

Original Article

Frequencies of house fly proto-Y chromosomes across populations are predicted by temperature heterogeneity within populations

Patrick D. Foy^{1,12}, Sara R. Loetzerich¹, David Boxler², Edwin R. BurgessIV^{3, ID}, R.T. Trout Fryxell^{4, ID}, Alec C. Gerry⁵, Nancy C. Hinkle⁶, Erika T. Machtlinger^{7, ID}, Cassandra Olds⁸, Aaron M. Tarone⁹, Wes Watson¹⁰, Jeffrey G. Scott^{11, ID} and Richard P. Meisel^{1,*, ID}

¹Department of Biology and Biochemistry, University of Houston, Houston, TX 77204, USA,

²Institute of Agriculture and Natural Resources, University of Nebraska, Lincoln, NE 69101, USA,

³Department of Entomology and Nematology, University of Florida, Gainesville, FL 32610, USA,

⁴Department of Entomology and Plant Pathology, University of Tennessee, Knoxville, TN 37996, USA,

⁵Department of Entomology, University of California Riverside, Riverside, CA 92521, USA,

⁶Department of Entomology, University of Georgia, Athens, GA 30602, USA,

⁷Department of Entomology, The Pennsylvania State University, University Park, PA 16802, USA,

⁸Department of Entomology, Kansas State University, Manhattan, KS 66506, USA,

⁹Department of Entomology, Texas A&M University, College Station, TX 77843, USA,

¹⁰Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC 27695, USA,

¹¹Department of Entomology, Cornell University, Ithaca, NY 14853, USA,

¹²Current affiliation: Department of Biology, Indiana University, Bloomington, IN 47405, USA

*Corresponding author: Email: rpmeisel@uh.edu

Corresponding Editor: Melissa Wilson

Abstract

Sex chromosomes often differ between closely related species and can even be polymorphic within populations. Species with multifactorial sex determination segregate for multiple different sex-determining loci within populations, making them uniquely informative of the selection pressures that drive the evolution of sex chromosomes. The house fly (*Musca domestica*) is a model species for studying multifactorial sex determination because male-determining genes have been identified on all six of the chromosomes, which means that any chromosome can be a “proto-Y.” Natural populations of house flies also segregate for a recently derived female-determining locus, meaning house flies also have a proto-W chromosome. The different proto-Y chromosomes are distributed along latitudinal clines on multiple continents, their distributions can be explained by seasonality in temperature, and they have temperature-dependent effects on physiological and behavioral traits. It is not clear, however, how the clinal distributions interact with the effect of seasonality on the frequencies of house fly proto-Y and proto-W chromosomes across populations. To address this question, we measured the frequencies of house fly proto-Y and proto-W chromosomes across nine populations in the United States of America. We confirmed the clinal distribution along the eastern coast of North America, but it is limited to the eastern coast. In contrast, the annual mean daily temperature range predicts proto-Y chromosome frequencies across the entire continent. Our results therefore suggest that temperature heterogeneity can explain the distributions of house fly proto-Y chromosomes in a way that does not depend on the cline.

Key words: Diptera, insect, *Musca domestica*, polygenic sex determination

Introduction

Sex chromosomes and sex-determining genes often differ between species, even those that are closely related (Bachetrog et al. 2014; Beukeboom and Perrin 2014). Evolutionary changes in sex chromosomes typically occur via one of the two mechanisms (Abbott et al. 2017). First, an autosome can fuse to a sex chromosome, turning the autosome into a neo-sex chromosome. Second, an autosome can acquire a new master sex-determining locus, turning the autosome into a proto-sex

chromosome (and allowing the ancestral sex chromosome to “revert” to an autosome). Sex-specific selection pressures are thought to be important for the invasion and fixation of both neo- and proto-sex chromosomes because sex-linkage can resolve intralocus sexual antagonism (Charlesworth and Charlesworth 1980; Roberts et al. 2009; Mank et al. 2014; van Doorn and Kirkpatrick 2007; van Doorn 2014). In addition, if sex ratios are distorted from their evolutionary stable equilibrium, a new sex determiner on a proto-sex

chromosome can be favored if it increases the frequency of the sex that is below its equilibrium value (Bull 1983; Werren and Beukeboom 1998). Notably, ecological factors can modulate the effects of these sex-specific selection pressures, but the extent to which ecological selection pressures drive sex chromosome evolution are not yet resolved (Meisel 2022).

We used the house fly (*Musca domestica*) as a model species to explore how ecological factors affect sex chromosome evolution. House fly is well-suited for this purpose because it has a highly polymorphic multifactorial sex determination system (Hamm et al. 2015). Male-determining factors have been genetically mapped to all six of the house fly chromosome pairs, and a single gene (*Mdmd*) has been implicated as the male-determiner on at least four of the six chromosomes (Sharma et al. 2017). Each of these *Mdmd*-bearing chromosomes is a young proto-Y chromosome (Meisel et al. 2017; Son and Meisel 2021). Nearly every male housefly in North America carries one or both the two most abundant proto-Y chromosomes, the Y chromosome (Y^M) and third chromosome (III^M), and males with other proto-Y chromosomes are rarely found (Hamm et al. 2015). In North America, and throughout the world, males can additionally carry multiple proto-Y chromosomes and even be homozygous for a proto-Y chromosome (e.g. Y^M/Y^M or III^M/III^M genotype), demonstrating that these chromosomes are not fixed for recessive lethal alleles found in other Y chromosomes (Hamm and Scott 2009; Hamm et al. 2015).

