

Amm_transcriptome_results_update2

We investigated the issue of second cloud of female-biased genes in brain tissues, it turned out that 30% transcripts do not have any BLAST hit (88%) with new XX genome.

We removed the non-frog transcripts, and redone all analysis using the updated transcriptome.

A. Tissue specificity index tau

B. Sex-biased gene expression along developmental stages, across adult tissues

C. Transcription degeneration

D. Coding sequence divergence and Faster-X

A. Tissue specificity index tau

Approach:

1. Define in total 11 tissues, since somatic tissues and early embryonic tissues have very small number of sex-biased genes, we treat them as single tissue regardless of sex, e.g. G23, G27, G31, liver and brain; for late larva tissues and gonad tissues, we treat tissue with separate sexes, e.g. ovary, testis, XX G43, XY G43, XX G46 and XY G46 (see Brown & Bachtrog 2014).
2. Use Kallisto to quantify transcript expression, and then calculate Tau from the generated output of TPM matrix.
3. Use tissue specificity index Tau formula below (Mank et al. 2008; Brown & Bachtrog 2014):

$$\tau = \frac{\sum_{i=1}^N 1 - \frac{\log E_i}{\log E_{max}}}{N - 1},$$

i is one tissue, E_i is the expression (non-normalized TPM value) of certain transcript.

4. Keep transcripts which are expressed in at least one of the 11 tissues with $TPM \geq 1$.

Figure 1. Histogram distribution of Tau for all transcripts in Ammarnas transcriptome.

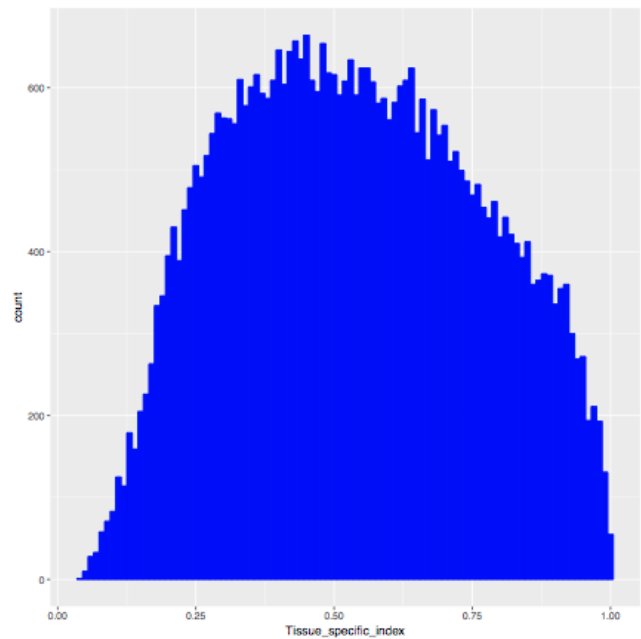


Figure 2. Boxplot of sex-biased genes and the tissue specificity index Tau at stage G46 (gene expression from sex-reversed XX male is not included here).

Sex-biased genes are more tissue specific than unbiased genes (note, male-bias genes have higher tau than female-biased genes).

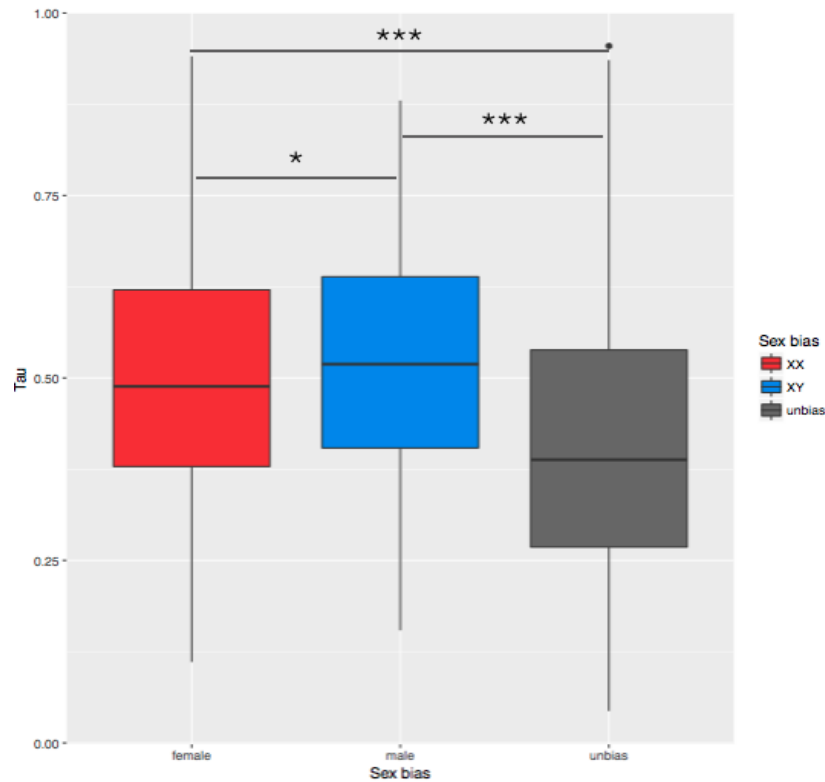
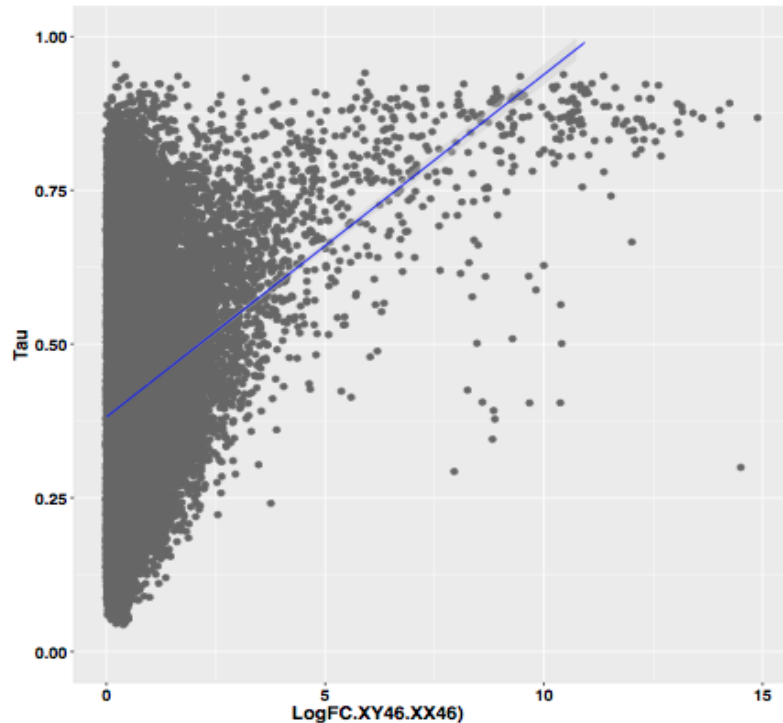


