

Y chromosome diversity

We have done simulations of background selection on the Y-chromosome to test whether our observed diversity, θ_Y , is consistent with the effects of background selection on this chromosome.

Let's ask: given our simulation parameters, what does the theory predict?

Here are the input parameters:

- $Ne_Y = 1/4Ne_A$, where the initial $Ne_A = 1000$
- $\mu = 0.005$ (neutral mutation rate, per gamete, per generation)
- $U = 0.02$ (deleterious mutation rate, per *diploid*; so, twice the per gamete rate)
- $r = 0$ (recombination rate)
- $h = 1$ (dominance)

Under the BGS model, the expected π is $E[\pi] = \pi_0 e^{-\frac{U}{2sh+r}}$, U is the mutation rate to strongly-deleterious alleles, π_0 is the value expected in the absence of BGS (*i.e.* $\pi_0 = \theta = 4Ne\mu$), s and h are the selection and dominance coefficients, and r is the recombination rate.

For our parameters, we have: $\pi_0 = \theta = 4Ne\mu = 5$, so

$$E[\pi] = 5e^{-\frac{0.02}{0.1}} = 1.839397$$

```
5*(exp(-(0.02/((2*0.01)))))
```

```
## [1] 1.839397
```

Analogously for the autosomes, with $r = 0.5$, we have

$$\pi_0 = \theta = 4Ne\mu = 20$$

and

$$E[\pi] = 20e^{-\frac{0.02}{0.1+0.5}} = 19.24537$$

```
20*(exp(-(0.02/((2*0.01)+0.5))))
```

```
## [1] 19.24537
```

and for the X chromosome

$$\pi_0 = \theta = 4Ne\mu = 15$$

and

$$E[\pi] = 15e^{-\frac{0.02}{0.1+0.5}} = 14.43403$$

```
15*(exp(-(0.02/((2*0.01)+0.5))))
```

```
## [1] 14.43403
```

Expected Y/A and X/A diversity

OK, so then the expected ratios of diversity under background selection are:

$$\frac{E[\pi]_Y}{E[\pi]_A} = \frac{1.839397}{19.24537} = 0.095$$

and

$$\frac{E[\pi]_X}{E[\pi]_A} = \frac{14.43403}{19.24537} = 0.75$$

Simulated Y/A and X/A diversity

How does this compare with the simulation output?

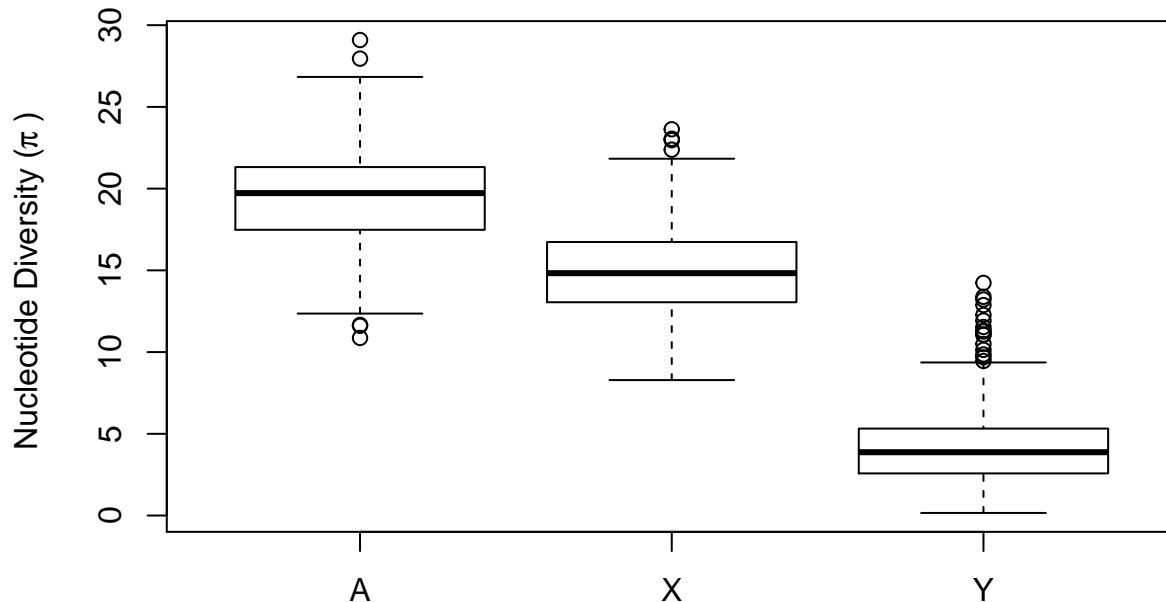
import the simulation data

```
#autosomes
a <- read.csv("/Users/josh/Dropbox/temp/Ydiversity/simdata/a_fixed.csv", header = T)
#X-chromosome
x <- read.csv("/Users/josh/Dropbox/temp/Ydiversity/simdata/x_fixed.csv", header = T)
#Y-chromosome
y <- read.csv("/Users/josh/Dropbox/temp/Ydiversity/simdata/y_fixed.csv", header = T)
```

Let's see some plots of the simulated π_Y , π_X , and π_A

```
boxplot(a$thetapi, x$thetapi, y$thetapi, names = c("A","X","Y"),
        ylab = expression(paste("Nucleotide Diversity (", pi, ")")))

```



The following π values are the means from all of the simulations, for each chromosome type. You can see for the autosomes and the X chromosome, they are quite close to the theoretically expected values calculated above.

$$\pi_{A_{sim}} = 19.5$$

$$\pi_{X_{sim}} = 14.9$$

$$\pi_{Y_{sim}} = 4.2$$

```
mean(a$thetapi)
```

```
## [1] 19.56666
```

```
mean(x$thetapi)
```

```
## [1] 14.91575
```

```
mean(y$thetapi)
```

```
## [1] 4.22218
```

However, the estimate for the Y chromosome is much higher than predicted the by model.

Moreover, ratios for the simulated data:

$$X/A = \frac{\pi_{X_{sim}}}{\pi_{A_{sim}}} = 0.76$$

and

$$Y/A = \frac{\pi_{Y_{sim}}}{\pi_{A_{sim}}} = 0.21$$

are not different from neutral expectations

```
mean(x$thetapi)/mean(a$thetapi)
```

```
## [1] 0.7623043
```

```
mean(y$thetapi)/mean(a$thetapi)
```

```
## [1] 0.2157844
```

This is basically unchanged (well, even higher for the Y) if one weights these means by the number of polymorphic sites (called “npoly” in the data)

Stephen: I don't know why this is? It might be some issue with the simulation somehow still treating sites on the Y as independent (this is my suspicion), but I have been investigating all day, and not quite sure anymore...