Ecological factors are thought to create heterogeneous selection pressures responsible for the maintenance of multifactorial sex determination across house fly populations. Y^M and III^M form latitudinal clines in Europe, Japan, and North America, with Y^M most common in northern populations and III^M predominating in the south (Franco et al. 1982; Denholm et al. 1986; Tomita and Wada 1989; Hamm et al. 2005). Consistent with this geographical distribution, the Y^M chromosome confers greater tolerance to extreme cold and preference for cooler temperatures, while III^M confers improved tolerance to extreme heat and preference for warmer temperatures (Delclos et al. 2021). In addition, higher Y^M frequency is associated with locations with low seasonality of temperatures, or small differences between minimum and maximum values of monthly high and low temperatures (Feldmeyer et al. 2008).

Sex ratio selection is also thought to affect the evolution of multifactorial sex determination in house flies. In some house fly populations, males can carry multiple proto-Y chromosomes (e.g. both Y^M and III^M , homozygous for Y^M or III^M), which could create male-biased sex ratios and selection in favor of a female-determining factor (Bull and Charnov 1977; Bulmer and Bull 1982; Eshel 1975). Indeed, such a female-determiner exists in house fly populations, in the form of a dominant allele of the house fly ortholog of *transformer* (*Md-tra^D*) on chromosome IV, which causes embryos to develop into females even if they carry multiple male-determining chromosomes (Hediger et al. 2010). Only a single haplotype of the *Md-tra^D* allele has been found throughout the world, suggesting that the allele is of recent origin (Scott et al. 2014). Chromosome IV with the *Md-tra^D* allele is therefore a recently derived proto-W chromosome. The frequency of *Md-tra^D* is correlated with the frequency of males with multiple male-determining chromosomes across populations, suggesting that there is a selection for balanced sex ratios (Meisel et al. 2016).

We aimed to test if the frequencies of Y^M , III^M , and *Md-tra^D* across North American populations could be explained by climatic variables. Previous studies linking the frequencies of male-determining chromosomes to climatic variation in North America have been limited to a latitudinal cline along the eastern coast (Hamm et al. 2005; Hamm and Scott 2008), and only Japanese and African populations were sampled to study how *Md-tra^D* frequencies vary across climates (Feldmeyer et al. 2008). However, the frequencies of Y^M , III^M , and *Md-tra^D* vary in non-clinal patterns across regions of North America outside the eastern coast (McDonald et al. 1975; Meisel et al. 2016), suggesting that different climatic variables may predict their distributions outside the cline. To address this question, we genotyped male and female house flies from nine different locations across the United States of America, and we tested if the frequencies of Y^M , III^M , and *Md-tra^D* were associated with a variety of climatic variables.

Materials and methods

House fly collections

House flies were collected from dairy farms, poultry farms, and other locations where flies are present in nine different populations across the United States of America (Supplementary Table S1). Collections were performed in May–June 2021 using sweep nets. Flies were allowed to lay eggs in laboratories near the collection sites. The resulting pupae were then shipped to Cornell University (Ithaca, NY), where colonies from each collection site were established. Pupae from each of those colonies were then shipped to the University of Houston (Houston, TX) within four generations of establishing the laboratory colonies. Previous work found that III^M increases in frequency, and Y^M decreases, in laboratory rearing conditions (Hamm et al. 2009), suggesting that our estimates of III^M or Y^M frequencies may be over- or under-estimates, respectively. Pupae that were shipped to Houston were raised into adults in laboratory conditions (22 °C), and the emerging adults were frozen for genotyping.

DNA extraction and genotyping

DNA was extracted from individual frozen house fly heads using the hot sodium hydroxide and tris, HotSHOT, protocol (Truett et al. 2000). We performed PCR to test for the presence of Y^M using the A12CMF1 and A12CMR1 primer pair (Hamm et al. 2009). We were unable to design a PCR primer pair that could reliably identify the III^M chromosome. We used the GM2IIIF1 and GM2IIIR2 primer pair as a positive control to confirm successful DNA amplification from males (Hamm et al. 2009). We tested for *Md-tra^D* in females using a primer pair that amplifies a region of exon 3 of the *Md-tra* gene containing the diagnostic deletion (Hediger et al. 2010; Meisel et al. 2016). We tested for the presence of *Mdmd* in females using the *Mdmd*_F1 and *Mdmd*_R4 primer pair (Sharma et al. 2017).

Each male was genotyped for the presence of Y^M . From these data, we determined the frequencies of males with Y^M and males without Y^M (Fig. 1A). The latter group (males without Y^M) presumably consists of III^M males, but some males with Y^M may also carry III^M . We used population genetic simulations (see below) to estimate the frequency of III^M along with the frequencies of all possible male genotypes in each population (Fig. 1B).

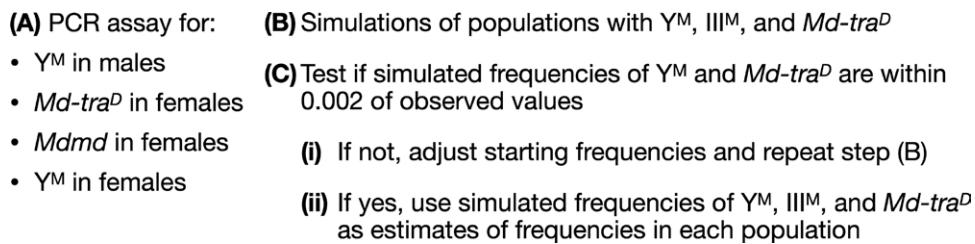


Figure 1. Approach to estimate the frequencies of Y^M , III^M , and $Md-tra^D$ in each of the nine sampled populations. **(A)** PCR was used to determine the frequencies of proto-sex chromosomes and sex-determining loci/alleles. **(B)** Simulations were performed to determine equilibrium frequencies of proto-sex chromosomes in randomly mating populations. **(C)** If those equilibrium frequencies deviated from the observed frequencies (i), then the starting frequencies were adjusted and the simulations were repeated. If the equilibrium frequencies were similar to the observed frequencies in a population (ii), then the equilibrium frequencies were used as estimates of the frequencies of Y^M , III^M , and $Md-tra^D$ in a population.

