



Published in final edited form as:

Front Ecol Evol. 2022 ; 10: . doi:10.3389/fevo.2022.832378.

Diptera and *Drosophila* Karyotype Databases: A Useful Dataset to Guide Evolutionary and Genomic Studies

Magnolia W. Morelli¹, Heath Blackmon², Carl E. Hjelman^{1,*}

¹Department of Biology, Utah Valley University, Orem, UT, United States,

²Department of Biology, Texas A&M University, College Station, TX, United States

Abstract

Karyotypes and chromosome data have been widely used in many subfields of biology over the last century. Unfortunately, this data is largely scattered among hundreds of articles, books, and theses, many of which are only available behind paywalls. This creates a barrier to new researchers wishing to use this data, especially those from smaller institutions or in countries lacking institutional access to much of the scientific literature. We solved this problem by building two datasets for true flies (Order: Diptera and one specific to *Drosophila*). These datasets are available via a public interactive database that allows users to explore, visualize and download all data. The Diptera karyotype databases currently contain a total of 3,474 karyotype records from 538 publications. Synthesizing this data, we show several groups are of particular interest for future investigations by whole genome sequencing.

Keywords

Diptera; *Drosophila* ; karyotype; sex chromosome; chromosome number

INTRODUCTION

For more than a century, karyotypes have been used to determine chromosome number within species (Metz, 1914), in addition to identifying sex chromosome systems, aneuploidy events, and structural mutations of species or populations (Powell, 1997; The Tree of Sex Consortium, 2014; Blackmon et al., 2017). Chromosome number is a simple and fundamental piece of genomic information collected for a species. Understanding the mode and tempo of chromosome number evolution across the tree of life is an active area of research (Blackmon et al., 2019; Ruckman et al., 2020; Sylvester et al., 2020). Additionally, a major application of chromosome number is used to measure the quality

This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

*Correspondence: Carl E. Hjelman, Carl.Hjelman@uvu.edu.

AUTHOR CONTRIBUTIONS

HB and CH conceived of the project and built the databases. All authors wrote and edited the manuscript.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

and “completeness” of genome assemblies; better assemblies should have contig/scaffold numbers close to the haploid number of chromosomes for the species. In the case of sex chromosome systems, karyotyping can reveal the presence of heteromorphic sex chromosome systems (XX/XY, ZZ/ZW, X0, Z0, and multiple/complex), and be used to identify recent chromosomal changes, such as the formation of a neo-sex chromosome system through the fusion of an autosome to the ancestral sex chromosomes. Despite the importance of karyotype data for basic and applied research in the life sciences, this data remains largely scattered among hundreds of articles, books, and theses, many of which are only available behind paywalls.

Karyotype data in the insect order Diptera has been abundantly produced over the last century, due to the presence of easily visualized and differentiated polytene chromosomes (Metz, 1914, 1916; Painter, 1933; Painter, 1934; Koltzoff, 1934; Ward, 1949). In fact, polytene chromosomes were key to *Drosophila melanogaster* becoming one of the most widely used and important model organisms in genetics and evolutionary biology (Powell, 1997; Zhimulev et al., 2004). The banding patterns of these chromosomes allowed early researchers to make genetic maps, identify mutations within species, and investigate how populations diverge (Dobzhansky, 1946; Powell et al., 1972; Anderson et al., 1975). Polytene chromosomes are not only confined to *Drosophila*, but are documented throughout the entire order of Diptera, leading to many studies of polytene chromosomes in Culicidae (mosquitos), Chironomidae (midges), and Simuliidae (black flies) (Kitzmilller, 1963; Grossbach, 1977; Rothfels, 1979; Rao and Rai, 1987; Michailova et al., 2001; Adler et al., 2004; Petrova and Zhirov, 2014). Extensive karyotype work, has shown that Diptera have conserved (syntenic) regions, often referred to as Muller elements A–F, which suggest conservation of gene groupings and linkage (White, 1949; The Tree of Sex Consortium, 2014; Vicoso and Bachtrog, 2015). As these six elements are shown to be highly conserved in many species, it has been hypothesized that most species would not have haploid chromosome numbers which exceed six (Sved et al., 2016). This karyotype data has been reported in books, peer-reviewed journals, and proceedings. Unfortunately, a substantial number of these are not available online, and some are not in print today.