Figure 3. Correlation between absolute values of gene expression ratio $\text{Log}_2(\text{XY46}/\text{XX46})$ (include sex-biased and unbiased genes) and tissue specificity index tau.

`lm(tau~sqrt(abs(g46_tau_sub$logFC.XY46.XX46))*bias, g46_tau_sub)`

Results show tau is significantly correlated with expression ratio, sex bias, as well as the interaction of the two.



Furthermore, we could also ask whether the evolutionary rate of coding sequence can be explained by tissue specificity, or sex bias, or the interaction of the two factors.

```
y1 <- lm(dNdS~sqrt(tau)*bias, g46_tau_dnds)
anova(y1)
#####
#Response: dNdS
#Df Sum Sq Mean Sq F value Pr(>F)
# sqrt(tau)      1  0.4781 0.47813 140.8440 < 2.2e-16 ***
# bias           2  0.0643 0.03216   9.4725 7.799e-05 ***
# sqrt(tau):bias  2  0.1263 0.06317  18.6083 8.735e-09 ***
# Residuals     6589 22.3678 0.00339
#####
```

If removing unbiased genes, the results largely remain, but the interaction does not influence the dNdS anymore.

```
y2 <- lm(dNdS~(tau)*bias, g46sub_tau_dnds)
```

```

anova(y2)
#####
#Response: dNdS
#Df Sum Sq Mean Sq F value Pr(>F)
# tau      1 0.3678 0.36779 114.6115 < 2.2e-16 ***
# bias      1 0.0533 0.05330  16.6078 4.793e-05 ***
# tau:bias   1 0.0034 0.00338   1.0528  0.305
# Residuals 1827 5.8629 0.00321
#####

```

At stage G46, the coding sequence evolutionary rate can be explained by tissue specificity, sex bias, and also their interaction.

Figure 4. Boxplot of sex-biased genes and the tissue specificity index Tau in gonad tissues. Here again, sex-biased genes are more tissue specific than unbiased genes, but within sex-biased genes, opposite as G46, female-biased genes has higher tau than male-bias.

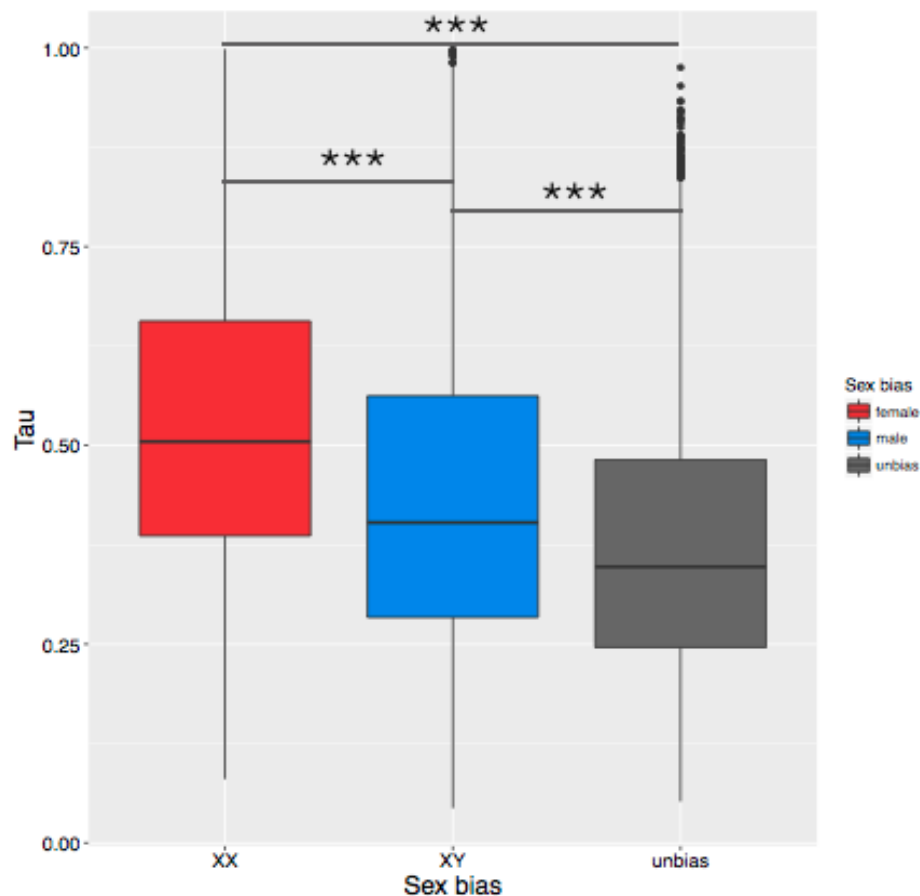
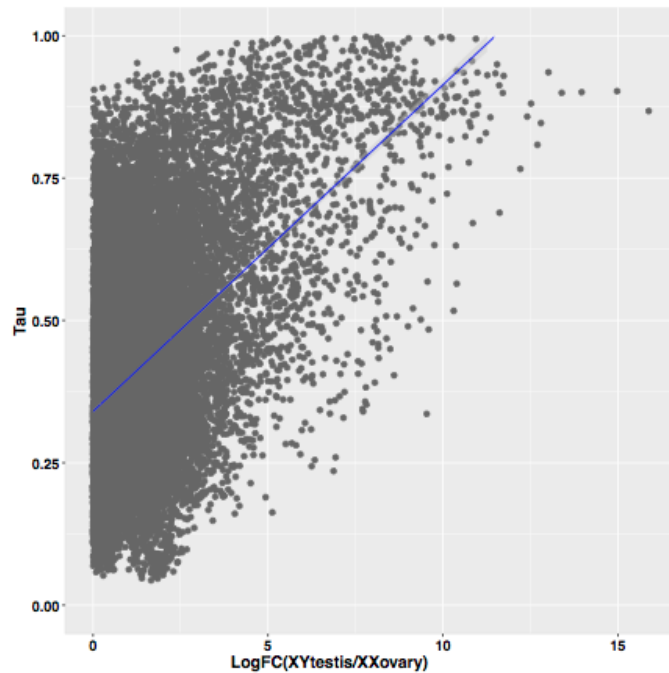


