobryidae, the closest relative springtail order with standard XX/X0 sex determination. Illumina genome sequencing and assembly was performed for males of each species, and average read coverage calculated for each contig. For the fly species, putative orthologs of *D. melanogaster* genes were identified via TBLASTN searches of each genome. Each ortholog was then assigned to one of the so-called Muller elements, *D. melanogaster* chromosomal linkage groups that have largely persisted over long evolutionary times in Diptera (*18, 19*). For

each Muller group in each species, the fraction of genes that are X-linked was estimated from read coverage distributions using improved methods based on Vicoso and Bachtrog (*19*). These methods provided clear estimates of X-linkage for nearly all our species, with exceptions in two species which were excluded from Fig. 2, non-gPGE *Exechia fusca*, likely due to poor assembly quality (Fig. S1, S2) and the Cecidomyiid *Lestremia cinerea* which showed three distinct peaks rather than two. The genome of the *Allacma fusca* springtail was sequenced and assembled in a very similar way, but instead of orthologs, we used genome annotations to estimate the gene density. We used publicly available genome assembly and annotation of *O. cincta* and *M. destructor* (Fig. 1b), and the X-linked scaffolds were identified using the mapping approach used for the other species with the addition of female DNA data to identify X-linked genes by relative coverage.

Among all non-gPGE fly species of Sciaroidea, we found very few X-linked genes, with the X chromosome in all species comprised mostly of genes from the diminutive F Muller element (<1% of all genes), consistent with the previous inference for the ancestral dipteran X chromosome (Fig. 2a) (*19*). Interestingly, in *Symmerus nobilis*, sister to all other Sciaroidea species, no Muller elements exhibited clear X-linked peaks, suggesting either homomorphic sex chromosomes or the lack of an X chromosome.

By contrast, for all six studied gPGE species in both the Sciaridae and Cecidomyiidae clades, genome-wide, we found large fractions of genes to be X-linked, including genes from all six Muller elements (Fig. 2a, 3). Notably, our results agree with previous results for *M. destructor*, identifying Muller elements C, D, F and E as partially X-linked (*19*), and our methods also detect partial X linkage for elements A and B. Both analysed springtails carried X-linked genes. However, while only 14.6% of genes in the genome of non-gPGE *Orchesella cincta* are X-linked, for the gPGE springtail *Alacma fusca*, 42.7% of annotated genes are X-linked (Fig. 2b).

Correspondence between X-linked genes within families indicates ancestrally gene-rich X chromosomes

Although we found an association between gene rich X chromosomes and gPGE in all three independent origins of this genetic system, the observed association could be explained by either X linkage facilitating the evolution of gPGE or vice versa. Consistent with the former, we see the same patterns of Muller group X-linkage within families (E>A>B in Sciaridae species; C>D>E>A>B in Cecidomyiidae). In addition, we found an association between X-linked gene subsets within individual Muller elements, as expected from ancestral linkage. For instance, the subsets of Muller B genes that are X-linked in the Sciaridae species *B. tilicola* and *T. splendens* significantly overlap, and the same is true for all partially X-linked Mullers in both Sciaridae (Fig. 3). By contrast, X-linked genes shared between Sciaridae and Cecidomyiidae are not overrepresented, supporting independent origins of the large X (Fig. S3).

Examination of Cecidomyiidae reveals an intriguing pattern. The deeply-diverged species *C. subobsoleta* and *M. destructor* show high correspondence between X-linked gene subsets, indicating substantial ancestral X-linkage. However, *P. nigripennis* shows divergent X linkage, with no significant pattern seen in shared X-linkage with other Cecidomyiids, and a relative increase in X-linkage on Muller elements A, B, and E. This pattern suggests turnover and increases in X linkage in this lineage since the divergence from *M. destructor* (or, less parsimoniously, parallel loss of A/B/E linkage in the other lineages) (Fig. 2a, 3).