In this study, we describe a new database containing karyotype information for the insect order Diptera, in addition to a separate database for the *Drosophila* genus. A broad understanding of karyotype evolution in Diptera may reveal key species to answer fundamental questions about genome biology, and result in species to be prioritized for sequencing. A freely accessible database for this order, similar to others compiled in Coleoptera, Polyneoptera, and Amphibians (Blackmon and Demuth, 2015a; Perkins et al., 2019; Sylvester et al., 2020) would accelerate future research in chromosome evolution and provide more equitable access to data.

MATERIALS AND METHODS

Collection of Karyotypes

We compiled karyotype information for Diptera using both online and library sources. Initially, we compiled karyotype information from previous compilations of *Drosophila* data (Ward, 1949; Patterson and Stone, 1952; Clayton and Wasserman, 1957; King, 1975;

Clayton, 1986). As these sources were not readily available online, we utilized the Texas A&M University Library system,¹ University of Texas Library system,² interlibrary loans, and the personal library of J. Spencer Johnston (Professor of Entomology, Texas A&M University). These records were supplemented with records collected by the Tree of Sex Consortium (³The Tree of Sex Consortium, 2014). Subsequently, we compiled records using Google Scholar searches combining Order (Diptera) or family names (for a total of 51 identified families) with “chromosome,” “chromosome number,” “karyotype,” “sex chromosome,” “Muller elements,” and “polytene chromosomes.” Additionally, we collected records by forward and backward citation tracking of previous compilations (Patterson and Stone, 1952; King, 1975).

References for records were compiled and condensed into one record for cases in which multiple reports are referring to the same data collection event. From each of the records obtained, we collected taxonomic information, chromosome number, chromosome morphology, sex system, and references (Table 1). Because *Drosophila* species have disproportionately more records than other families/genera and a different chromosome shape naming system, we created a separate database solely for this genus. While *Drosophila* species have their own database, their chromosome number data is also located in the more inclusive “Diptera Karyotype Database.”

Development of Online Database

We used the R package Shiny version 1.7.1 to create a dynamic and interactive database (Chang et al., 2021). The framework of this database is based on a data repository (all of the collected karyotype and taxonomic information) and three user modules. The first user module allows a researcher to narrow down the data of interest based on taxonomic rank (Family and Genus). The second module allows users to visualize their data; in this module users can choose from the variables of haploid number, sex chromosome system, family and genus. The website will build plots of the chosen variables using a framework built on ggplot2 version 3.3.5 functions (Wickham et al., 2016). After visualizing the data users can select the third module which will display a table of all selected data. This module also allows the user to download a CSV file for offline use in their own research. Both the Diptera Karyotype Database and *Drosophila* Karyotype Database are hosted at www.karyotype.org with yearly updates, pending new records.

RESULTS AND DISCUSSION

The current database contains 3,474 dipteran karyotypes which we compiled from 538 publications. While we did a thorough search of the literature, we have certainly not located all available data. Some papers are not easily located through internet searches or libraries, and/or may be in languages the authors were not comfortable translating. Users producing new records or who are aware of records missing from the database are encouraged to contact the authors to allow continued growth of this community resource.