Each female was genotyped for the presence of $Md-tra^D$, $Mdmd$, and Y^M . For each population, we determined the frequencies of: females without $Md-tra^D$; females with $Md-tra^D$ but not $Mdmd$; females with $Md-tra^D$ and Y^M ; and females with $Md-tra^D$ and $Mdmd$ but not Y^M (Fig. 1A). We also used these data to determine the frequencies of $Md-tra^D$, $Mdmd$ in females, and $Mdmd$ in $Md-tra^D$ females.

Population genetic simulations to determine genotype frequencies

There are 18 possible genotypes when considering all possible combinations of Y^M , III^M , and $Md-tra^D$ (Meisel 2021). Eight of the genotypes are male-determining, and they differ in whether the males carry zero, one, or two copies of Y^M and III^M (all males must carry at least one copy of either Y^M or III^M). The remaining 10 genotypes are female determining. One female genotype carries none of the Y^M , III^M , or $Md-tra^D$ chromosomes; the remaining nine female genotypes carry one copy of $Md-tra^D$ along with zero, one, or two copies of Y^M and III^M . Homozygosity for Y^M or III^M is possible because these chromosomes are not fixed for recessive lethal mutations (Hamm et al. 2015).

We used population genetic simulations to predict the frequencies of the 18 possible sex chromosome genotypes in each sampled population (Fig. 1B). Our PCR genotyping method is unable to detect the III^M chromosome nor is it able to diagnose specific genotypes, which means our genotype data are incomplete. To help overcome the deficiency in our genotyping protocol, we performed population genetic simulations to identify genotype frequencies that would produce the frequencies of Y^M and $Md-tra^D$ observed in our data (Fig. 1C). Our simulations used the same recursion equations that we have previously used to model randomly mating house fly populations, without selection (Meisel et al. 2016). We performed separate simulations for each of the nine sampled populations in order to estimate the frequencies of Y^M , III^M , $Md-tra^D$, and each genotype in each population (Supplementary Material on Dryad for code).

In the simulations, we calculated the frequency of a chromosome (Y^M or III^M) or allele ($Md-tra^D$) as follows. The frequency of Y^M was calculated as the number of Y^M chromosomes in a population divided by the sum of the number of Y^M and X chromosomes in that population. The frequency of the III^M chromosome was calculated as the number of third chromosomes with $Mdmd$ (i.e. III^M) divided by the total number of third chromosomes. The frequency of $Md-tra^D$ was calculated as the number of $Md-tra^D$ alleles divided by the total number of $Md-tra$ genes. The frequencies

of Y^M and III^M can take values between 0 and 1, while the $Md-tra^D$ frequency can take values between 0 and 0.25 (because it is a W chromosome in a randomly mating population with equal sex ratios).

We started each simulation with estimates of the frequencies of Y^M , III^M , and $Md-tra^D$ from our PCR assay. The following calculations were used to determine initial frequencies in our simulations, and the assumptions used only apply to the calculations of these initial frequencies. The initial frequency of Y^M (f_{YM}) was estimated as half of the frequency of males carrying a Y^M chromosome. This initial frequency assumes that all males carrying Y^M also carry an X chromosome, and all copies of Y^M are found in heterozygous individuals. The initial frequency of III^M (f_{III^M}) was estimated as $\frac{1}{2} - f_{YM}$, which assumes all males without Y^M carry one copy of III^M . Both of these calculations assume that f_{YM} and f_{III^M} are equal in males and females. The initial frequency of $Md-tra^D$ (f_{tra^D}) was estimated as one-quarter of the number of females carrying $Md-tra^D$, which requires no assumptions about genotypes because each female carrying $Md-tra^D$ must be heterozygous and males cannot carry $Md-tra^D$.

From the initial estimated frequencies of Y^M , III^M , and $Md-tra^D$, we calculated initial frequencies of each of the 18 possible genotypes. We first calculated the initial frequencies of each single chromosome genotype assuming random mating. For example, the frequency of the X/X genotype was estimated as $(1 - f_{YM})^2$; the X/ Y^M frequency was estimated as $2f_{YM}(1 - f_{YM})$; and the Y^M/Y^M frequency was estimated as f_{YM}^2 . All three genotypes are possible because homozygous Y^M/Y^M are viable and found in nature. Similar calculations were performed to estimate the frequencies of the third chromosome genotypes. The frequency of the $Md-tra^D/Md-tra^+$ genotype was estimated as $f_{tra^D}(1 - f_{tra^D})$, and the frequency of the $Md-tra^+/Md-tra^+$ genotype was estimated as $(1 - f_{tra^D})^2$. The initial frequencies of the 18 multi-chromosome genotypes were then calculated using the product of each single chromosome genotype, and each of the 18 frequencies were divided by the sum to obtain a new estimate that summed to one. The assumptions used to calculate the initial chromosome and genotype frequencies only apply to the initial frequencies used in the simulations.