At the same time, our data attest to substantial dynamism of the X chromosome in both gPGE families. Notably, such dynamism is not predicted by common models of X chromosomal evolution. Sex chromosome turnover is generally thought to be driven by sexual antagonism, since alleles that are beneficial in females but not in males benefit from being X-linked due to more frequent transmission through females (*20-22*). However, in the Sciarid and Cecidomyid systems, males transmit their entire maternal genome, thus X chromosomes are not more frequently female-transmitted than are autosomes (Fig. 1). Thus, the marked turnover of X

chromosomal gene complement in gPGE species is not predicted by standard models of sex chromosome evolution. However, the atypical dynamics of these systems could drive increased X linkage by atypical mechanisms. In particular, conflict between maternally and paternally derived genes promotes a reduction in expression of the paternal genome in males, as hypothesized for other PGE lineages (*12*). An increase in X linkage will increase the proportion of the genome that is exclusively expressed from the maternal copy in males.

Concluding remarks

In this study, we find that species in the gPGE groups Cecidomyiidae and Sciaridae have, on average, X chromosomes 30 times more gene-rich than non-gPGE Sciaroidea species, while the X chromosome gene content of the gPGE springtail species has more than doubled in comparison to the diploid outgroup. These findings represent the first empirical evidence that Intragenomic Conflict drives the evolution of abnormal sex determining systems such as HD/gPGE. Given the widespread and repeated evolution of male haploidy, and its association with many unique ecological and life history strategies, our findings point to an important role for intragenomic conflict in shaping biology at all levels from molecule to organism to community.

Materials and Methods

Specimens and sequencing:

In order to contrast X chromosomes of gPGE species to their diplodiploid relatives, we collected and sequenced males of 18 species, 14 belonging to the superfamily Sciaroidea spanning nearly all families within, two outgroup species in the dipteran families Anisopodidae and Bibionidae (Sciaroidea and these families are both in the infraorder Bibionomorpha), and two belonging to the springtail species *Allacma fusca* and *Orchesella cincta.* Eleven dipteran specimens were collected and provided by Jan Ševčík, *Catotricha subobsoleta* by Scott

Fitzgerald and *Bolitophila hybrida* by Nikola Burdíková, and *Bradysia tilicola* is cultured at the University of Edinburgh. Springtails were provided by Jacintha Ellers. Both specimens were flash-frozen and stored at -80°C. We used publicly available genome assemblies for the Cecidomyiid *Mayetiola destructor* (GCA_000149195.1) and for the springtail *Orchesella cincta* (GCA_001718145.1). For *M. destructor,* publicly available male (SRR1738190) and female reads (SRR1738189) were used, and for *O. cincta*, female reads (SRR2222657) were used.

For 15 dipteran species, DNA extractions (Qiagen DNAeasy Blood & Tissue kit), library preparation (Illumina TruSeq kit), and sequencing (Illumina Hi-Seq) were performed by Iridian Genomes. Genomes were assembled using Megahit 1.13 (*23*) by Brian Couger at Oklahoma State University. For the two collembolan species and the Sciaridae *Bradysia tilicola*, DNA was extracted using a modified extraction protocol from DNAeasy Blood & Tissue kit (Qiagen, The Netherlands) and Wizard Genomic DNA Purification kit (Promega). TruSeq DNA Nano gel free libraries (350 bp insert) were generated by Edinburgh Genomics (UK) and sequenced on the Illumina HiSeq X (for springtails) or NovaSeq S1 (for *B. tilicola*) generating short reads (150 bp paired-end). The genome for *B. tilicola* was assembled using Megahit 1.2.9 (*23*). The genome of springtail *A. fusca* was assembled using SPAdes v3.13.1 (*24*). Both genomes of *B. tilicola* and *A. fusca* assemblies were decontaminated with blobtools (*25*). The assembly of *A. fusca* was annotated using braker (version 2.1.5) (*26*). For the other springtail *O. cincta* we used a publicly available genome assembly and annotation (GCA_001718145.1). We assessed the quality of all genomes using BUSCO (*27*), to determine the proportion of single copy orthologs expected to be present in either insects (insecta_odb10 for fungus gnat species) or arthropods (for springtails) in the genome assemblies. Two genomes lacking a substantial fraction of complete BUSCO genes, *Exechia fusca*, or with irregular genome coverage patterns, *Lestremia cinerea* were excluded from downstream analysis (Fig. S1 and S2).