Figure 5. Correlation between absolute values of gene expression ratio $\text{Log}_2(\text{male/female})$ (only for sex-biased and unbiased genes) and tissue specificity index tau.

`lm(tau~sqrt(abs(logFC.XYtestis.XXovary))*bias, gonad_tau)`

Tau is significantly correlated with expression ratio, sex bias and the interaction.



Similarly, for gonad tissues, we could also ask whether the evolutionary rate of coding sequence can be explained by tissue specificity, or sex bias, or the interaction of the two factors.

```
y2 <- lm(dNdS ~ sqrt(tau) * bias, gonad_tau_dnds)
anova(y2)
#####
#Response: dNdS
#Df Sum Sq Mean Sq F value Pr(>F)
# sqrt(tau)      1  0.2805 0.280491 80.0801 < 2.2e-16 ***
# bias           2  0.0020 0.001013  0.2892  0.748840
#sqrt(tau):bias  2  0.0378 0.018880  5.3904  0.004583 **
# Residuals     5774 20.2242 0.003503
#####
```

If removing unbiased genes, redo the linear regression, the results remain the same.

```
y3 <- lm(dNdS ~ sqrt(tau) * bias, gonadsub_tau_dnds)
anova(y3)
#Response: dNdS
#Df Sum Sq Mean Sq F value Pr(>F)
# sqrt(tau)      1 0.1280 0.127956 35.2817 3.152e-09 ***
# bias           1 0.0014 0.001430 0.3944 0.530034
# sqrt(tau):bias 1 0.0376 0.037630 10.3757 0.001289 **
# Residuals     3277 11.8847 0.003627
```

At stage G46, the coding sequence evolutionary rate can be explained by tissue specificity, and the interaction of tissue specificity but not sex bias. It is interesting to see that here sex bias does not significantly explain the difference in evolutionary rate of coding sequence. I think it will be quite interesting to see the patterns in other tissues too. This is different from pattern in G46.

More analysis will come along the way, now it is meant to get some idea on the patterns and trigger further related questions.

B. Sex-biased gene expression along developmental stages, across adult tissues

Approach: We cleaned the transcriptome by BLAST the transcriptome with the new XX genome, and removed the transcripts which were not found in the XX genome. This transcriptome with removed-non-frog-transcripts is used, we quantify the transcript abundance with Kallisto.

EdgeR is applied to analyze differential or sex-biased gene expression across developmental stages, as well as adult tissues.

The selection criteria to remove low expressed transcripts are:

- 1) For each transcript, we select the average LogCPM above 0;
- 2) Additionally, we select LogCPM >1 is present in at least half of the tissues per sex.

Figure 6. Number of sex-biased genes throughout development and adult tissues (FDR < 0.05, Log2 >=1).

Sex-biased genes increase dramatically throughout development, and show stronger difference in adult gonadal tissues, but not in somatic tissues.