We performed simulations for 10 generations of random mating using those initial genotype frequencies and previously developed recursion equations (Meisel et al. 2016) to determine the equilibrium frequencies of each chromosome and genotype, given the initial genotype frequencies (see code in Supplementary Material). We found that 10 generations were sufficient time for the simulated populations to reach

equilibrium frequencies without changes from one generation to the next (Supplementary Fig. S1-S9). We compared the resulting values of f_{Y^M} and f_{tra^D} after 10 generations with the observed values measured in the respective natural population. We tested if the simulated values were within 0.002 of the observed values (Fig. 1C). If the simulated frequency of a chromosome was less than the observed frequency, we increased the initial frequency and repeated the simulation. Conversely, if the simulated frequency was greater than the observed frequency, we decreased the initial frequency and repeated the simulation. We repeated this process until the simulated frequencies of f_{Y^M} and f_{tra^D} matched the observed frequencies within 0.002. We used these simulated frequencies at equilibrium (i.e. after 10 generations) as estimates of the chromosome and genotype frequencies in downstream analyses.

Climate data

We tested if climatic features were associated with the frequencies of sex chromosomes and genotypes across the sampled populations. To do so, we obtained weather data from the nearest NOAA station to the collection site measured between 1991 and 2020 (Table 1; Supplementary Table S1). From these data, we extracted many of the same features as a previous analysis comparing the frequencies of house fly sex chromosomes and climatic data (Feldmeyer et al. 2008), and we used annual precipitation (Precip) instead of humidity measurements. All temperatures were provided in Fahrenheit, and we converted the values to Celsius for analysis.

We performed a principal component analysis (PCA) to test if climate features were associated with the frequencies of Y^M , III^M , $Md\text{-}tra^D$, males with multiple proto-Y chromosomes, and males with both Y^M and III^M . To those ends, we used the `prcomp()` function in R to perform a PCA on the annual climate data, with variables scaled to have unit variance. We then tested if each principal component (PC) is correlated with Y^M or III^M frequencies across the population. All code to perform these analyses is available in [Supplementary Material on Dryad](#).

Results

We used PCR assays to determine the frequencies of male house flies carrying the Y^M chromosome across nine locations (populations) sampled in 2021 in the United States of America

Table 1. Climate features analyzed.

| Name | Description |
|----------------------------|---|
| T_{mean} | Annual mean average daily temperature |
| T_{min} | Annual mean minimum daily temperature |
| T_{max} | Annual mean maximum daily temperature |
| Daily_{TR} | Annual mean daily temperature range |
| Precip | Annual precipitation |
| T_{active} | The mean temperature of the warmest month |
| Season_1 | Coefficient of variation of average monthly temperatures |
| Season_2 | Average difference between highest and lowest monthly maximum temperature and difference between highest and lowest monthly minimum temperature |

(Fig. 2A). We also used PCR to determine the frequencies of female house flies carrying $Md\text{-}tra^D$, $Mdmd$, and Y^M in the same nine populations (Fig. 2A). We then performed population genetic simulations to identify frequencies of Y^M , III^M , and $Md\text{-}tra^D$ in each population that could produce the observed frequencies we observed in our PCR assays (Fig. 1; Supplementary Fig. S1-S9). The III^M chromosome was at the highest frequency in two of the three southernmost populations (CA and FL). In the FL population, males were predicted to be almost entirely III^M , and there were very few $Md\text{-}tra^D$ females. In contrast, the northernmost population (PA) was predicted to have almost entirely Y^M males. Our simulations produced a positive correlation between the frequencies of males with multiple male-determining chromosomes (Y^M and/or III^M) and females carrying $Md\text{-}tra^D$ ($r^2 = 0.975$, $P = 4.5 \times 10^{-7}$; Fig. 2C), as expected in natural populations where sex ratio selection maintains equal numbers of males and females (Meisel et al. 2016).

We compared our estimates of the frequencies of Y^M , III^M , $Md\text{-}tra^D$, and four sex chromosome genotypes in the CA and NC populations with previous measurements in nearby populations from California and North Carolina (Table 2). A population was sampled from Chino, CA in 1982 and 2014 (Meisel et al. 2016), ~50 km from our CA collection site (which we sampled in 2021). There was a high frequency of $Md\text{-}tra^D$ in both populations. However, the Chino population had a higher frequency of Y^M chromosomes, while the CA population we sampled in 2021 had a higher III^M frequency. In contrast, we observed similar frequencies of Y^M , III^M , and $Md\text{-}tra^D$ in the NC population we sampled in 2021 and the populations sampled in 2002, 2006, and 2007. All of the NC populations were sampled in close proximity within Wake County.

We tested for associations between climatic variables and the frequencies of sex chromosomes across the nine populations we sampled. To those ends, we performed a PCA using eight climate features measured across the nine populations. The first three PCs explain >98% of the variance in the data (Supplementary Table S2), with PC1 and PC2 explaining nearly 90% of the variance (Fig. 3A). PC1 captured variation in seasonality (Season_1 and Season_2), along with minimum, maximum, and average temperatures (T_{mean} , T_{min} , T_{max} , and T_{active}) across populations. PC2 captured variation in precipitation (Precip) and annual mean daily temperature range (Daily_{TR}) across populations. We tested for correlations between PC1 or PC2 and the frequencies of sex chromosomes across populations. Only PC2 and Y^M frequency had a significant correlation, with Y^M frequency decreasing as PC2 values increased ($\rho = -0.767$, $P = 0.0214$). We additionally constructed linear models in which we tested if PC1, PC2, and their interaction predicted the frequencies of Y^M or III^M . The only significant relationship in these models was between PC2 and III^M frequency ($F = 8.372$, $P = 0.034$). Therefore, there is evidence that Y^M and III^M frequencies across populations were associated with PC2, which captured variation in precipitation and daily temperature range. These relationships between PC2, Daily_{TR} , and Y^M/III^M frequencies can be seen in plots of Y^M or III^M frequency against Daily_{TR} (Fig. 3B-C).