¹ <https://library.tamu.edu/>

² <https://www.lib.utexas.edu/>

³ <http://treeofsex.org/>

A total of 36% (1,244) of the records we found are from *Drosophila*. These records are available as part of the Diptera database but also the more focused *Drosophila* Karyotype Database. When examining all data in a phylogenetic perspective it appears some clades may have less stable karyotypes than others [phylogeny from Wiegmann et al. (2011)]. For instance, in Figure 1, we can compare two families like Simuliidae (237 records) and Chironomidae (181 records). Generally assuming one record per species, that suggests we cover approximately 14% of the 1,700 estimated species in Simuliidae and merely 1–2% of the estimated 10,000 species in Chironomidae. We might expect Simuliidae to exhibit more variation in chromosome number since we have considerably more records for this group, both in raw number and taxon representation. However, our synthesis of all available data indicates Chironomidae, with just 181 records, has haploid numbers ranging from two to eight, while Simuliidae has only haploid counts of two and three, despite more extensive sampling, suggesting that rates of karyotype evolution are accelerated in Chironomidae relative to Simuliidae. Combining this data in the future with species level phylogenies would allow the application of probabilistic models of chromosome evolution (Zenil-Ferguson et al., 2018; Blackmon et al., 2019). This approach may identify which clades have more or less stable genome structure, and the underlying causes of these differences in the tempo of evolution and potential correlations with speciation rates.

Our database confirms that the vast majority of species (96%; 3,341) have a haploid chromosome count of six or less, suggesting a strong conservation of the six Muller elements (Figure 1). When visualizing the distribution of chromosome number on the phylogeny and focusing on records with six or fewer haploid chromosomes, it is apparent that species within Cyclorrhapha consistently have higher average chromosome number than those of earlier branching clades (Figure 1). This pattern suggests that chromosome number may have “settled” at the level of the six Muller elements in the cyclorrhaphous Diptera from an ancestral condition with fewer chromosomes. In fact, when viewed phylogenetically, chromosome number seems to have increased in steps, with the earliest branching Brachycera having chromosome numbers intermediate between Cyclorrhapha and earlier branching families, like Simuliidae and Culicidae. Further comparative phylogenetic approaches are necessary to validate this pattern of chromosome numbers in the phylogeny of Diptera.

While there is exceptional conservation of the six Muller elements in most of the Diptera phylogeny, there are a number of species which exceed this number. The 133 species with more than six chromosomes are not phylogenetically clustered; rather, 16 families (~25% of all families with data) have at least one species with greater than 6 chromosomes. Three families of orthorrhaphous dipterans (Bombyliidae, Tabanidae, and Stratiomyidae) have more than 20 species with greater than six chromosomes. It has been hypothesized that the paucity of species with more than six chromosomes in Diptera is driven by reliance on retrotransposons (evidence from *Drosophila* species) or tandemly repeated sequences (evidence from lower Diptera) to protect chromosome ends (Diptera lack telomerase) (Mason et al., 2011; Sved et al., 2016). Using our database provides a straightforward workflow to identify several distantly related species all of which overcame this apparent threshold on chromosome number. In fact, one of these species *Hermetia illucens* (black soldier fly) has a haploid chromosome number of seven. This species was recently

sequenced (Generalovic et al., 2021) combining this with additional species could help researchers understand how species can remodel their genome in the absence of telomerase.

Our database contains sex chromosome system information for 2,128 species (Figure 2). The vast majority of species have an XX/XY system (2,053 species). The remaining records include 11 multi-XY(X_1X_2Y , $X_1X_2Y_1Y_2$, XY_1Y_2), 57 XO, and 7 ZW systems. Loss of the Y chromosome, though rare, is relatively widespread. The 57 species with XO sex chromosome systems are spread across seven families with the most species (18) found in Cecidomyiidae. Interestingly Cecidomyiidae exhibits striking variation in sex chromosome systems. One species, *Phytophaga destructor*, exhibits homomorphic sex chromosomes likely an indication of sex chromosome turnover. Two other species (*Mayetiola destructor* and *M. hordei*) possess a multi sex chromosome system ($X_1X_1X_2X_2/X_1X_2Y$). These systems are usually found in cases where an autosome has fused to the Y chromosome (although fissions of the X chromosomes could presumably result in a X_1X_2Y system) in a species that was ancestrally XX/XY (a state unlikely to be ancestral in this group since 18 of 21 species are XX/XO). Its presence in these two species is even harder to explain because there is no reduction in its haploid number in comparison to related species; both species in the genus *Mayetiola* have a haploid number of four as do 17 of the other 19 species in this family.

