**The Probability of Fusions Joining Sex Chromosomes and Autosomes**

Nathan W. Anderson and Heath Blackmon

Department of Biology; Texas A&M University; College Station, TX 77843, USA

Author for correspondence: HB, *coleoguy@gmail.com*

# Abstract

Chromosome fusion and fission are primary mechanisms of karyotype evolution. In particular, the fusion of a sex chromosome and an autosome has been proposed as a mechanism to resolve intralocus sexual antagonism. If sexual antagonism is common throughout the genome, we should expect to see an excess of fusions that join sex chromosomes and autosomes. Here, we present a null model that provides the probability of a sex chromosome autosome fusion, assuming all chromosomes have an equal probability of being involved in a fusion. This closed-form expression is applicable to both male and female heterogametic sex chromosome systems and can accommodate unequal proportions of fusions originating in males and females. We find that over 25% of all chromosomal fusions are expected to be between a sex chromosome and an autosome whenever the diploid autosome count is fewer than 16, regardless of sex chromosome system. We also demonstrate the utility of our model by analyzing two empirical datasets: one from *Drosophila* and one from *Habronattus*.

*Keywords: sexual antagonism; chromosome fusion; sex determination systems; chromosome number*

# Introduction

The fusion and fission of chromosomes are two of the primary mechanisms that restructure the genome into discrete chromosomes (Blackmon et al. 2019). Early on, it was recognized that both fusions and fissions might be selectively favored because they modify linkage among loci (White 1977; Stebbins and others 1971). In particular, the fusion of a sex chromosome and an autosome (SA-fusion) has been proposed to resolve sexual antagonism. Sexual antagonism occurs when an allele is beneficial for one sex and deleterious for the other. Relocating sexually antagonistic alleles to sex chromosomes minimizes the deleterious effects of these alleles. Therefore, SA fusions are predicted to be more common than autosome autosome fusions (AA-fusions) (Charlesworth and Charlesworth 1980). Limited empirical examples have shown instances where autosomes, which are enriched for sexually antagonistic loci, have recently fused with sex chromosomes (Zhou and Bachtrog 2012). For instance, a recent fusion between the X chromosome and an autosome in *Drosophila americana* is proposed to have been driven by selection to reduce recombination between the sex determining locus and sexually antagonistic locus located on the autosome (McAllister 2003). Additionally, an apparent surplus in X chromosome autosome fusions in jumping spiders, *Habronattus*, is hypothesized to result from a mechanism of isolating male-beneficial sexually antagonistic alleles on the neo-Y chromosome (Maddison and Leduc-Robert 2013). Further empirical studies suggest that sexual antagonism may be common throughout the genome (Innocenti and Morrow 2010; Cheng and Kirkpatrick 2016). However, there remains significant debate on the ubiquity of sexually antagonistic variation (Kasimatis, Ralph, and Phillips 2019; Ponnikas et al. 2018). A strong measure of the frequency of significant sexually antagonistic variation across the genome would be an excess of SA-fusions relative to AA-fusions across large clades. We derive equations describing the probability of each type of fusion necessary to perform such a test and illustrate two approaches to using these equations with empirical datasets.

# The Model

The probability of SA-fusions is a function of the sex chromosome system and the number of autosomes in the genome. To facilitate tests of the balance between SA-fusions and AA-fusions, we have derived a closed form expression of the probability of a SA-fusion under a null model where any chromosome is equally likely to fuse with any other non-homologous chromosome. Our result is applicable to XO, XY and multi-XY (e.g. XXY or XXXYY) sex determination systems and, with slight modification, to ZW systems. We ignore fusions among homologous chromosomes, including fusions that join an X and Y chromosome, because this would lead to unbalanced gametes during meiosis and, presumably, these would be non-viable.

When any two chromosomes fuse, there are 3 possibilities. The two chromosomes could both be autosomes (AA-fusion), they could both be sex chromosomes (SS-fusion), or one could be a sex chromosome and the other an autosome (SA-fusion). We will denote our three possibilities as events , , and , respectively. Given that a fusion has occurred, we are interested in the probability it is a SA-fusion. Or, equivalently, we are interested in the expected proportion of all fusions which are SA-fusions. Unfortunately, this proves difficult to calculate directly. We avoid this using the complement rule. We define the probability that any given fusion is a SA-fusion as:

(1)

It is quite possible that the sexes may make unequal contributions to the fusions entering a species (Pennell et al. 2015). We include the term , representing the proportion of fusions that occur in females to account for this possibility. We use a subscript and for sire and dam when referring to sex specific values to avoid any confusion stemming from using subscript and . In the supplemental material we provide a step by step derivation but breifly we present the following expression for the expected proportion of fusions which occur between two sex chromosomes and two autosomes:

(2)

(3)

Given a species with diploid autosome count, X chromosome count in males, Y chromosome count in males, female diploid number, and male diploid number,

Each fraction represents the probability of two types of chromosomes fusing using a counting argument. For instance, the faction represents in [3] and can be more intuitively written as . The product of the probability of the first chromosome involved in a fusion being an autosome, the number of autosomes divided by the total number of chromosomes, and the probability the second chromosome involved is also an autosome. We subtract 2 from both the top and the bottom in the latter case to account for the first chromosome ‘chosen’ being unable to fuse with itself and its homolog.