The CA population appeared to be an outlier in many respects, which could have driven some of the patterns we observed. For example, the CA population had higher frequencies of III^M chromosomes, males with multiple male determiners, and females with $Md\text{-}tra^D$, when compared to

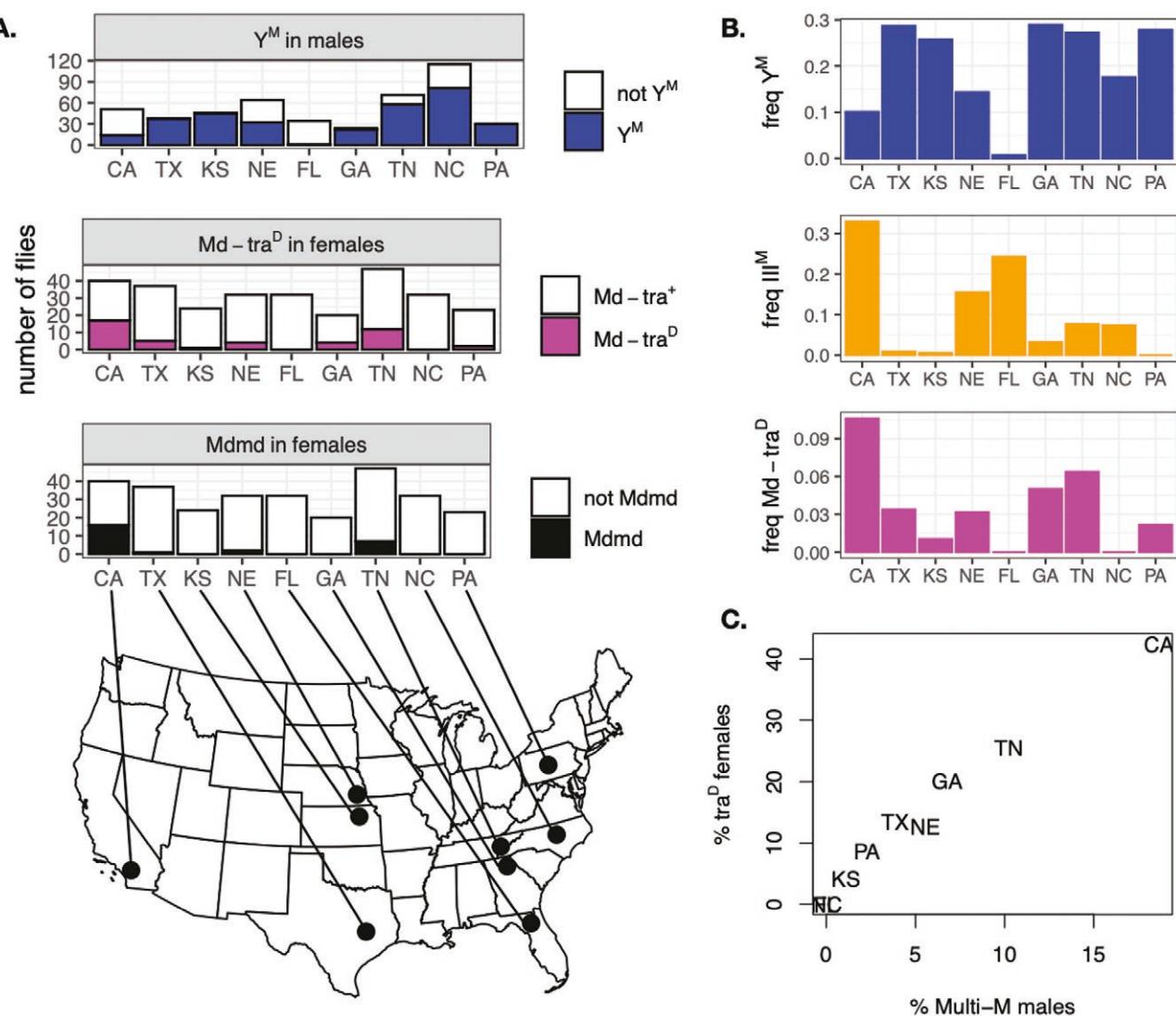


Figure 2. Observed and inferred frequencies of sex chromosomes, determiners, and alleles across nine populations. **(A)** The number of males genotyped with Y^M (blue bars), females with $Md\text{-}tra^D$ (magenta bars), and females with $Mdmd$ (black bars) from each of the nine populations are plotted. The number of flies without each chromosome, allele, or gene is shown in white bars. The sampling locations for each population are indicated by dots on the map. **(B)** The estimated frequencies of Y^M , III^M , and $Md\text{-}tra^D$ in each of the nine populations are plotted. Estimated frequencies are from population genetics simulations which produced observed frequencies shown in panel A. **(C)** The relationship between the percent of females carrying $Md\text{-}tra^D$ and the percent of males with multiple male-determining chromosomes (Y^M and/or III^M) are plotted for the nine populations based on estimated genotype frequencies.

Table 2. Frequencies of sex chromosomes and genotypes in CA and NC populations sampled in different years.

| Genotype/allele | Sex | Percent in population (%) | | | | | | |
|--|--------|---------------------------|-------------------------|---------|----------------------|----------------------|----------------------|---------|
| | | Chino 1982 ^a | Chino 2014 ^a | CA 2021 | NC 2002 ^b | NC 2006 ^c | NC 2007 ^c | NC 2021 |
| X/Y ^M ; III/III | male | 68.0 | 76.0 | 16.2 | 77.7 | 77.8 | 95.3 | 70.4 |
| X/X; III ^M /III | male | 0.0 | 8.0 | 65.2 | 20.0 | 19.4 | 2.3 | 29.6 |
| X/Y ^M ; III ^M /III | male | 12.0 | 4.0 | 8.8 | 2.3 | 1.4 | 0.0 | 0.0 |
| Y ^M /Y ^M ; III/III | male | - | 12.0 | 0.4 | 0.0 | 1.4 | 2.3 | 0.0 |
| Y ^M | male | - | 92.0 | 27.4 | 80.0 | 80.6 | 97.6 | 70.4 |
| III ^M | male | - | 12.0 | 83.4 | 22.3 | 20.8 | 2.3 | 29.6 |
| $Md\text{-}tra^D$ | female | - | 33.6 | 42.4 | 5.9 | - | 4.2 | 0.0 |

(a)Meisel et al. (2016).