The distribution of sex chromosome systems also highlights the importance of the family Tephritidae. While this family is relatively unremarkable when we consider haploid number (range from 4 to 7), sex chromosome systems in this group appear to be highly unstable. In this family, we have sex chromosome information for 71 species, yet we find all of the diversity of systems exhibited by the entire order of Diptera. The XX/XY system is still the most common, with 59 species spread wide taxonomically across 14 genera. The next most common is ZZ/ZW sex chromosome system, with seven species in five genera, which are likely from single origin (Frias, 1992). However, in the Tephritis genus, there is a species with an XO system, showing that even sister species can have different heterogametic sexes. We also find complex XX/XY sex chromosome systems in three species (two in the genus *Anastrepha* and one in the genus *Rhagoletis*). Finally, the XX/XO sex chromosome system is found in just two species *Rhagoletis meigeni* and *Tephritis arnicae*.

In insects, it has been hypothesized that complex sex chromosome systems often arise through fusions of autosomes with sex chromosomes (Blackmon and Demuth, 2015b). In fact, a recent study of Polyneoptera showed that species with complex sex chromosome systems have systematically lower numbers of autosomes (Sylvester et al., 2020). However, synthesizing all of our collected records we find Diptera with complex sex chromosome systems actually have a higher haploid number (mean of 5.5) than do species with a simpler and likely ancestral XX/XY sex chromosome system (mean of 5.2). A recent study of patterns of sex chromosome autosome fusions in *Drosophila* has suggested that the generation of these complex sex chromosome systems may be limited in species with achiasmatic meiosis (Anderson et al., 2020). However, at most this is only a partial explanation as not all Diptera possess achiasmatic meiosis.

Despite reductions in the cost of sequencing, we are still far from being able to sequence and *de novo* assemble all genomes. Moving forward, we believe one of the most powerful uses of this data will be to identify species whose genomes promise to be most useful in investigating the dynamics of genome structure evolution.

ACKNOWLEDGMENTS

We thank J. Spencer Johnston for the use of his personal library, James Alfieri and other members of the Blackmon Lab for reviewing an early version of the manuscript, and an editor for additional references that contained karyotype data.

FUNDING

This work was supported by National Institute of General Medical Sciences at the National Institutes of Health NIGMS R35GM138098 to HB.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: www.karyotype.org.

REFERENCES

- Adler PH, Currie DC, Wood DM, Idema RM, and Zettler LW (2004). *The Black Flies (Simuliidae) of North America*. Ithaca: Cornell University Press.
- Anderson NW, Hjelman CE, and Blackmon H (2020). The probability of fusions joining sex chromosomes and autosomes. *Biol. Lett* 16:20200648. doi: 10.1098/rsbl.2020.0648 [PubMed: 33232649]
- Anderson W, Dobzhansky T, Pavlovsky O, Powell J, and Yardley D (1975). Genetics of Natural Populations XLII. three decades of genetic change in *Drosophila pseudoobscura*. *Evolution* 29, 24–36. doi: 10.2307/2407139 [PubMed: 28563287]
- Blackmon H, and Demuth JP (2015a). Coleoptera karyotype database. *Coleopt. Bull* 69, 174–175. doi: 10.1649/0010-065X-69.1.174
- Blackmon H, and Demuth JP (2015b). Genomic origins of insect sex chromosomes. *Curr. Opin. Insect. Sci* 7, 45–50. doi: 10.1016/j.cois.2014.12.003 [PubMed: 32846676]
- Blackmon H, Justison J, Mayrose I, and Goldberg EE (2019). Meiotic Drive Shapes rates of karyotype evolution in mammals. *Evolution* 73, 511–523. doi: 10.1111/evo.13682 [PubMed: 30690715]
- Blackmon H, Ross L, and Bachtrog D (2017). Sex determination, sex chromosomes, and karyotype evolution in insects. *J. Hered* 108, 78–93. doi: 10.1093/jhered/esw047 [PubMed: 27543823]
- Chang W, Cheng J, Allaire JJ, Sievert C, Schloerke B, Xie Y, et al. (2021). Shiny: Web Application Framework for R. R Package Version 1.7.1. Available online at: <https://CRAN.R-project.org/package=shiny> (accessed date 2021-October-02).
- Clayton FE (1986). “Overview of Chromosomal Evolution in the Family Drosophilidae,” in *The Genetics and Biology of Drosophila* no. 3, eds Ashburner M, Carson HL, and Thompson JN (New York, NY: Academic), 1–38.
- Clayton FE, and Wasserman M (1957). Chromosomal studies of several species of *Drosophila*. *Univ. Texas Publs* 5721, 125–131.
- Dobzhansky T (1946). Genetics of Natural Populations. Xiii. Recombination and Variability in Populations of *Drosophila pseudoobscura*. *Genetics* 31, 269–290. doi: 10.1093/genetics/31.3.269 [PubMed: 20985721]
- Frias L (1992). Some aspects of the evolutive biology of Chilean species of Tephritidae (Diptera). *Acta Entomologica Chilena* 1992, 69–78.