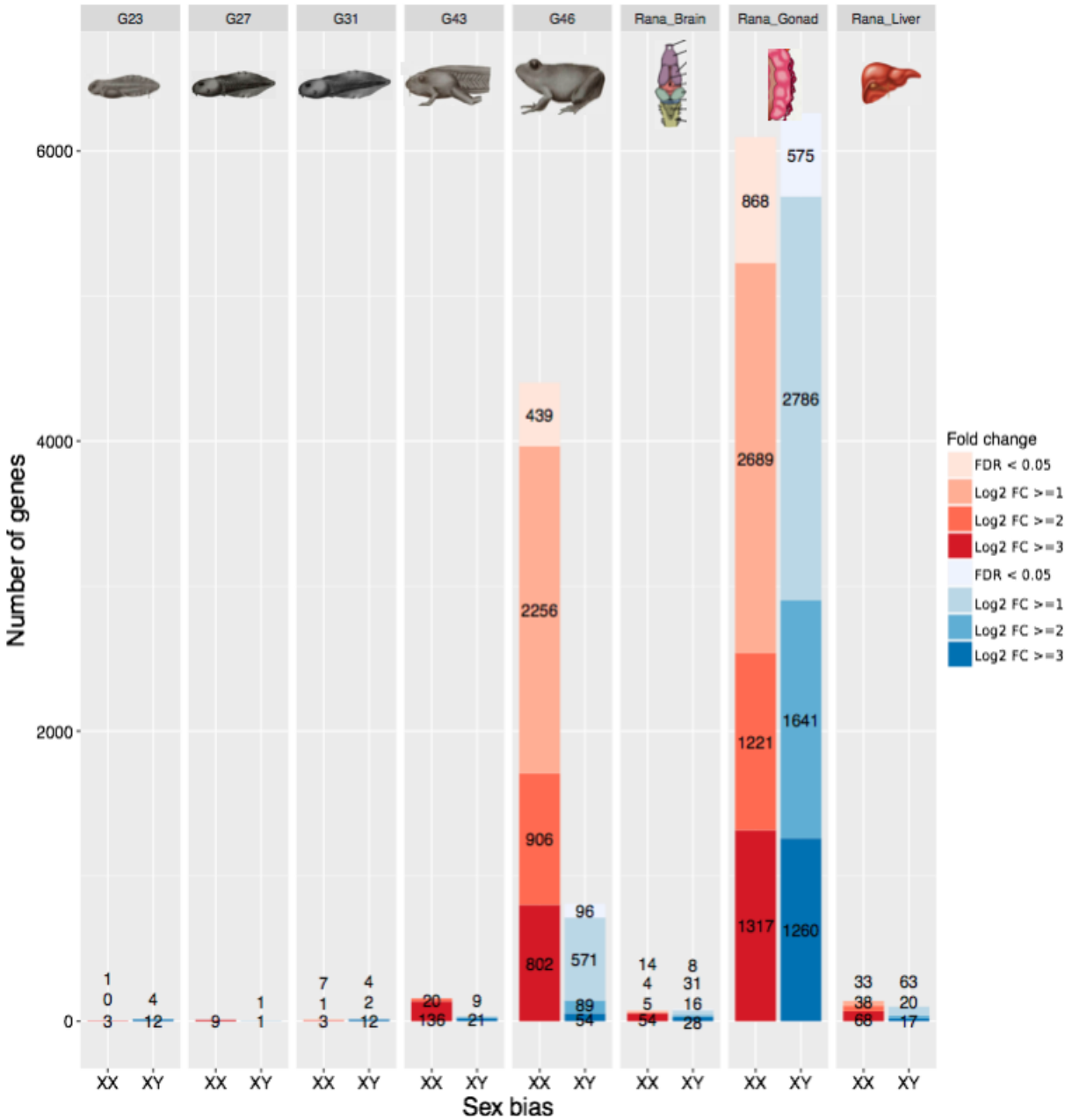


Figure 7. Venn diagram on sex-biased genes throughout development. (a) Shared sex-biased genes, (b) shared female-biased genes and (c) shared male-biased genes throughout development stages.

There is none shared sex-biased genes across five developmental stages, but there are certain number of shared in adjacent stages.

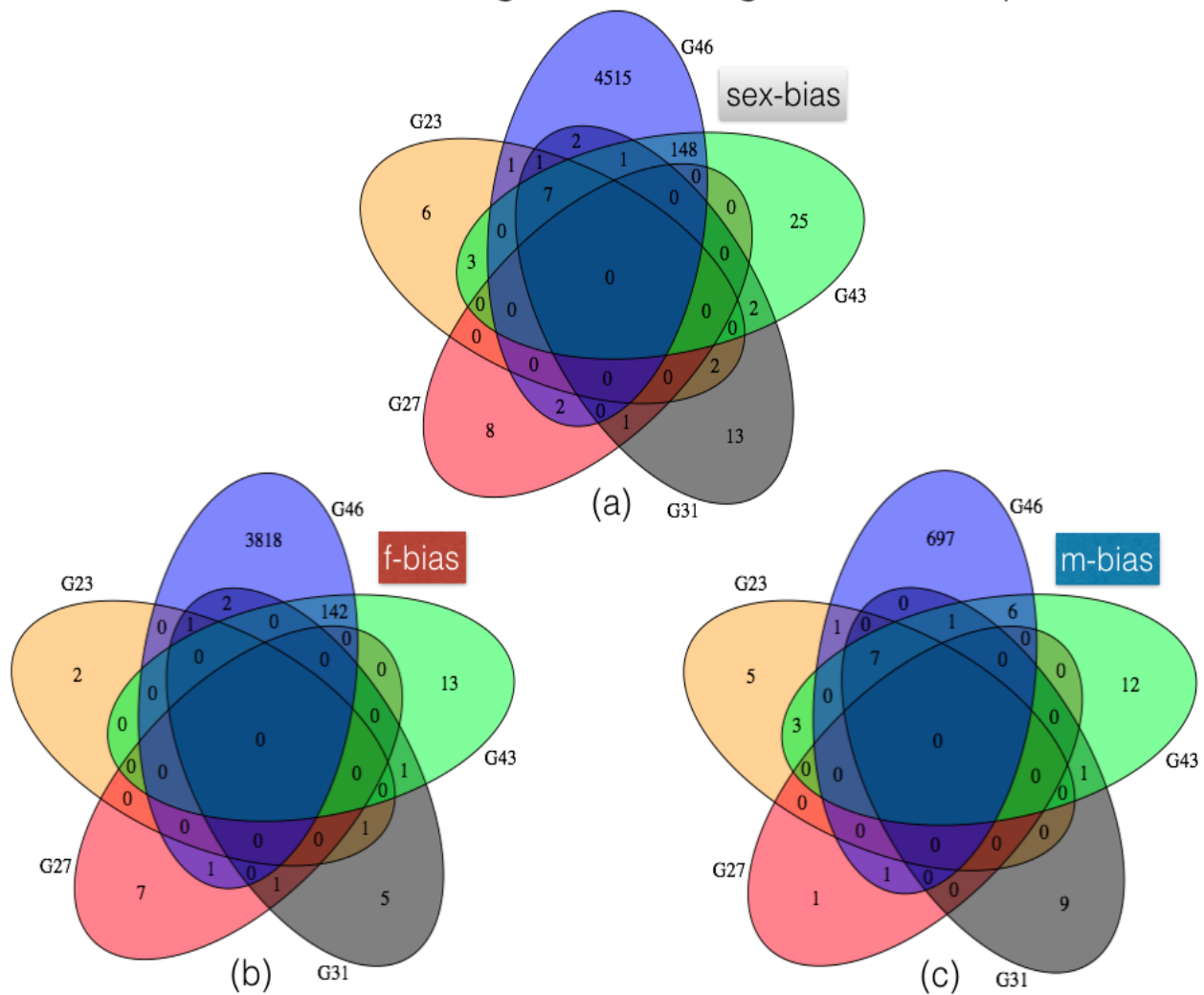


Figure 8. Venn diagram on sex-biased genes among adult tissues. (a) Shared sex-biased genes, (b) shared female-biased genes and (c) shared male-biased genes among three adult tissues, brain, gonad and liver.

