Supplemental Methods

Fitness effects and genotypic fitness

To generate the fitness of the 18 possible multi-chromosome genotypes (Table 1), I first assigned fitness effects to each single chromosome genotype (e.g., for the X and Y^M chromosomes there are three possible genotypes: X/X, X/Y^M, and Y^M/Y^M). The single chromosome genotype is analogous to a single locus genotype in other population genetic models (e.g., Kidwell *et al.* 1977). I generated the single chromosome genotype fitness values by assigning a sex-specific fitness effect, s_{ij}, to each j proto-Y or proto-W chromosome (Y^M, III^M, or IV^F) in each sex i. The fitness effect of IV^F was only assigned for females because the female-determining proto-W chromosome (carrying *Md-tra*^{D}) cannot be found in male genotypes (Table 1). Thus, there are five fitness effects: Y^M in males ($i=\mathrm{male},\ j=\mathrm{Y^M}$), Y^M in females ($i=\mathrm{female}, j=\mathrm{Y^M}$), III^M in males ($i=\mathrm{male}, j=\mathrm{III^M}$), III^M in females ($i=\mathrm{female}, j=\mathrm{III^M}$), and IV^F in females ($i=\mathrm{female}, j=\mathrm{IV^F}$). Each s_{ij} was drawn from a uniform distribution between -1 and 1. Positive fitness effects ($s_{ij}>0$) mean that the proto-Y or proto-W chromosome increases fitness (i.e., it is beneficial), and negative values ($s_{ij}<0$) mean that it decreases fitness (i.e., it is deleterious).

The fitness effects of the proto-Y and proto-W chromosome were then used to calculate the fitness of each single chromosome genotype, using one of four dominance scenarios. The calculations depend on whether the proto-Y chromosome has beneficial ($s_{ij}>0$) or deleterious ($s_{ij}<0$) effects, which ensures that the maximum genotypic fitness for a given chromosome and sex is equal to 1 (Table 2). First, in the additive model, the sex-specific single chromosome fitness values are defined such that the heterozygous genotype (X/Y^M or III/III^M) is intermediate between the two homozygous genotypes. Second, when the proto-Y chromosomes have

dominant fitness effects, the single chromosome fitness of the heterozygote is equal to the proto-Y homozygote. Third, when the effects of the proto-Y chromosomes are recessive, the single chromosome fitness of the heterozygote is equal to the proto-X homozygote and different from the proto-Y homozygote (Table 2). Fourth, the Y^M and III^M proto-Y chromosomes have overdominant effects in males, but are additive in females. In the overdominant model, the fitness effect of recessive deleterious proto-Y alleles is equal in magnitude to the fitness benefit of dominant advantageous proto-Y alleles. This simplifying assumption may not be biologically realistic, and future work could examine the implications of this assumption. In all four models, there is a single fitness effect associated with carrying a copy of the IV^F proto-W chromosome in females; only females can carry IV^F and it is impossible to be homozygous for IV^F (Table 1). In all four dominance models, the fitness of each of the multi-chromosomal genotypes (Table 1) is calculated as the product of the three relevant sex-specific single chromosome genotype fitness values.

I performed a normalization to ensure that the the maximum multi-chromosome genotype fitness in each sex is equal to 1. This is necessary because not all multi-chromosome genotypes are possible for both males and females, which means that the maximum product of all single chromosome genotypes can be <1 even though the maximum single chromosome genotype fitness values are set to 1. To ensure that the maximum female multi-chromosome genotype is 1, I divided each female multi-chromosome genotype fitness by the maximum female multi-chromosome genotype fitness. I did the same for males.

All four dominance scenarios (additive, dominant, recessive, or overdominant) attribute a single sex-specific fitness effect to each proto-Y and proto-W chromosome, whether it is Y^M, III^M, or IV^F. I am therefore assuming that all copies of each proto-sex chromosome in the population carry an identical suite of beneficial and deleterious alleles (i.e., all copies of III^M are identical, all copies of Y^M are identical, and all copies of IV^F are identical). In addition, all models assume complete linkage between any allele(s) under selection and the *Mdmd* locus on the Y^M and III^M

chromosomes. Similarly, I assume complete linkage between the female-determining *Md-tra*^{D} locus and any alleles on the IV^{F} chromosome. Complete linkage may be achieved if, for example, chromosomal inversions suppress recombination between the proto-Y and proto-X (or proto-Z and proto-W) chromosomes in heterozygotes (Bergero and Charlesworth 2009; Wright *et al.* 2016). Free recombination in sex chromosome homozygotes would not affect genetic variation on a sex chromosome because all copies of each chromosome are assumed to carry the same alleles.