- Generalovic TN, McCarthy SA, Warren IA, Wood JMD, Torrance J, Sims Y, et al. (2021). A high-quality, chromosome-level genome assembly of the black soldier fly (*Hermetia Illucens* L.). *G3* 11:5. doi: 10.1093/g3journal/jkab085
- Grossbach U (1977). "The Salivary Gland of Chironomus (Diptera): A Model System for the Study of Cell Differentiation," in *Biochemical Differentiation in Insect Glands*, ed. Beermann W (Berlin: Springer), 147–196. doi: 10.1007/978-3-540-37332-2_3
- King RC (1975). *Handbook of Genetics, Volume 3: Invertebrates of Genetic Interest*. New York, NY: Plenum Press.
- Kitzmiller JB (1963). Mosquito cytogenetics. a review of the Literature, 1953–62. *Bull. World Health Organ* 29, 345–355. [PubMed: 14058227]
- Koltzoff N (1934). The Structure of the Chromosomes in the Salivary Glands of *Drosophila*. *Science* 80, 312–313. doi: 10.1126/science.80.2075.312 [PubMed: 17769043]
- Mason JM, Reddy HM, and Frydrychova RC (2011). Telomere Maintenance in Organisms without Telomerase. *Chromosoma* 2011 323–346.
- Metz CW (1914). Chromosome Studies in the Diptera. I. A preliminary survey of five different types of chromosome groups in the genus *Drosophila*. *J. Exp. Zool* 17, 45–59. doi: 10.1002/jez.1400170103
- Metz CW (1916). Chromosome Studies on the Diptera. III. additional types of chromosome groups in the drosophilidae. *Am. Nat* 50, 587–599. doi: 10.1086/279569
- Michailova P, Ilkova J, Petrova N, and White K (2001). Rearrangements in the salivary gland chromosomes of *chironomus riparius* Mg.(Diptera, Chironomidae) Following Exposure to Lead. *Caryologia* 54, 349–363. doi: 10.1080/00087114.2001.10589246
- Painter TS (1933). A new method for the study of chromosome rearrangements and the plotting of Chromosome Maps. *Science* 78, 585–586. doi: 10.1126/science.78.2034.585 [PubMed: 17801695]
- Painter TS (1934). A new method for the study of chromosome aberrations and the plotting of chromosome maps in *Drosophila melanogaster*. *Genetics* 19, 175–188. doi: 10.1093/genetics/19.3.175 [PubMed: 17246718]
- Patterson JT, and Stone WS (1952). *Evolution in the Genus Drosophila*. New York, NY: Macmillan Company.
- Perkins RD, Gamboa JR, Jonika MM, Lo J, Shum A, Adams RH, et al. (2019). A database of amphibian karyotypes. *Chromosome Res.* 27, 313–319. doi: 10.1007/s10577-019-09613-1 [PubMed: 31338646]
- Petrova NA, and Zhirov SV (2014). Characteristics of the karyotypes of three subfamilies of chironomids (diptera, chironomidae: tanypodinae, diamesinae, prodiamesinae) of the World Fauna. *Entomol. Rev* 94, 157–165. doi: 10.1134/S001387381402002X
- Powell JR (1997). *Progress and Prospects in Evolutionary Biology: The Drosophila Model*. New York, NY: Oxford Press.
- Powell JR, Levene H, and Dobzhansky T (1972). Chromosomal Polymorphism in *Drosophila pseudoobscura* used for diagnosis of geographic origin. *Evolution* 26, 553–559. doi: 10.2307/2407052 [PubMed: 28563351]
- Rao PN, and Rai KS (1987). Comparative karyotypes and chromosomal evolution in some genera of nematoceros (diptera: nematocera) Families. *Ann. Entomol. Soc* 80, 321–332. doi: 10.1093/aesa/80.3.321
- Rothfels KH (1979). Cytotaxonomy of Black Flies (Simuliidae). *Annu. Rev. Entomol* 24, 507–539. doi: 10.1146/annurev.en.24.010179.002451
- Ruckman SN, Jonika MM, Casola C, and Blackmon H (2020). Chromosome Number Evolves at Equal Rates in Holocentric and Monocentric Clades. *PLoS Genet.* 16:e1009076. doi: 10.1371/journal.pgen.1009076 [PubMed: 33048946]
- Sved JA, Chen Y, Shearman D, Frommer M, Gilchrist AS, and Sherwin WB (2016). Extraordinary conservation of entire chromosomes in insects over long evolutionary Periods. *Evolution* 70, 229–234. doi: 10.1111/evo.12831 [PubMed: 26639450]
- Sylvester T, Hjelman CE, Hanrahan SJ, Lenhart PA, Johnston JS, and Blackmon H (2020). Lineage-specific patterns of chromosome evolution are the rule not the exception in Polyneoptera Insects. *Proc. Royal Soc. B* 287:20201388. doi: 10.1098/rspb.2020.1388