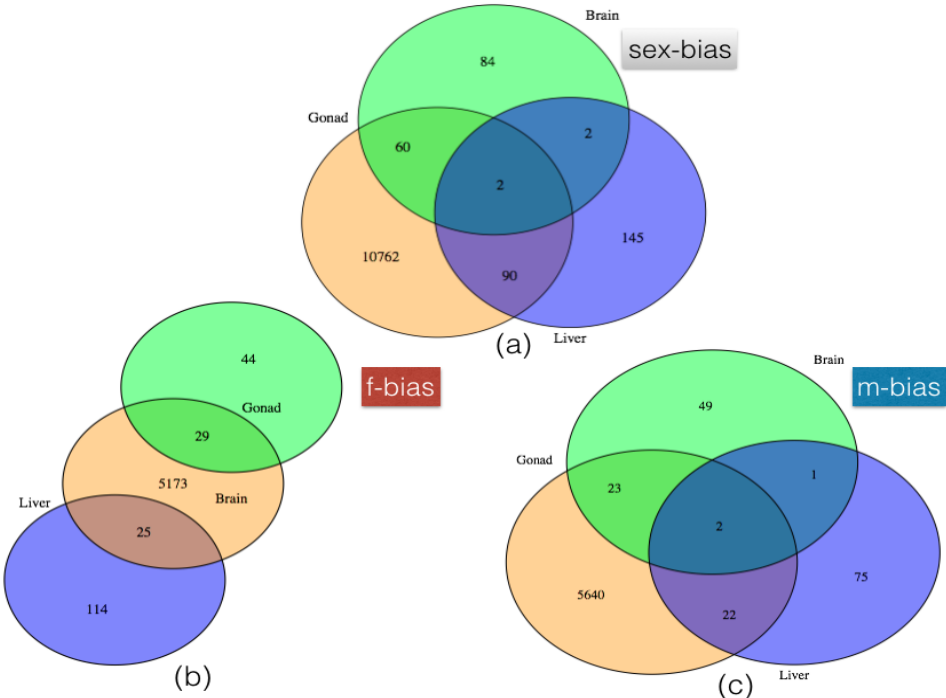


Figure 9. Venn diagram on sex-biased genes between G46 and gonad tissues. (a) Shared sex-biased genes, (b) shared female-biased genes and (c) shared male-biased genes between G43 and gonad tissues which have high sex-biased genes.

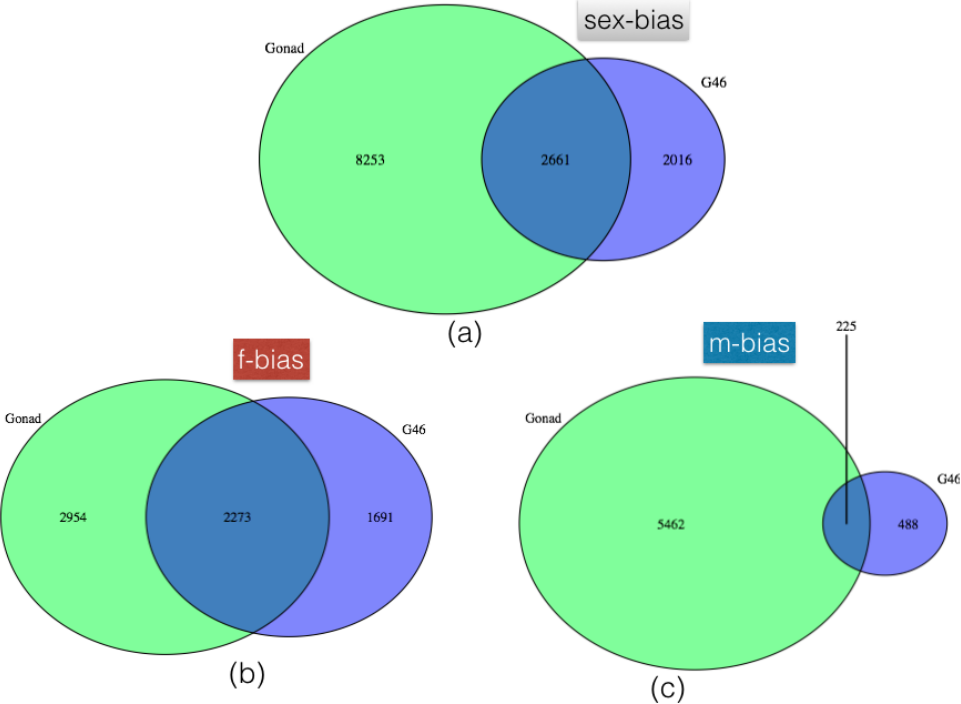
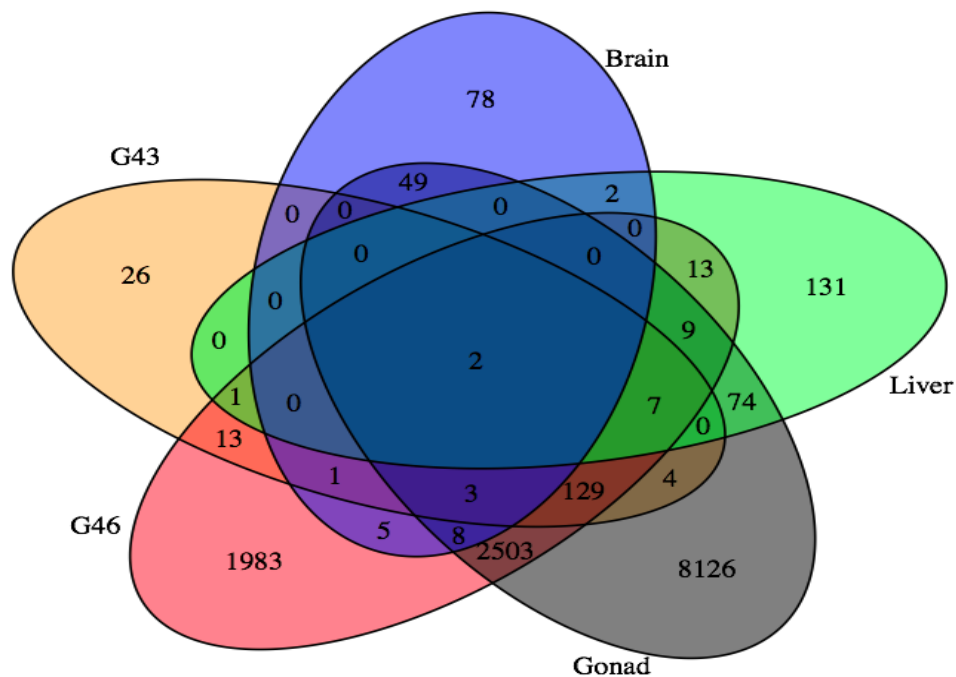