Simulations with infinite population size

I used forward simulations to determine how the fitness effects of the proto-sex chromosomes affect their frequencies in populations. These simulations were performed with non-overlapping generations and random mating, assuming a population of infinite size (simulations with finite population sizes are described later). Each generation of a simulation consists of two discrete steps (Figure 1A). First, the frequency of each of the 18 genotypes is multiplied by its corresponding fitness value, which models differential survival across genotypes. Second, recursion equations developed by Hamm (2008) and previously implemented by Meisel *et al.* (2016) are used to model the production of progeny by random mating (Supplemental Table S1). After 1,000 generations of selection and random mating, the frequency of each genotype and proto-sex chromosome was calculated. This process was performed for 1,000,000 fitness arrays for each dominance scenario (additive, dominant, recessive, and overdominant) and four possible initial genotype frequencies (16,000,000 total simulations).

I considered four possible initial genotype frequencies for each simulation. The first three initial frequencies are based on the observed frequencies of Y^{M}, III^{M}, and *Md-tra*^{D} from three North American populations (Meisel *et al.* 2016). These populations were sampled in Chino, CA (Meisel *et al.* 2016), Wake County, NC (Hamm and Scott 2008), and Chemung County, NY

(Scott *et al.* 2013). Initial genotype frequencies were calculated based on the observed frequencies of Y^M, III^M, and *Md-tra*^{D}, assuming random mating (Meisel *et al.* 2016). These three populations were chosen because Y^M, III^M, and *Md-tra*^{D} have all remained at a frequency >1% across multiple years of sampling (Hamm *et al.* 2005; Hamm and Scott 2008; Meisel *et al.* 2016). I used frequencies from actual populations as the initial frequencies because I am specifically interested in how selection can maintain PSD at the frequencies observed in natural populations. The fourth starting values consist of all 18 genotypes at the same initial frequency (i.e., 1/18 for each genotype). This allows me to evaluate how selection pressures can drive the proto-sex chromosomes to the frequencies observed in natural populations if they start at arbitrary values that deviate from the observed frequencies.

To determine whether the fitness effects maintain PSD, I first selected fitness arrays that produced frequencies of each proto-sex chromosome (Y^M, III^M, IV^F, and their homologous chromosomes, X, III, and IV) that are all >0.1% after 1,000 generations. This criterion ensures that all three chromosomes are polymorphic with both alleles at a frequency that is measurable (i.e., 1/1,000) given the sampling schemes used in previous studies of natural house fly populations (Hamm *et al.* 2005; Hamm and Scott 2008; Meisel *et al.* 2016). I refer to genotype fitness arrays with all proto-sex chromosomes present at a frequency >0.1% as maintaining PSD, although not necessarily at the frequencies observed in natural populations.

From the genotype fitness arrays that maintain PSD (i.e., all proto-sex chromosomes >0.1%), I selected those that produce proto-sex chromosome frequencies most similar to the frequencies observed in natural populations. To do so, I calculated the mean squared error (MSE) between the frequencies of Y^M, III^M, and IV^F after 1,000 generations in a simulation (S) and the observed frequencies (O) in a natural population (CA, NC, or NY):

$$\mathrm{MSE} = \frac{1}{3}[(O_{Y^M} - S_{Y^M})^2 + (O_{III^M} - S_{III^M})^2 + (O_{tra^D} - S_{tra^D})^2].$$

When using simulations that started with proto-sex chromosome frequencies observed in a population (e.g., CA), I only used MSE to compare with observed frequencies from that same population. When simulations started with equal frequencies of all 18 genotypes, I compared simulated chromosomes frequencies with the observed frequencies in each of the three populations (CA, NC, and NY).

I retained the 1,000 fitness arrays that produce proto-sex chromsome frequencies with the lowest MSE for a given dominance model, population, and starting genotype frequency. I refer to each of these 1,000 fitness arrays as the best-fitting arrays for each population. Simulations for the 1,000 best-fitting arrays were then run for 1,000,000 generations and compared to the 1,000 generation simulations to assess the stability of chromosomal frequencies. I also selected 1,000 random fitness arrays and 1,000 fitness arrays that maintain PSD to perform 1,000,000 generation simulations.