- The Tree of Sex Consortium (2014). Tree of sex: a database of sexual systems. *Sci. Data* 1:140015. doi: 10.1038/sdata.2014.15 [PubMed: 25977773]
- Vicoso B, and Bachtrog D (2015). Numerous transitions of sex chromosomes in diptera. *PLoS Biol.* 13:e1002078. doi: 10.1371/journal.pbio.1002078 [PubMed: 25879221]
- Ward CL (1949). Karyotype variation in *Drosophila*. *Univ. Texas Publ* 4920, 70–79.
- White MJD (1949). Cytological evidence on the phylogeny and classification of the diptera. *Evolution* 1949, 252–261. doi: 10.1111/j.1558-5646.1949.tb00025.x
- Wickham H, Chang W, and Wickham MH (2016). Package ‘Ggplot2.’ create elegant data visualisations using the grammar of graphics. Version 2, 1–189.
- Wiegmann BM, Trautwein MD, Winkler IS, Barr NB, Kim JW, Lambkin C, et al. (2011). Episodic radiations in the fly tree of life. *PNAS* 108, 5690–5695. doi: 10.1073/pnas.1012675108 [PubMed: 21402926]
- Zenil-Ferguson RJ, Burleigh G, and Ponciano JM (2018). Chromploid: an r package for chromosome number evolution across the plant tree of life. *Appl. Plant Sci* 6:e1037. doi: 10.1002/aps3.1037 [PubMed: 29732267]
- Zhimulev IF, Belyaeva ES, Semeshin VF, Koryakov DE, Demakov SA, Demakova OV, et al. (2004). Polytene chromosomes: 70 years of genetic research. *Int. Rev. Cytol* 241, 203–275. doi: 10.1016/s0074-7696(04)41004-3 [PubMed: 15548421]