Figure 10. Shared sex-biased genes between tissues of G43, G46, and adult gonad, liver and brain tissues.



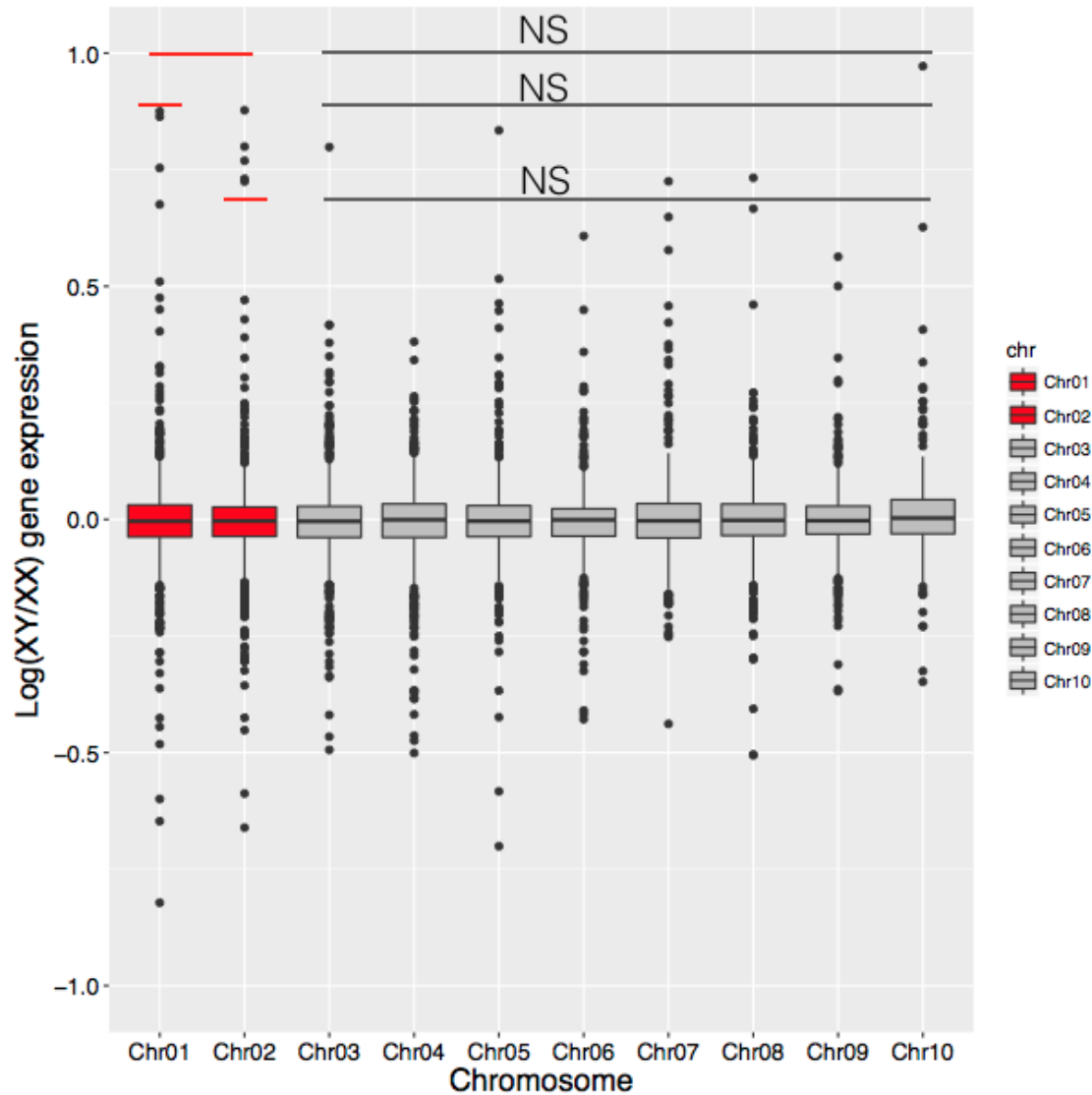
It is interesting to know the proportion of sex-biased genes in certain comparisons, will add that later on.

C. Transcription degeneration

To investigate possible transcriptional degeneration on the Y chromosome at early sex chromosome evolution, we compare gene expression ratio of XY individuals with testis and XX individuals with testis, $\text{Log}_2(\text{XY}/\text{XX})$.

Figure 10. Boxplot of gene expression ratio between individuals of XY with testis and XX with testis. Note, chromosome 1 and 2 are both shown to have male specific haplotypes.