Assigning ancestral linkage groups:

The X chromosome in each fly species was identified using two strategies— Muller group linkage and genomic read coverage, similar to strategies implemented in Vicoso and Bachtrog 2015. Muller elements are six chromosomal elements first characterized in *Drosophila* that are well-conserved within Diptera and are thus informative about chromosomal linkage (*28*). The *D. melanogaster* proteome (flybase r6.32) (*28)* was searched against each assembled Bibionomorphan genome translated into 6 frames using TBLASTN. Top hits for each *D. melanogaster* gene were identified and corresponding Bibionomorphan genes were classified by the Muller element of their closest *D. melanogaster* ortholog. The X chromosomes in springtails were identified using the coverage approach only.

Identifying X-linkage via coverage

Our second strategy implemented DNA coverage levels to characterize autosomal and X-linked sequence. Because the X chromosome is present in a single copy in males, in males, sequence that is X-linked is expected to be at half coverage compared to autosomal sequence. Male DNA reads (trimmed to 50nts) for each Bibionomorphan were mapped to their respective genome assemblies using Bowtie, discarding reads that mapped to multiple locations in the genome. Because some Bibionomorphan contigs contained large amounts of repetitive sequence that prevented reads from mapping singly, we corrected coverage estimates to only account for singularly mappable positions on the contigs. To do this, we simulated 50nt reads from every mappable position on each contig, mapped them back to the genome from which they were generated from using Bowtie, and subtracted the number of reads from each contig that were unable to map singularly from the contig length. This provided us with an adjusted length that excluded sequence content that could not be mapped to singularly to use for adjusting coverage estimates. Coverage was calculated as: (Read count x read length) / (Contig length - number of multiply mapping reads for that contig + 1). Because male and female DNA sequence for *M.*

destructor is available, the comparison of male to female read coverage was used in addition to using linkage information previously established by physical mapping to more stringently classify X-linkage.

To classify Bibionomorphan genes by coverage as either autosomal or X-linked, we used a multi-step protocol. First, we used standard methods to (i) identify the highest peak in the coverage distribution; and (ii) identify the highest secondary peak near half or twice the coverage of the highest peak (as expected if a minority/majority of the genome is X-linked, respectively). Second, the autosomal peak was selected by either choosing the higher peak (for species in which there was effectively no second peak, methodologically defined as when the higher peak was >10X the height of the other) or the peak at higher coverage (for species with two large peaks, i.e., those with substantial fractions of X-linked genes). For further analysis, the expected X-linked coverage value was set as one-half that of the autosomal value.

Next, Muller-distribution specific peaks were located by searching the highest peaks closest to the full genome distribution X and autosomal peak estimates. Muller element coverage distributions were deemed as having two peaks by the same relative peak height comparison used on the full genome distributions, and if the coverage values in between peaks were non-monotonic. Genes per Muller element distribution were assigned as X-linked or autosomal via k-means clustering, using the Muller-specific X and autosomal peaks as initial cluster centers. The doubled standard deviation of the proportion of X-linked genes in each distribution was estimated as a proxy for the 95% confidence interval.

Testing for ancestral Muller group linkage

To test for evidence of ancestral X linkage, we compared various pairs of species. We studied each Muller element for which both compared species had partial X linkage, in which the ancestral linkage groups have broken up and are now partially X-linked and partially autosomal. Genomes of each species pair were reciprocally blasted to defined putative pairwise orthologs

using TBLASTX. Only best reciprocal hits and orthologs that blasted to the same *D. melanogaster* gene were included in further analysis. Each ortholog pair was then assigned based on its inferred X/autosomal linkage for both species (X-linked/X-linked, X-linked/autosomal, autosomal/X-linked, or autosomal/autosomal). Association between X-linkage across between-species orthologs was tested by a Chi square test.

Acknowledgements and funding sources

We would like to acknowledge Scott Fitzgerald, Nikola Burdíková, Jacintha Ellers for their work collecting specimens. Genome sequencing of all dipteran specimens besides *B. tilicola* was funded by Iridian Genomes. NA and SWR were supported by NSF Award #1616878. Springtail sequencing and KSJ were supported by a European Research Council Starting Grant (PGErepro, to LR). CH was supported by the National Sciences and Engineering Research Council of Canada postgraduate scholarship and the Darwin Trust of Edinburgh. Springtail collection and LR were supported by a Natural Environment Research Council Independent Research Fellowship (NE/K009516/1) and Dorothy Hodgkin Fellowship (DHF\R1\180120).