Simulations with finite population sizes

I assessed how well the 1,000 best-fitting fitness arrays for each population, starting frequency, and dominance model maintain PSD when population sizes are finite. Finite population sizes introduce stochasticity to the change in allele frequencies across generations (i.e., genetic drift), in contrast to the deterministic effects of constant fitness values in a population of infinite size (Hartl and Clark 2007). It is not possible to calculate the probability of fixation for alleles in a finite population when there is complex sex-linked inheritance, as in the house fly PSD system (Meisel *et al.* 2016). To overcome this limitation, I estimated the probability of fixation using simulations that modeled finite populations by including multinomial sampling each generation.

I simulated finite populations by adding one additional step to the simulations described above in order to model a population size (N) of 10,000 individuals (Figure 1B). This population size was chosen because it is small enough to capture the effects of drift within the time scale of

my simulations. Each simulation was started with the frequencies of the proto-sex chromosomes observed in the population where the fitness array was identified as best-fitting. In each generation, following multiplication by the fitness array (i.e., natural selection), I used multinomial sampling to calculate the new frequencies of all 18 genotypes. Multinomial sampling for all genotype was performed with 10,000 trials (i.e., $N=10^4$) and a probability of success equal to the frequency of each genotype after selection. The resulting array of 18 genotype frequencies was then divided by the sum of all values to ensure that the genotype frequencies sum to 1. This array of genotype frequencies (after selection and multinomial sampling) was next used as input into the same recursion equations described above to calculate genotype frequencies after random mating (Supplemental Table S1). The process (selection, multinomial sampling, and random mating) was repeated for 1,000 generations. I simulated 100 replicate finite populations for each fitness array. I then used the frequency with which fixation and loss occur in those 100 replicates as an estimate of the probability of fixation and loss of proto-sex chromosomes within 1,000 generations.

I determined a null expectation for fixation or loss of proto-sex chromosomes in finite populations by using simulations without selection (i.e., genetic drift and no fitness differences across genotypes). To those ends, I performed simulations with 10,000 individuals for 1,000 generations, as was done in the simulations with selection and drift. I performed 1,000 replicate "drift-only" simulations with starting values at the observed frequencies of the proto-sex chromosomes in each population (CA, NC, and NY). These drift-only simulations included the same steps as the simulations with both selection and drift, except that there is no differential survival (i.e., natural selection) step in the drift-only simulation. After 1,000 generations, I calculated the frequency of each proto-Y and proto-W chromosome (Y^M, III^M, and IV^F), as well as the frequency with which each chromosome was lost, across the 1,000 simulations.

Bergero, R., and D. Charlesworth, 2009 The evolution of restricted recombination in sex chromosomes. Trends Ecol. Evol. 24: 94–102.

Hamm, R. L., 2008 Exploring the population genetics of the house fly sex determining genes, M and F [Ph.D.]: Cornell University.

Hamm, R. L., and J. G. Scott, 2008 Changes in the frequency of Y^M versus III^M in the housefly, *Musca domestica* L., under field and laboratory conditions. Genet. Res. 90: 493–498.

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

Hartl, D. L., and A. G. Clark, 2007 *Principles of Population Genetics*. Sinauer.

Kidwell, J. F., M. T. Clegg, F. M. Stewart, and T. Prout, 1977 Regions of stable equilibria for models of differential selection in the two sexes under random mating. Genetics 85: 171–183.

Meisel, R. P., T. Davey, J. H. Son, A. C. Gerry, T. Shono *et al.*, 2016 Is multifactorial sex determination in the house fly, *Musca domestica* (L.), stable over time? J. Hered. 107: 615–625.

Scott, J. G., C. A. Leichter, F. D. Rinkevich, S. A. Harris, C. Su *et al.*, 2013 Insecticide resistance in house flies from the United States: resistance levels and frequency of pyrethroid resistance alleles. Pestic. Biochem. Physiol. 107: 377–384.

Wright, A. E., R. Dean, F. Zimmer, and J. E. Mank, 2016 How to make a sex chromosome. Nat. Commun. 7: 12087.