No significant differences between gene expression ratio on chr 1 and autosomes, or chr 2 and autosomes, or chr1+chr2 and autosomes.



D. Coding sequence divergence and Faster-X

Approach: find the longest ORFs per transcript, use 1:1 ortholog with *X.tropicalis* to locate the genome locations for each transcript, finally use PLINK to perform codon alignment and finally calculate dN, dS, dN/dS with the Codeml model in PAML.

Figure 11. dN/dS ratio of sex-biased and unbiased genes throughout development and three adult tissues. The sample number are included in each category.

M= male-bias, F= female-bias, U= unbiased, Signif. codes: 0 '***', 0.001 '**', 0.01 '*', 0.05 '.

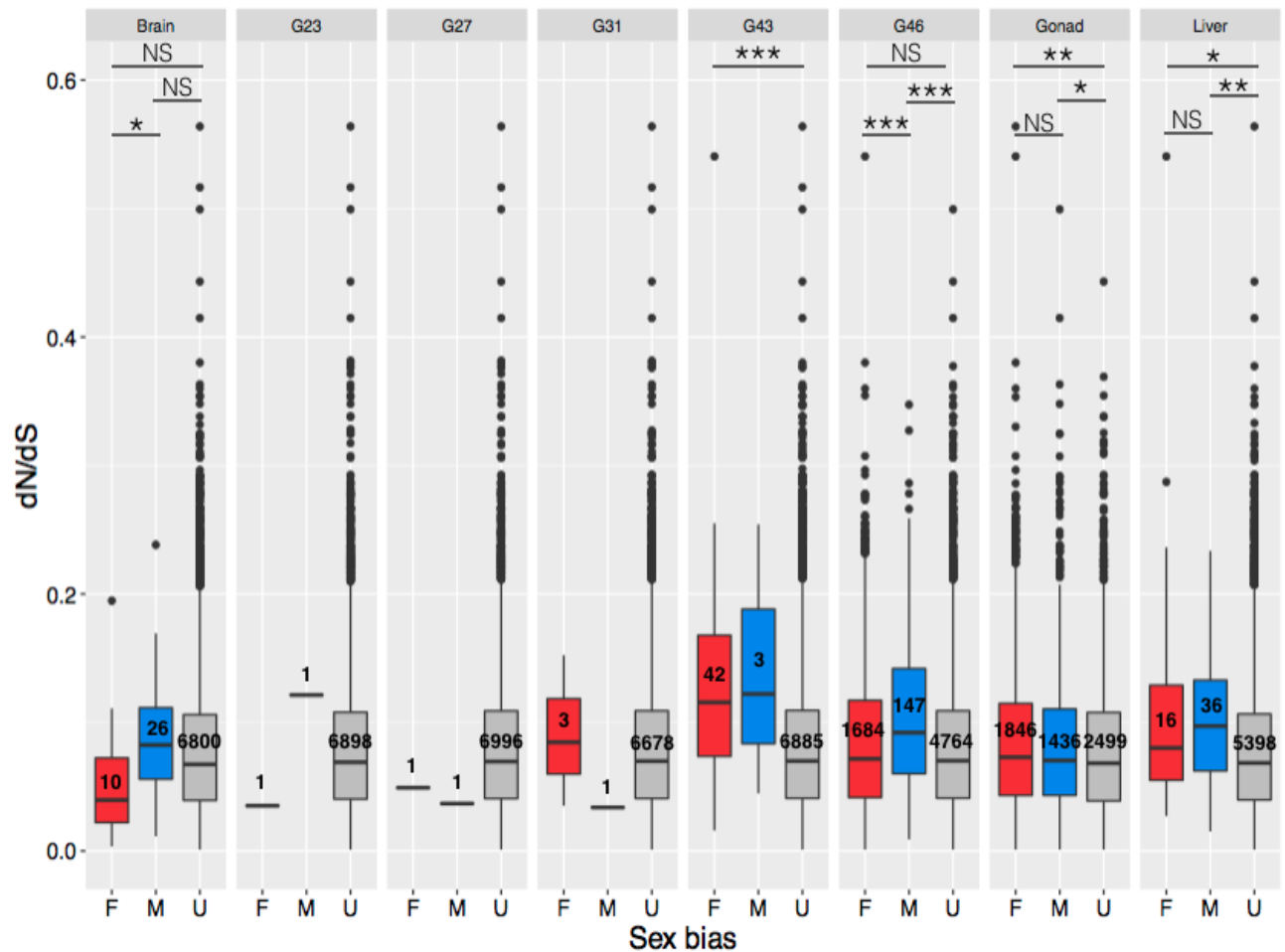
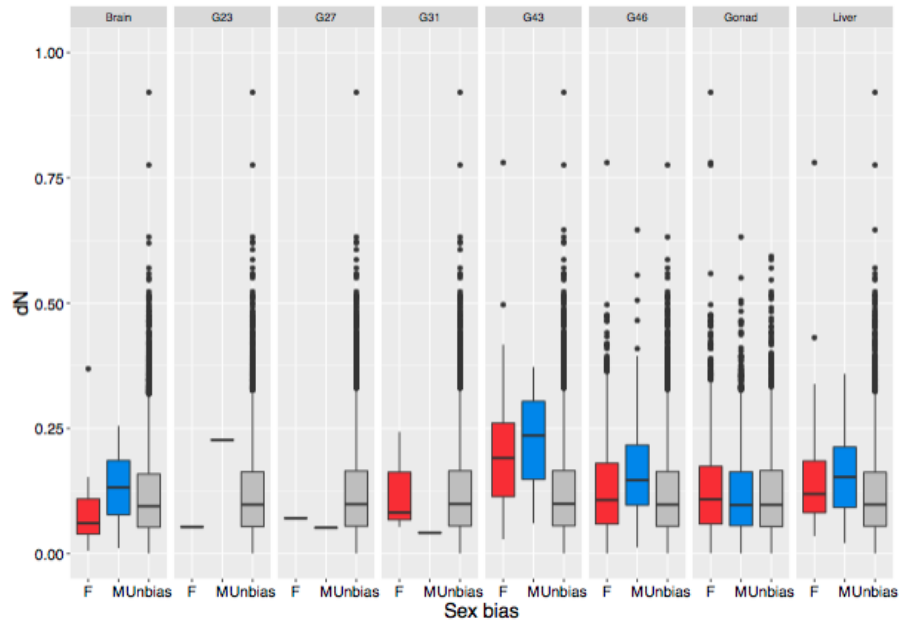
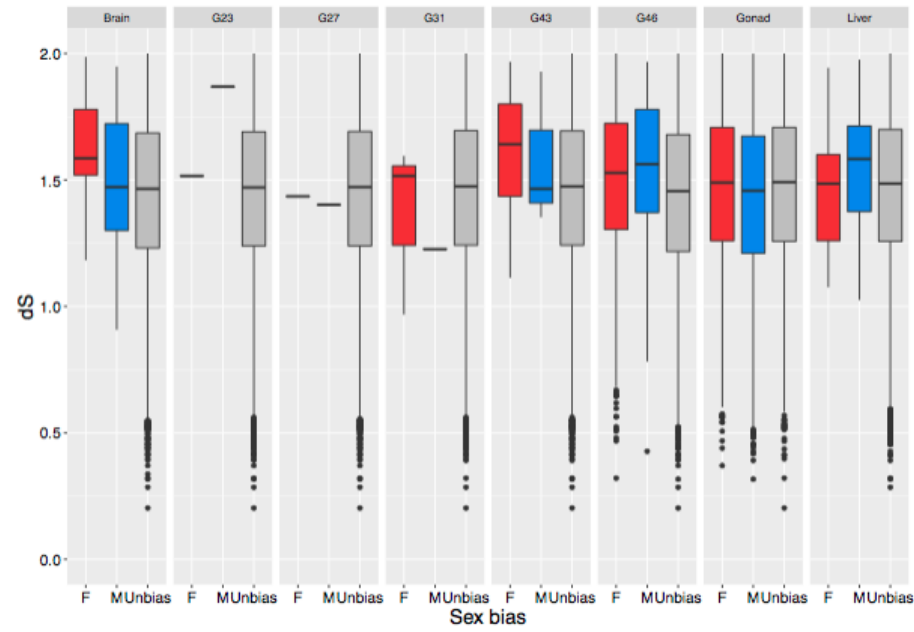