References

1. T.-L. Ashman, D. Bachtrog, H. Blackmon, E. E. Goldberg, M. W. Hahn, M. Kirkpatrick, J. Kitano, J. E. Mank, I. Mayrose, R. Ming, S. P. Otto, C. L. Peichel, M. W. Pennell, N. Perrin, L. Ross, N. Valenzuela, J. C. Vamosi, Tree of Sex: A database of sexual systems. *Sci. Data*. *1*, 140015 (2014).

2. S. W. Brown, Automatic frequency response in the evolution of male haploidy and other coccid chromosome systems. *Genetics*. *49*, 797–817 (1964).

3. W. D. Hamilton, Extraordinary sex ratios. *Science. 156*, 477–488 (1967).

4. D. L. Hartl, S. W. Brown, The origin of male haploid genetic systems and their expected sex ratio. *Theor. Popul. Biol. 1*, 165–190 (1970).

5. J. J. Bull, An advantage for the evolution of male haploidy and systems with similar genetic transmission. *Heredity. 43*, 361–381 (1979).

6. B. B. Normark, Maternal kin groups and the origins of asymmetric genetic systems: Genomic imprinting, haplodiploidy, and parthenogenesis. *Evolution. 60*, 631–642 (2006).

7. A. Gardner, L. Ross, Mating ecology explains patterns of genome elimination. *Ecol. Lett. 17*, 1602–1612 (2014).

8. H. Blackmon, N. B. Hardy, L. Ross, The evolutionary dynamics of haplodiploidy: Genome architecture and haploid viability. *Evolution. 69*, 2971–2978 (2015).

9. D. B. Goldstein, Deleterious mutations and the evolution of male haploidy. *Am. Soc. Nat. 144*, 176–183 (1994).

10. B. B. Normark, Haplodiploidy as an outcome of coevolution between male-killing cytoplasmic elements and their hosts. *Evolution. 58*, 790–798 (2004).

11. D. Haig, The evolution of unusual chromosomal systems in coccoids: extraordinary sex ratios revisited. *J. Evol. Biol. 6*, 69–77 (1993).

12. L. Ross, I. Pen, D. M. Shuker, Genomic conflict in scale insects: The causes and consequences of bizarre genetic systems. *Biol. Rev. 85*, 807–828 (2010).

13. D. Haig, The evolution of unusual chromosomal systems in sciarid flies: intragenomic conflict and the sex ratio. *J. Evol. Biol.* 6, 249–261 (1993).

14. A. Burt, R. Trivers, “Genomic Exclusion” in *Genes in Conflict* (Harvard University Press, Cambridge. Ed. 2006), pp. 395-404.

15. C.W. Metz. Chromosome behavior, inheritance and sex determination in Sciara. *American Naturalist*. *72*, 485–520 (1938).

16. R. L. Gallun, H. H. Hatchett. Genetic evidence of elimination of chromosomes in the Hessian fly. *Annals of the Entomological Society of America. 62*, 1095-1101 (1969).

17. R. Dallai, P. P. Fanciulli, F. Frati. Aberrant spermatogenesis and the peculiar mechanism of sex determination in Symphypleonan Collembola (Insecta). *The Journal of Heredity. 91(5)*, 351-8 (2000)

18. H. J. Muller, Bearings of the ‘Drosophila’ work on systematics. *The New Systematics*, 185–268 (1940).

19. B. Vicoso, D. Bachtrog, Numerous transitions of sex chromosomes in diptera. *PLoS Biol. 13* (2015).

20. B. B. Normark, The evolution of alternative genetic systems in insects. *Annu. Rev. Entomol. 48,* 397–423 (2003).

21. M. Kozielska, F. J. Weissing, L. W. Beukeboom, I. Pen, Segregation distortion and the evolution of sex-determining mechanisms. *Heredity. 104,* 100–112 (2010).