Substituting these results into [1] and rearranging yields:

(4)

Equations [[2]](#eq11), [3], and [[4]](#eq13) have six parameters: , , , , and . We have eliminated one parameter, the number of X chromosomes in females, by noting . Although illustrated for male heterogametic systems, these formulations can be converted for use in ZW sex chromosome systems as well. Taking equations [[2]](#eq11), [3], and [[4]](#eq13) and exchanging and , replacing with ,replacing with , and replacing with , generates equations that provide probabilities for ZO, ZW, and multi-ZW systems. Additionally, setting (because there is no homogametic diploid individuals) and replacing both and Y for V generates equations which are accurate for UV sex chromosome systems, only if there are an equal number of U and V chromosomes. We have provided R functions that calculate P(SA), P(SS), and P(AA)in the R package evobirR (citation). When using these functions, the only arguments required are the diploid number, the sex chromosome system, and which is assumed to be 50% by default.

# Empirical Applications

*Habronotus:* In a recent study of jumping spiders, *Habronattus*, the large disparity between the number of SA-fusions (8-15) and AA-fusion (1) and SS-fusions (1) all in a system with 26 autosomes is presented as evidence that SA-fusions are being favored (Maddison and Leduc-Robert 2013). The intuition that this imbalance in the occurrence of fusions is unlikely can be rigorously tested with our null model. Using our equations [[2]](#eq11)-[[4]](#eq13), and a multinomial distribution, we are able to calculate the exact empirical p-value of having observed eight or more SA-fusions out of a total of 10 fusions. We assume an XXO sex chromosome system and a diploid autosome count of 26 (this karyotype was the most common in the ancestral state estimation performed in the study). . This confirms that *Habronattus* spiders do in fact have an excess of SA-fusions.

*Drosophila*: In the previous example, we calculated the expected proportion of the different types of fusions based on the ancestral, and most common, karyotype inferred in the *Habronattus* clade. However, across the entire clade, a variety of karyotypes exist. We envision the primary use of equation [[4]](#eq11) will be to calculate the expected proportion of fusions that are SA-fusions across large clades. We can do this by employing a biologically realistic Markov model of possible fusions and fissions (figure 2), and leveraging stochastic mappings generated under such a model to extract the proportion of time that lineages in a clade spent with each possible chromosome number and sex chromosome system (Huelsenbeck, Nielsen, and Bollback 2003; Revell 2012). These proportions can then be used in conjunction with equation [4] to generate a weighted sum that describes the expected proportion of all observed fusions that are SA-fusions (figure [3](#approach)). The resulting expected value can then be compared to the observed proportion of SA-fusions inferred from the stochastic mappings. An additional advantage of this approach is that it can incorporate uncertainty in both ancestral state reconstructions and phylogenetic history. To illustrate this approach, we analyzed data from *Drosophila*.

In the analysis below we used a dated ultrametric phylogeny of Drosophila from a recent study of genome size evolution (CITE Carls Underreplication paper). We use chromosome number and sex chromosome system data from the same paper and a comparative genomic analysis of 11 Drosophila species (CITE Schaeffer 2008 polytene...). This yielded a dataset consisting of 120 species with a diploid number ranging from 6 to 12. The sex chromosome system of eleven of the species was NeoXY while the remainder were XY. We performed 1000 stochastic mappings using the make.simmap function in the R package phytools (CITE Liam). Stochastic mapping was accomplished using a transition matrix that matches the Markov model presented in Figure 2. We performed a number of preliminary analyses where we assessed the impact of: 1) the prior placed on the root of the tree and 2) the inclusion of . We found that our results where qualitatively the same under all evaluated conditions. The results we present are based on fixing the root of the tree with a diploid number of 12 and an XY sex chromosome system, and including the parameter in the model. The prior on the root of the tree is supported by comparative genomic studies (CITE that same polytene paper Schaeffer). The inclusion of is based on our concern that some species may harbor an undocumented neoXY. We found that including in the model elevated our estimate of the proportion of SA fusions but not sufficiently to change the interpretation of the results.