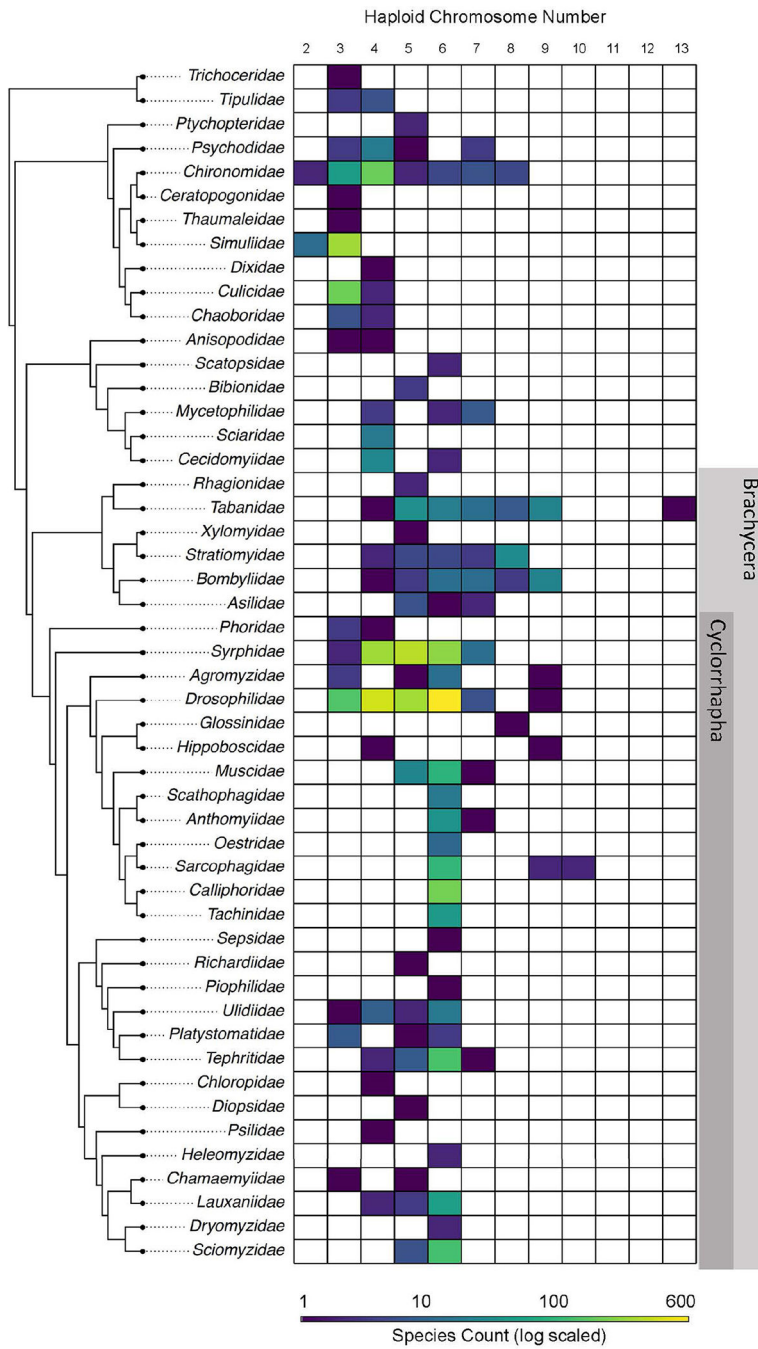


FIGURE 1 | Phylogenetic distribution of haploid chromosome count in Diptera. In each row the number of species with a specific haploid chromosome count is indicated by color. Note the color ramp is on a log scale to allow simultaneous visualization of families like Drosophilidae and Piophilidae which have orders of magnitude differences in the number of records available.