22. D. Bachtrog, Sex determination: why so many ways of doing it? *PLoS Biol. 12*, e1001899 (2014).

23. D. Li, R. Luo, C. M. Liu, C. M. Leung, H. F. Ting, K. Sadakane, H. Yamashita, T. W. Lam. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods. 102*, 3-11 (2016).

24. A. Bankevich, S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, P. A. Pevzner. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*. *19*, 455-477 (2012).

25. D. R. Laetsch, M. L. Blaxter. BlobTools: Interrogation of genome assemblies [version 1; peer review: 2 approved with reservations]. *F1000Research. 6*, 1287 (2017).

26. K. J. Hoff, A. Lomsadze, M. Borodovsky, M. Stanke. Whole-Genome Annotation with BRAKER. *Methods Mol Biol*. *1962*, 65-95 (2019).

27. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics. 19*, 3210-2 (2015).

28. S. J. Marygold, M. A. Crosby, J. L. Goodman, The FlyBase Consortium. Using FlyBase, a Database of Drosophila Genes and Genomes. Methods in Molecular Biology in *Drosophila*. (Humana Press, New York, NY. 2016), vol 1478.

29. J. Ševčík, D. Kaspřák, M. Mantič, S. Fitzgerald, T. Ševčíková, A. Tóthová, M. Jaschhof, Molecular phylogeny of the megadiverse insect infraorder Bibionomorpha *sensu lato* (Diptera). *PeerJ*. 4, e2563 (2016).

Figures

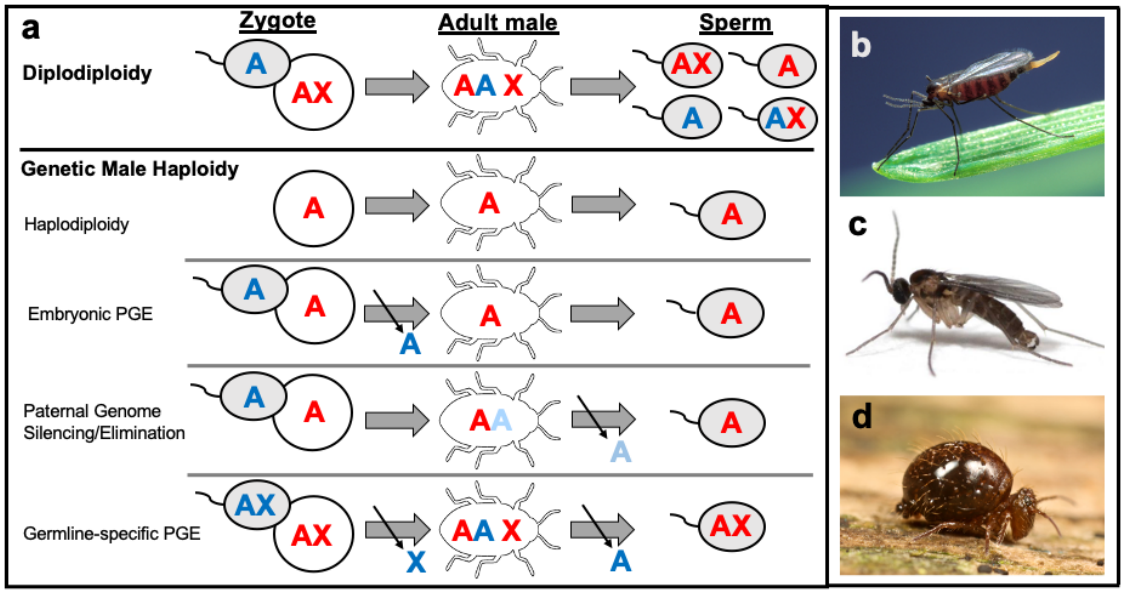

Figure 1. Schematic of different genetic systems discussed. a) Male production and spermatogenesis under diplodiploidy, and various forms of male genetic haploidy are shown. Blue and red letters indicate paternal and maternally derived material respectively. A and X represent autosomes and X chromosomes, respectively. Shown are haplodiploidy (HD); embryonic paternal genome elimination (ePGE); paternal genome silencing/elimination, in which genome PGE is coupled to somatic silencing of the paternal genome (indicated by the blue “A” in males); and germline-specific PGE (gPGE), as observed in Sciarids, Cecidomyids and Symphypleonan springtails, wherein males are produced by somatic loss of the paternal X chromosome(s), and the paternal genome is eliminated in spermatogenesis. b-d) representative species with gPGE: b) the Hessian fly *Mayetiola destructor* (Ceccidiomyiidae); c) the fungus gnat *Bradysia tilicola* (Sciaridae) and d) the springtail *Allacma fusca* (Sminthuridae).