(b)Hamm et al. (2005).

(c)Hamm and Scott (2008).

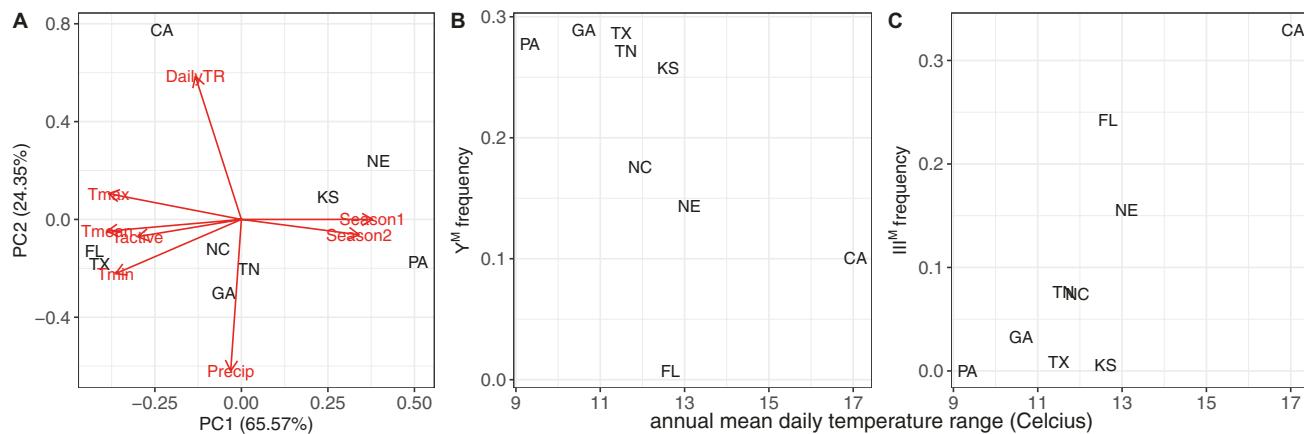


Figure 3. Associations between climate features and proto-Y chromosome frequencies across the nine sampled populations. Each population is represented by the two-letter abbreviation of the state from which it was collected. **(A)** Populations are plotted according to the first two principal components (PCs) based on climate features. The variance explained by each PC is shown along each axis. The loadings of each climate feature are indicated by labeled vectors. Vector labels are described in Table 1. **(B-C)** The relationships between the predicted frequency of Y^M or III^M and the annual mean daily temperature range are plotted for each population.

all other populations (Fig. 2). In addition, the CA site was an outlier along PC2 because it had a higher daily temperature range and lower precipitation than the other populations (Fig. 3). When we excluded the CA site from our climate PCA, we observed similar loadings of the climate variables: PC1 explained 71.69% of the variance and captured variation in seasonality and minimum/maximum temperature, while PC2 explained 22.13% of variance and captured variation in daily temperature range and precipitation (Supplementary Fig. S10). We also observed a significant negative correlation between daily temperature range and Y^M frequency ($\rho = -0.881$, $P = 0.007$) even when the CA population was excluded. Therefore, the relationships between climate features and proto-Y chromosome frequencies were not driven solely by the CA population.

Discussion

We observed substantial variation in the frequencies of Y^M , III^M , and $Md\text{-}tra^D$ across populations of house flies in North America (Fig. 2). Along the eastern coast, Y^M was most common in the north (PA), III^M was most common in the south (FL), and both Y^M and III^M were found in the central (NC) population, consistent with the previously documented cline (Hamm et al. 2005). However, moving west, we found that the clinal distribution eroded, and latitude was not associated with the frequencies of Y^M and III^M . For example, the GA, TN, NE, and CA populations all had moderate to high frequencies of III^M , Y^M , and $Md\text{-}tra^D$, without any relationship to latitude. In addition, the presence of all three chromosomes/alleles in CA is consistent with previous observations (Meisel et al. 2016).

Two components of our data collection likely affected the accuracy with which we measured the frequencies of Y^M , III^M , $Md\text{-}tra^D$, and the 18 possible genotypes. First, we sampled flies after multiple generations of lab breeding, which could lead to deviations from the frequencies in natural populations. Second, we used population genetic simulation models to estimate chromosome and genotype frequencies in each population based on PCR assays for the presence of Y^M , $Md\text{-}md$, and $Md\text{-}tra^D$ from the individual flies we sampled (Fig. 1).

Our PCR assays and simulations likely do not measure allele, chromosome, and genotype frequencies as accurately as more direct genotyping assays that were used in prior studies (Hamm et al. 2005; Feldmeyer et al. 2008; Meisel et al. 2016). We also assumed that Y^M and III^M were the only possibly male-determining, proto-Y chromosomes in North America, which is consistent with prior observations in populations sampled from across the continent (Hamm et al. 2015; Meisel et al. 2016).