Figure 12. dN (a) and dS (b) values of sex-biased and unbiased genes throughout development and three adult tissues.
This is for supplementary materials i think, need to add the significance levels.

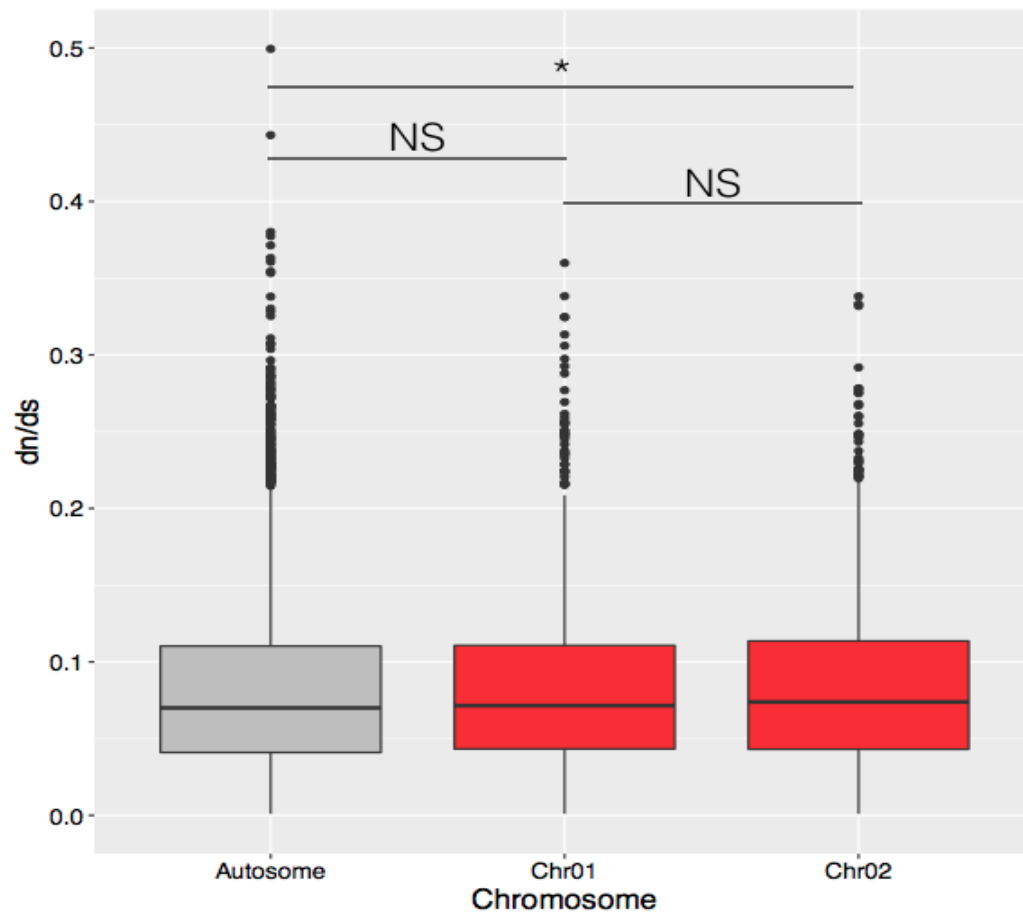


(a)



(b)

Figure 12. dN/dS ratio of genes on sex chromosomes (chromosome 1 and chromosome 2) and autosomes.



Here we update the results with transcriptome without non-frog transcripts, we still need to do get XX and XY transcriptome, by replacing SNPs of current transcriptome by mapping XX individual and XY individual respectively.