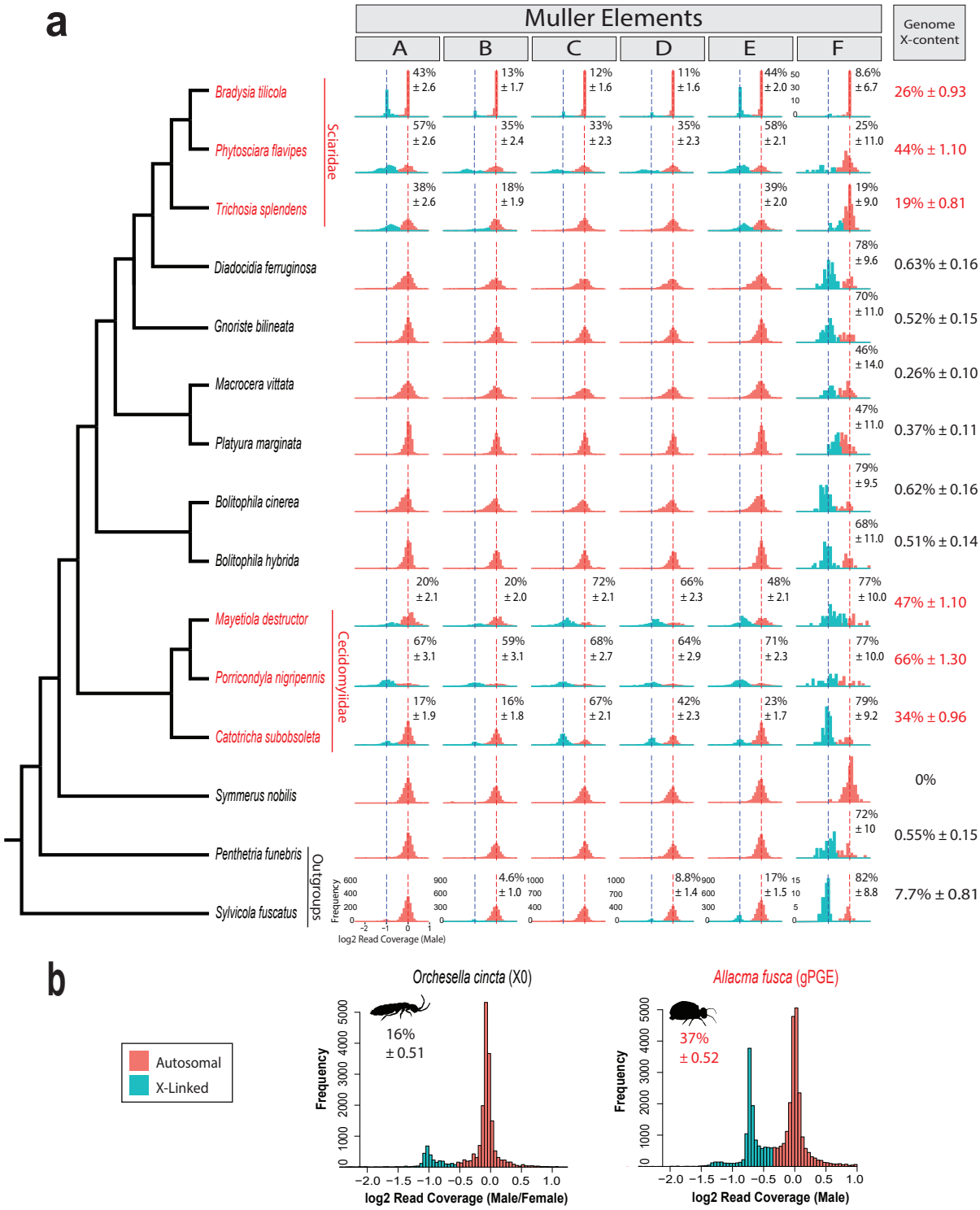

Figure 2. Frequency of X-linked and autosomal genes in gPGE species and related diplodiploid species, assessed by DNA read coverage. a) Sciaroidea and outgroups; phylogenetic tree based on Ševčík et al. (*29*). Histograms for each Muller element show log2 male read coverage normalized by putative median