Despite the limitations in our experimental design, we believe that our estimates of allele, chromosome, and genotype frequencies were sufficiently accurate for the analyses we performed. One evidence for accuracy is that we detect substantial differences in frequencies across populations (Fig. 2), which suggests that shared laboratory selection did not drive the convergent evolution of frequencies for all collections. In addition, our estimated frequencies of Y^M , III^M , and $Md\text{-}tra^D$ are largely concordant with prior estimates from the same county in North Carolina (Table 2). Moreover, previous work found that Y^M is more common than III^M in Texas (McDonald et al. 1975), consistent with our results (Fig. 2). Furthermore, lab selection under similar conditions was previously shown to increase the frequency of the III^M chromosome (Hamm et al. 2009), yet we were still able to detect high frequencies of Y^M in many of the populations (Fig. 2).

Despite the concordance with prior results, there are discrepancies between the frequencies we predicted and those previously observed in California. The CA population we sampled was estimated to have a high frequency of III^M (Fig. 2B), but previous collections from a nearby population found that Y^M was at a higher frequency than III^M (Meisel et al. 2016). This discrepancy between our observations and prior measurements from California can likely be explained by a >50 km distance between our site and the previous sampling locations. Similar differences in the frequencies of house fly male-determining chromosomes have been observed over relatively short distances in Japan and Spain (Tomita and Wada 1989; Li et al. 2022). Therefore, small-scale variations in Y^M and III^M frequencies appear to be a global phenomenon, in addition to the large-scale variation observed across the entire continent (Fig. 2).

Our results contribute to the body of evidence that climatic factors affect the frequencies of Y^M and III^M in natural populations of house flies. In addition to the clinal distribution we confirmed in eastern North America (Fig. 2), we also detected consistent associations between Y^M or III^M frequencies, annual mean daily temperature range (Daily_{TR}), and precipitation (Fig. 3). In particular, we predicted that Y^M was at the highest frequency when daily temperature heterogeneity was lowest, and III^M frequency was higher with more daily temperature variation. The same general patterns were observed when the CA population, which had extreme values relative to other populations, was excluded.

Our results differ from a prior observation that the frequencies of non- Y^M , male-determining chromosomes (e.g. III^M) in Africa and Europe were higher when seasonality in temperature was highest (Feldmeyer et al. 2008). Feldmeyer et al. (2008) measured seasonality as the difference between the minimum and maximum values of the monthly minimum and maximum temperatures. We found this measure of seasonality was orthogonal to daily temperature range (Fig. 3A). However, both seasonality and daily temperature range are measures of temperature heterogeneity across time, suggesting that temperature variation more generally may be an important selection pressure that affects proto-Y chromosome frequencies across house fly populations.

There is growing evidence that ecological factors contribute to sex chromosome evolution (Meisel 2022). Our results contribute to evidence that temperature variation across the species range predicts the frequencies of house fly proto-Y chromosomes (Denholm et al. 1986; Franco et al. 1982; Feldmeyer et al. 2008; Hamm et al. 2005; Tomita and Wada 1989). In addition, the house fly proto-Y chromosomes affect thermal traits in ways that are consistent with their clinal distributions (Delclos et al. 2021). These temperature-dependent phenotypic effects likely create variation in the fitness effects of the proto-Y chromosomes, which in turn allow for the maintenance of multiple male-determining loci across populations. This is a special case of local adaptation maintaining genetic variation across populations (Wadgymar et al. 2022). It may also be possible for these differences in fitness effects to promote divergence between populations and subsequent speciation. These links between ecological adaptation and proto-Y chromosomes could therefore provide a mechanism to explain the disproportionate effects of sex chromosomes on speciation (Payseur et al. 2018).

It remains unclear if or how ecological selection pressures that affect sex chromosome evolution are related to sex-specific selection pressures that are predicted to be important for sex chromosome evolution. The house fly proto-Y chromosomes are disproportionately found in males relative to females, but they can also be carried by females who have an *Md-tra^B* allele. Population genetic models predict that the proto-Y chromosomes could have male-beneficial, female detrimental sexually antagonistic fitness effects, which could contribute to the maintenance of the polymorphism within populations (Meisel et al. 2016; Meisel 2021). However, there is no direct evidence for sexually antagonistic effects of the proto-Y chromosomes, let alone sexual antagonism that depends on temperature or any other ecological factor. Future work is therefore needed to evaluate if the well-documented temperature-dependent fitness effects of house fly proto-Y chromosomes have any relationship to

their hypothesized sexually antagonistic effects. Such evidence would provide an important link between the effects of sexual antagonism and ecological variation on sex chromosome evolution.

Supplementary Material

Supplementary material can be found at <http://www.jhered.oxfordjournals.org/> and Dryad (<https://doi.org/10.5061/dryad.15dv41p5x>).

Funding

This material is based upon work supported by the National Science Foundation under Grant No. DEB-1845686. This research was supported in part by the US Department of Agriculture under multistate agreement S-1076.

References

Abbott JK, Nordén AK, Hansson B. Sex chromosome evolution: historical insights and future perspectives. *Proc Biol Sci*. 2017;284:20162806.

Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman T-L, Hahn MW, Kitano J, Mayrose I, Ming R, et al; Tree of Sex Consortium. Sex determination: why so many ways of doing it? *PLoS Biol*. 2014;12:e1001899.

Beukeboom LW, Perrin N. 2014. *The Evolution of Sex Determination*. Oxford, UK: Oxford University Press.

Bull JJ. 1983. Evolution of sex determining mechanisms. Menlo Park, CA, USA: Benjamin/Cummings.

Bull JJ, Charnov EL. Changes in the heterogametic mechanism of sex determination. *Heredity*. 1977;39:1–14.

Bulmer MG, Bull JJ. Models of polygenic sex determination and sex ratio control. *Evolution*. 1982;36:13–26.

Charlesworth D, Charlesworth B. Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. *Genet Res*. 1980;35:205–214.