Across our 1000 stochastic mappings we find that the average number of SA fusions observed is 4.49 and that this equates to a proportion of 0.155 (credible interval 0.12 – 0.22). Using our formula as described above we also calculated the expected proportion of SA fusions. The mean expected SA fusion proportion was 0.43 (credible interval 0.42-0.44). Comparing these distributions (figure 4) we find that they have zero overlap and that the empirical dataset shows far fewer SA fusions than would be expected by chance.

# Discussion

The need for a quantitative null model of the probability of SA-fusions is illustrated by examining the expected probability of SA-fusions across a range of observed chromosome numbers and sex chromosome systems. In figure [1](#autosomenum), we show that when the autosome number is small, a large proportion of fusions are expected to be SA-fusions even under a null model which assumes they are not selectively favored. In fact, for the XY sex chromosome system the probability of a given fusion being an SA-fusion does not drop below 25% until the diploid autosome count is greater than or equal to 16. In systems with XXY sex chromosomes, the case is even more extreme. The probability of SA-fusion does not drop below 25% until the diploid autosome count is greater than 22. Therefore, evaluating the proportion of SA-fusions and determining whether there is evidence for positive selection on these fusions can only be accomplished in light of a quantitative null model which takes account of chromosome number and sex chromosome system.

PARAGRAPH ABOUT DROSOPHILA RESULT MAKE SURE TO BRING IN CITATIONS TO THE COYNE CHARLESWORTH ETC PAPER WERE THEY THOUGHT ABOUT THIS APPROACH WITHOUT TOOLS WE HAVE.

We have developed a flexible equation used to calculate the probability of SA-fusions under common sex chromosome systems (male or female heterogametic). This model will allow for quantitative analyses of fusions across large clades and provide a way to test the long-standing hypothesis that SA-fusions are selectively favored for their ability to resolve sexual antagonism. In some clades where chromosome number is high (e.g. Lepidoptera and Isoptera) our model shows that SA-fusions should be rare (Blackmon, Ross, and Bachtrog 2017). In these cases, several SA-fusions within a clade may well suggest that these fusions are selectively favored. However, this model also shows that for clades with very few chromosomes (e.g. Diptera and Hemiptera), we should expect many SA-fusions even if they are not selectively favored (Blackmon, Ross, and Bachtrog 2017). Therefore, SA-fusions should only be considered as evidence for sexual antagonism when they occur at a higher rate than expected for the chromosome numbers and sex chromosome systems that have been present during the evolution of a clade.

# Appendix

Recall equation [1]:

(1)

It is quite possible that the sexes may make unequal contributions to the fusions entering a species (Pennell et al. 2015). Because of this, we must account for the differing probabilities in males and females. We have added the term , representing the proportion of fusions that occur in females to account for this possibility. We use a subscript and for sire and dam when referring to sex specific values to avoid any confusion stemming from using subscript and . While there is a detailed derivation in the appendix our result is essentially of the form:

(1)

We will find an expression for each of the probabilities. Each term is calculated using a counting argument. For instance, the probability of ‘choosing’ an autosome is the number of autosomes available over the total number of chromosomes available to be ‘chosen’. A graphical description of each counting problem is given in figure A1, in the case of a XXYYY sex chromosome system with 6 autosomes, but the generalization follows easily.

(A1)

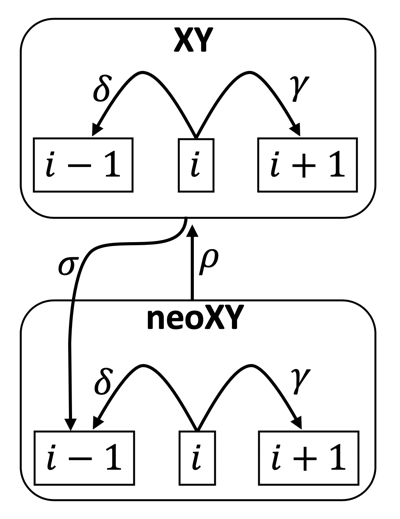
(A2)

(A3)

(A4)

Probability of a random fusion joining a sex chromosome and autosome. On the vertical axis we plot the proportion of all fusions that are SA-fusions while on the horizontal axis we plot the diploid autosome count. Each sex chromosome system is indicated by a unique color.

**Figure 1 Probability of a random fusion joining a sex chromosome and autosome.** On the vertical axis we plot the proportion of all fusions that are SA-fusions while on the horizontal axis we plot the diploid autosome count. Each sex chromosome system is indicated by a unique color.