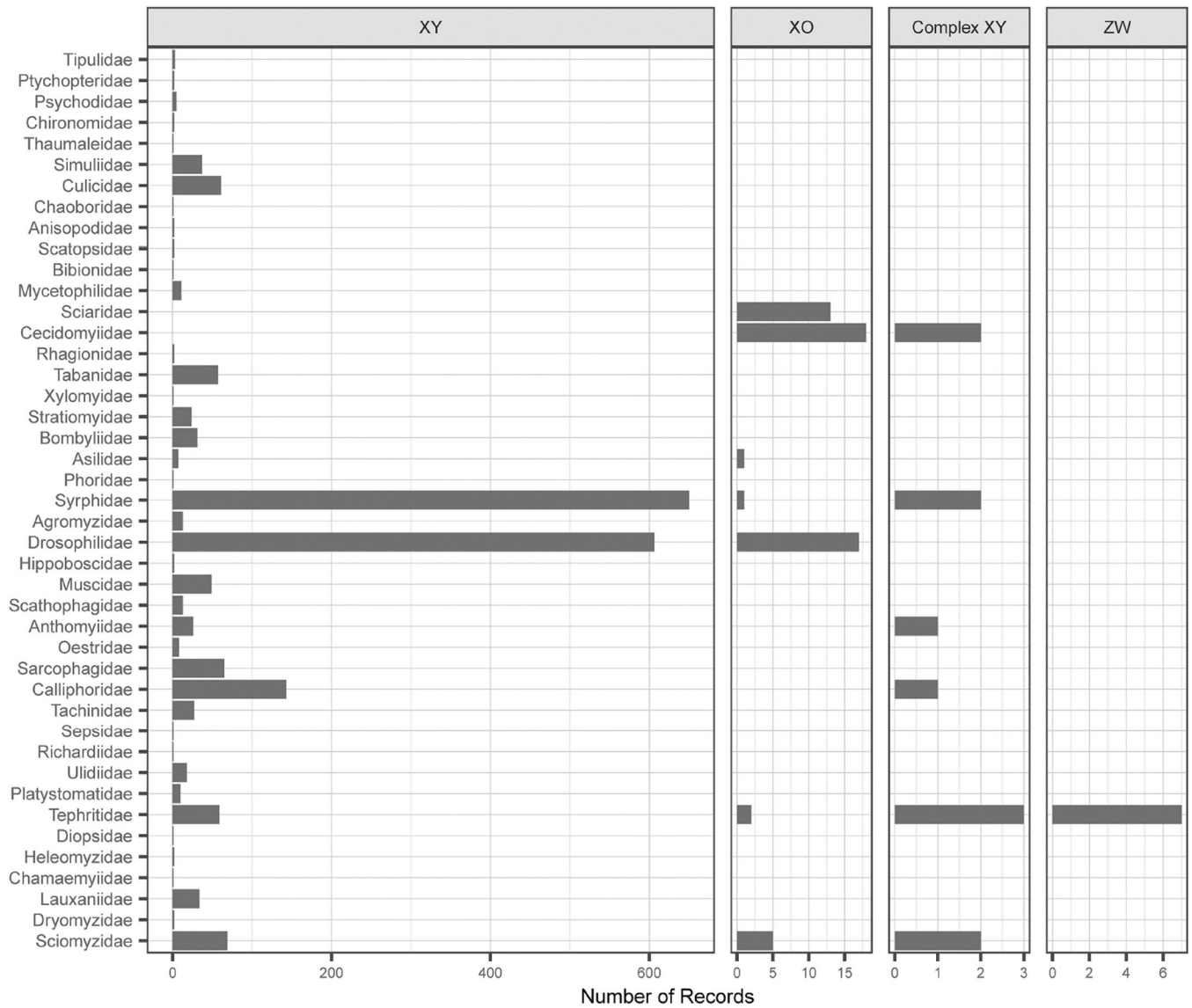


FIGURE 2 | Distribution of sex chromosome systems across families of Diptera. The order of families is the same as in Figure 1. Note scales on horizontal axis vary in size to allow visualization of rare sex chromosome systems.

TABLE 1 |

Traits recorded in karyotype databases.

Trait	States
Taxonomy	Family/Subfamily; Tribe; Genus; Species
Haploid Chromosome Number	2–13
Chromosome Shapes	Submetacentric; Metacentric; Subtelocentric; Telocentric; Subacrocentric; Acrocentric; Dot; Sex Chromosome Shape
Chromosome Shapes (<i>Drosophila</i> specific)	V for Metacentric; J for Submetacentric; Rod for Telocentric/Acrocentric
Sex Chromosome System	Complex XY; XO; XY; ZW; Monogenic
References	Citation for the sources of data included in the database