Delclos PJ, Adhikari K, Hassan O, Cambrie JE, Matuk AG, Presley RI, Tran J, Sriskantharajah V, Meisel RP. Thermal tolerance and preference are both consistent with the clinal distribution of house fly proto-Y chromosomes. *Evol Lett*. 2021;5:495–506.

Denholm I, Franco MG, Rubini PG, Vecchi M. Geographical variation in house-fly (*Musca domestica* L.) sex determinants within the British Isles. *Genet Res*. 1986;47:19–27.

Eshel I. Selection of sex-ratio and the evolution of sex-determination. *Heredity*. 1975;34:351–361.

Feldmeyer B, Koziel ska M, Kuijper B, Weissing FJ, Beukeboom LW, Pen I. Climatic variation and the geographical distribution of sex-determining mechanisms in the housefly. *Evol Ecol Res*. 2008;10:797–809.

Franco MG, Rubini PG, Vecchi M. Sex-determinants and their distribution in various populations of *Musca domestica* L. of Western Europe. *Genet Res*. 1982;40:279–293.

Hamm RL, Gao J-R, Lin GG-H, Scott JG. Selective advantage for III^M males over Y^M males in cage competition, mating competition, and pupal emergence in *Musca domestica* L. (Diptera: Muscidae). *Environ Entomol*. 2009;38:499–504.

Hamm RL, Meisel RP, Scott JG. The evolving puzzle of autosomal versus Y-linked male determination in *Musca domestica*. *G3* 2015;5:371–384.

Hamm RL, Scott JG. Changes in the frequency of Y^M versus III^M in the housefly, *Musca domestica* L., under field and laboratory conditions. *Genetics Res*. 2008;90:493–498.

Hamm RL, Scott JG. A high frequency of male determining factors in male *Musca domestica* (Diptera: Muscidae) from Ipswich, Australia. *J Med Entomol*. 2009;46:169–172.

Hamm RL, Shono T, Scott JG. A cline in frequency of autosomal males is not associated with insecticide resistance in house fly (Diptera: Muscidae). *J Econ Entomol.* 2005;98:171–176.

Hediger M, Henggeler C, Meier N, Perez R, Saccone G, Bopp D. Molecular characterization of the key switch *F* provides a basis for understanding the rapid divergence of the sex-determining pathway in the housefly. *Genetics.* 2010;184:155–170.

Li X, Lin F, van de Zande L, Beukeboom LW. Strong variation in frequencies of male and female determiners between neighboring housefly populations. *Insect Sci.* 2022;29:1470–1482.

Mank JE, Hosken DJ, Wedell N. Conflict on the sex chromosomes: cause, effect, and complexity. *Cold Spring Harbor Perspect Biol.* 2014;6:a017715–a017715.

McDonald IC, Overland DE, Leopold RA, Degruyillier ME, Morgan PB, Hofmann HC. Genetics of house flies: variability studies with North Dakota, Texas, and Florida populations. *J Hered.* 1975;66:137–140.

Meisel RP. The maintenance of polygenic sex determination depends on the dominance of fitness effects which are predictive of the role of sexual antagonism. *G3* 2021;11:jkab149.

Meisel RP. Ecology and the evolution of sex chromosomes. *J Evol Biol.* 2022;35:1601–1618.

Meisel RP, Davey T, Son JH, Gerry AC, Shono T, Scott JG. Is multifactorial sex determination in the house fly, *Musca domestica* (L.), stable over time? *J Hered.* 2016;107:615–625.

Meisel RP, Gonzales CA, Luu H. The house fly Y Chromosome is young and minimally differentiated from its ancient X Chromosome partner. *Genome Res.* 2017;27:1417–1426.

Payseur BA, Presgraves DC, Filatov DA. Sex chromosomes and speciation. *Mol Ecol.* 2018;27:3745–3748.

Roberts RB, Ser JR, Kocher TD. Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. *Science.* 2009;326:998–1001.

Scott JG, Warren WC, Beukeboom LW, Bopp D, Clark AG, Giers SD, Hediger M, Jones AK, Kasai S, Leichter CA, et al. Genome of the house fly, *Musca domestica* L., a global vector of diseases with adaptations to a septic environment. *Genome Biol.* 2014;15:466.

Sharma A, Heinze SD, Wu Y, Kohlbrenner T, Morilla I, Brunner C, Wimmer EA, van de Zande L, Robinson MD, Beukeboom LW, et al. Male sex in houseflies is determined by *Mdmd*, a paralog of the generic splice factor gene CWC22. *Science.* 2017;356:642–645.

Son JH, Meisel RP. Gene-level, but not chromosome-wide, divergence between a very young house fly proto-Y chromosome and its homologous proto-X chromosome. *Mol Biol Evol.* 2021;38:606–618.

Tomita T, Wada Y. Multifactorial sex determination in natural populations of the housefly (*Musca domestica*) in Japan. *Japan J Gen.* 1989;64:373–382.

Truett GE, Heeger P, Mynatt RL, Truett AA, Walker JA, Warman ML. Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). *Biotechniques.* 2000;29:52, 54.

van Doorn GS. Evolutionary transitions between sex-determining mechanisms: a review of theory. *Sex Dev.* 2014;8:7–19.

van Doorn GS, Kirkpatrick M. Turnover of sex chromosomes induced by sexual conflict. *Nature.* 2007;449:909–912.

Wadgymar SM, DeMarche ML, Josephs EB, Sheth SN, Anderson JT. Local adaptation: causal agents of selection and adaptive trait divergence. *Annu Rev Ecol Evol Syst.* 2022;53:87–111.

Werren JH, Beukeboom LW. Sex determination, sex ratios, and genetic conflict. *Annu Rev Ecol Syst.* 1998;29:233–261.