**Figure 2 Model for the evolution of karyotype data in Drosophila.** At any instance in time a lineage will have chromosome and either an XY or neoXY sex chromosome system. A lineage can make four possible transitions: the fusion of two autosomes, the fission of an autosome, fusion of an autosome and a sex chromosome, and the transition from neoXY to XY.

Estimating P(SA) across a clade. On the left a stochastic map showing chromosome number and sex chromosome system. In the table on the right we have calculated the proportion of time that each state is present in the clade and then calculated P(SA) for each of these states. These P(SA) values along with the proportions are used to generate the expected P(SA) for the clade as a whole.

Figure 3 Estimating across a clade. On the left a stochastic map showing chromosome number and sex chromosome system. In the table on the right we have calculated the proportion of time that each state is present in the clade and then calculated for each of these states. These values along with the proportions are used to generate the expected for the clade as a whole.

Ashman, Tia-Lynn, Doris Bachtrog, Heath Blackmon, Emma E Goldberg, Matthew W Hahn, Mark Kirkpatrick, Jun Kitano, et al. 2014. “Tree of Sex: A Database of Sexual Systems.” *Scientific Data* 1: 140015.

Bachtrog, Doris, Judith E Mank, Catherine L Peichel, Mark Kirkpatrick, Sarah P Otto, Tia-Lynn Ashman, Matthew W Hahn, et al. 2014. “Sex Determination: Why so Many Ways of Doing It?” *PLoS Biol* 12 (7): e1001899.

Blackmon, Heath, Joshua Justison, Itay Mayrose, and Emma E Goldberg. 2019. “Meiotic Drive Shapes Rates of Karyotype Evolution in Mammals.” *Evolution* 73 (3): 511–23.

Blackmon, Heath, Laura Ross, and Doris Bachtrog. 2017. “Sex Determination, Sex Chromosomes, and Karyotype Evolution in Insects.” *Journal of Heredity* 108 (1): 78–93.

Charlesworth, D, and B Charlesworth. 1980. “Sex Differences in Fitness and Selection for Centric Fusions Between Sex-Chromosomes and Autosomes.” *Genetics Research* 35 (2): 205–14.

Cheng, Changde, and Mark Kirkpatrick. 2016. “Sex-Specific Selection and Sex-Biased Gene Expression in Humans and Flies.” *PLoS Genetics* 12 (9).

Hsu, Tao C, and Kurt Benirschke. 2013. *An Atlas of Mammalian Chromosomes*. Vol. 10. Springer Science & Business Media.

Huelsenbeck, John P, Rasmus Nielsen, and Jonathan P Bollback. 2003. “Stochastic Mapping of Morphological Characters.” *Systematic Biology* 52 (2): 131–58.

Innocenti, Paolo, and Edward H Morrow. 2010. “The Sexually Antagonistic Genes of Drosophila Melanogaster.” *PLoS Biology* 8 (3).

Kasimatis, Katja R, Peter L Ralph, and Patrick C Phillips. 2019. “Limits to Genomic Divergence Under Sexually Antagonistic Selection.” *G3: Genes, Genomes, Genetics* 9 (11): 3813–24.

Maddison, Wayne P, and Geneviev̀e Leduc-Robert. 2013. “Multiple Origins of Sex Chromosome Fusions Correlated with Chiasma Localization in Habronattus Jumping Spiders (Araneae: Salticidae).” *Evolution* 67 (8): 2258–72.

McAllister, Bryant F. 2003. “Sequence Differentiation Associated with an Inversion on the Neo-X Chromosome of Drosophila Americana.” *Genetics Society of America* 165 (3): 1317–28.

Pennell, Matthew W, Mark Kirkpatrick, Sarah P Otto, Jana C Vamosi, Catherine L Peichel, Nicole Valenzuela, and Jun Kitano. 2015. “Y Fuse? Sex Chromosome Fusions in Fishes and Reptiles.” *PLoS Genetics* 11 (5).

Ponnikas, Suvi, Hanna Sigeman, Jessica K Abbott, and Bengt Hansson. 2018. “Why Do Sex Chromosomes Stop Recombining?” *Trends in Genetics* 34 (7): 492–503.

Revell, Liam J. 2012. “Phytools: An R Package for Phylogenetic Comparative Biology (and Other Things).” *Methods in Ecology and Evolution* 3 (2): 217–23.

Stebbins, George Ledyard, and others. 1971. “Chromosomal Evolution in Higher Plants.” *Chromosomal Evolution in Higher Plants.*

White, Michael James Denham. 1977. *Animal Cytology and Evolution*. CUP Archive.

Zhou, Qi, and Doris Bachtrog. 2012. “Sex-Specific Adaptation Drives Early Sex Chromosome Evolution in Drosophila.” *Science* 337 (6092): 341–45.