Population Genetics of the Fragile Y Hypothesis

Heath Blackmon

last updated: 2016-02-05

Contents

Introduction

The fragile Y hypothesis provides a single explanation for rates of sex chromosome aneuploid offspring, origins of achiasmatic meiosis, fusions of autosomes and sex chromosomes, and the distribution of XO taxa across the tree of life. Briefly I posit that sexual antagonism leads to recurrent selection to reduce the recombining region of the sex chromosomes (PAR). However, this conflicts with the essential nature of the PAR - it must pair and recombine for proper segregation of the X and Y chromosomes during spermatogenesis. This leads to selection for gene movement off the Y, rejuvinations of the PAR via fusions or translocations with the autosomes, and evolution of alternative meiotic mechanisms like achiasmatic meiosis that render the PAR superfluos during spermatogenesis.

A variety of empirical data is consistant with this view of sex chromosome evolution. 1. Rate estimates of Y chromosome loss in Coleoptera indicate that clades with achiasmatic meiosis lose the Y less often than those that require a PAR. 2. Y chromosome losses in mammals are concentrated in lineages with small PARs 3. Y chromosome loss and achiasmatic meiosis are co-localized taxonomically in eutherian mammals. 4. Phylogenetic distribution of achiasmatic meiosis becomes reasonable. 5. Domestic animals with small PARs have more XO offspring However, a mathematical treatment of this hypothesis would reveal under what range of parameters we expect this process to play an important role in the evolution of genomes.

Model description

Model 1

I will evaluate the fragile Y hypothesis with a two-loci model where the first locus has two alleles A and a. This locus represents the sex determining locus and a is a dominant male determining allele. Therefor the possible genotypes at this locus are AA (female) and Aa (male). The second locus represents a locus with unresolved sexual antagonism, and has three alleles B the female benefit allele, b the male benefit allele, b' a second version of the male benefit allele that has been captured inside of a version spanning both loci - completely linking a copy of the a and b alleles. In the absence of the b' allele recombination occurs at a

rate r between locus 1 and locus 2. Finally, the b' allele will cary a cost during spermatogenesis causing the production of sperm with no sex chromosome at a rate of μ this represents the aneuploidy rate of the genotype.

Model 2

To test whether our expectations should differ when isolating inversions occur on the X or the Y we will evaluate a second model where the second locus has three alleles B the female benefit allele, b the male benefit allele, B' a second version of the female benefit allele that has been captured inside of a version spanning both loci - completely linking a copy of the A and B alleles. In the absence of the b' allele recombination occurs at a rate r between locus 1 and locus 2. Finally, the B' allele will cary a cost during spermatogenesis causing the production of sperm with no sex chromosome at a rate of μ this represents the aneuploidy rate of the genotype.

Model 3

I will perform a final analysis of this hypothesis to test whether our expectations should differ when the female is the heterogametic sex. For this I will employ a two-loci model where the first locus has two alleles A and a in this case though a is a dominant female determining allele. Therefor the possible genotypes at this locus are AA (male) and Aa (female). The second locus represents a locus with unresolved sexual antagonism, and has three alleles B the male benefit allele, b the female benefit allele, b' a second version of the female benefit allele that has been captured inside of a version spanning both loci - completely linking a copy of the a and b alleles. In the absence of the b' allele recombination occurs at a rate r between locus 1 and locus 2. Finally, the b' allele will cary a cost during oogenesis causing the production of eggs with no sex chromosome at a rate of μ this represents the aneuploidy rate of the genotype.

Setup and analysis of model 1

Production of gametes

Table 1 All 27 possible phased genotypes In determining the gamete pool we have to consider the gametes produced by each one of these possible genotypes.

	AB	Ab	Ab'	aB	ab	ab'
AB	X_1	X_7	X_{13}	X_{19}	X_{22}	X_{25}
Ab	X_2	X_8	X_{14}	X_{20}	X_{23}	X_{26}
Ab'	X_3	X_9	X_{15}	X_{21}	X_{24}	X_{27}
aB	X_4	X_{10}	X_{16}			
ab	X_5	X_{11}	X_{17}			
ab'	X_6	X_{12}	X_{18}			

Table 2 Females phased genotypes

	AB	Ab	Ab'
AB	X_1	X_7	X_{13}
Ab	X_2	X_8	X_{14}
Ab'	X_3	X_9	X_{15}

Table 3 Males phased genotypes

	AB	Ab	Ab'	aB	ab	ab'
AB				X_{19}	X_{22}	X_{25}
Ab				X_{20}	X_{23}	X_{26}
Ab'				X_{21}	X_{24}	X_{27}
aB	X_4	X_{10}	X_{16}			
ab	X_5	X_{11}	X_{17}			
ab'	X_6	X_{12}	X_{18}			

We are going to make a number of simplifying assumptions to make this a bit more tractable. First we will assume that AB/ab are the same as Aa/Bb this is acceptable since we aren't going to be concerned with cis/trans effects. We will also assume that AB/ab is the same as ab/AB. This is acceptable becuase we are not concerned with maternal/paternal effects.