autosomal coverage, with assigned X-linkage (blue bars) and autosomal linkage (red) indicated. Red dashed vertical lines indicate the expected autosomal coverage peak, blue dashed lines indicate the expected position of the X-linked peak, at half the coverage of the autosomes. Red and black species names and genome-wide estimates represent gPGE and diploid species, respectively. Percent estimates represent percent X-linkage for each Muller and across each full genome, with error represented by 2SD. b) Whole genome autosomal and X-linkage for springtails diplodiploid *Orchesella cincta* and gPGE *Allacma fusca*.

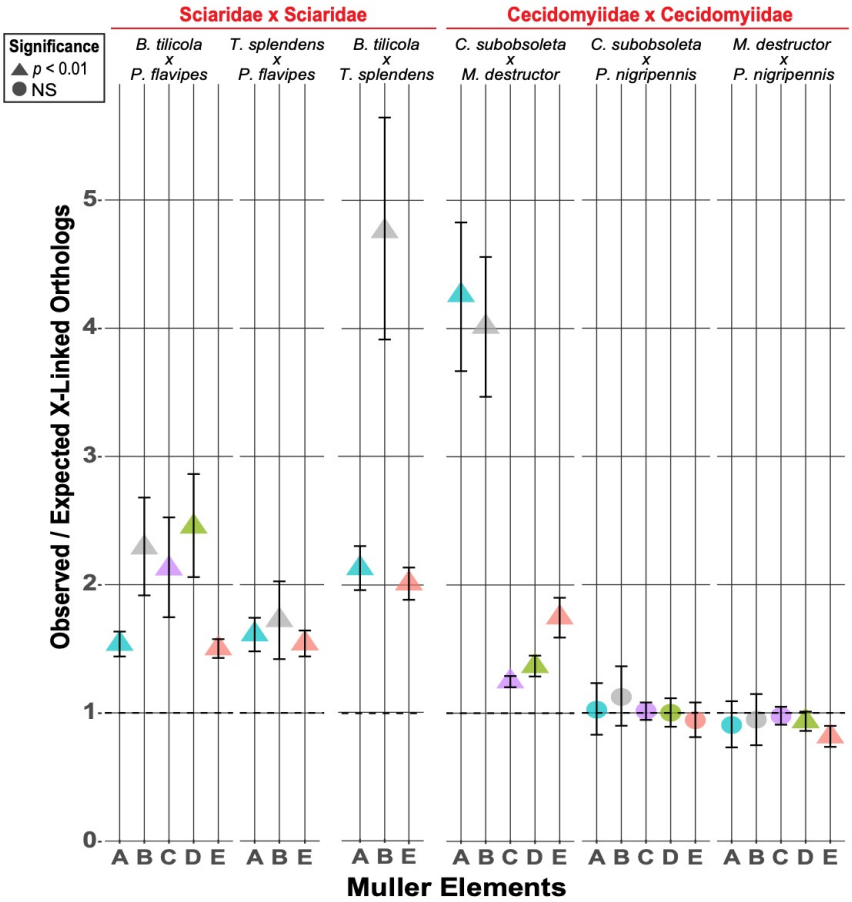

Figure 3. Number of ortholog pairs in which both genes are X-linked, compared to the null expectation, for pairs of gPGE species from the same family. Within-family comparisons are shown, between-family comparisons in Fig. S3. Color indicates Muller element. Muller elements for which species do not share X-linked orthologs are excluded, as is the F element. Shapes indicate significance via Chi square. Error bars represent 95% CIs computed from 10,000 bootstrap replicates. Expected value if no association between X-linked orthologs is 1.