Sperm Genotypes and Frequency. $AB = X_4\ X_5\ X_6\ X_{10}\ X_{16}\ X_{19}\ X_{20}\ X_{21}\ X_{22}\ X_{25}$

Ab Ab' aB ab ab'

r <- .5 AB <- .25 Ab <- .25 aB <- .25

```
Evolve <- function( r, v, t, AB, Ab, aB, gen){


  plot.valA <- plot.valB <- plot.valAB <- vector()



  # this loop goes through the generations
  for(i in 1:gen){


    #obtaining frequency of 'ab' gamete
    ab = 1 - AB - Ab - aB

    #genotype frequencies

    #      G_ABAB = AB*AB
    #      G_ABAb = AB*Ab
    #      G_ABaB = AB*aB
    #      G_ABab = AB*ab
    #      G_AbAB = Ab*AB
    #      G_AbAb = Ab*Ab
    #      G_AbaB = Ab*aB
    #      G_Abab = Ab*ab
    #      G_aBAB = aB*AB
    #      G_aBAb = aB*Ab
    #      G_aBaB = aB*aB
    #      G_aBab = aB*ab
    #      G_abAB = ab*AB
    #      G_abAb = ab*Ab
    #      G_abaB = ab*aB
    #      G_abab = ab*ab

    #Fitness values
    F_AB_AB = 1
```

```r
    F_AB_Ab = 1
    F_AB_aB = 1
    F_AB_ab = 1
    F_Ab_AB = 1
    F_Ab_Ab = 1 #(1-t)*(1/2)*G_abaB + G_abAB + G_aBab + G_aBAb + G_AbaB + G_AbAB + G_ABab + G_ABAb
    F_Ab_aB = 1
    F_Ab_ab = 1 #(1-t)*(1/2)*G_abaB + G_abAB + G_aBab + G_aBAb + G_AbaB + G_AbAB + G_ABab + G_ABAb
    F_aB_AB = 1
    F_aB_Ab = 1
    F_aB_aB = 1 #(1-t)*(1/2)*G_ABaB + G_ABaB + G_AbaB + G_Abab + G_aBAB + G_aBAb + G_abAB + G_abAb
    F_aB_ab = 1 #(1-t)*(1/2)*G_ABaB + G_ABaB + G_AbaB + G_Abab + G_aBAB + G_aBAb + G_abAB + G_abAb
    F_ab_AB = 1
    F_ab_Ab = 1 #(1-t)*(1/2)*G_abaB + G_abAB + G_aBab + G_aBAb + G_AbaB + G_AbAB + G_ABab + G_ABAb
    F_ab_aB = 1 #(1-t)*(1/2)*G_ABaB + G_ABaB + G_AbaB + G_Abab + G_aBAB + G_aBAb + G_abAB + G_abAb
    F_ab_ab = 1 #(1-t)*(1/2)*G_ABaB + G_ABaB + G_AbaB + G_Abab + G_aBAB + G_aBAb + G_abAB + G_abAb
    #+ (1-t)*(1/2)*G_abaB + G_abAB + G_aBab + G_aBAb + G_AbaB + G_AbAB + G_ABab + G_ABAb


    #Average fitness of F_i_j
    W <- sum(AB * AB * F_AB_AB, AB * Ab * F_AB_Ab,
             AB * aB * F_AB_aB, AB * ab * F_AB_ab,
             Ab * AB * F_Ab_AB, Ab * Ab * F_Ab_Ab,
             Ab * aB * F_Ab_aB, Ab * ab * F_Ab_ab,
             aB * AB * F_aB_AB, aB * Ab * F_aB_Ab,
             aB * aB * F_aB_aB, aB * ab * F_aB_ab,
             ab * AB * F_ab_AB, ab * Ab * F_ab_Ab,
             ab * aB * F_ab_aB, ab * ab * F_ab_ab)

    #Marginal fitness of haplotypes
    fAB = AB*(F_AB_AB) + Ab*(F_AB_Ab) + aB*(F_AB_aB) + ab*(F_AB_ab)

    fAb = AB*(F_Ab_AB) + Ab*(F_Ab_Ab) + aB*(F_Ab_aB) + ab*(F_Ab_ab)
    faB = AB*(F_aB_AB) + Ab*(F_aB_Ab) + aB*(F_aB_aB) + ab*(F_aB_ab)
    fab = AB*(F_ab_AB) + Ab*(F_ab_Ab) + aB*(F_ab_aB) + ab*(F_ab_ab)
    #Linkage Disequilibrium
    D = (AB*ab - Ab*aB)
    #Frequency of gametes in the next generation
    AB = (AB*fAB - r*D*F_AB_ab)/W
    Ab = (Ab*fAb + r*D*F_AB_ab)/W
    aB = (aB*faB + r*D*F_AB_ab)/W
    ab = (ab*fab - r*D*F_AB_ab)/W
    #Frequency of A allele
    pA = AB + Ab
    #Frequency of B allele
    pB = AB + aB
    plot.valA[i] <- pA
    plot.valB[i] <- pB
    plot.valAB[i] <- AB
}
results <- list()
results[[1]] <- plot.valA
results[[2]] <- plot.valB
results[[3]] <- plot.valAB
```

```
    return(results)
}
foo <- Evolve(r = 0.5, t = 9, v = 9, AB = 0.1, Ab = 0.3, aB = 0.1, gen = 50)
plot(foo[[1]], type="l", lwd = 4, col="red", ylim = range(0:5), xlab="Generations", ylab="Freq. of allel
lines(foo[[2]], type="l", lwd = 4, col ="green")
lines(foo[[3]], type="l", lwd = 4, col ="orange")
```