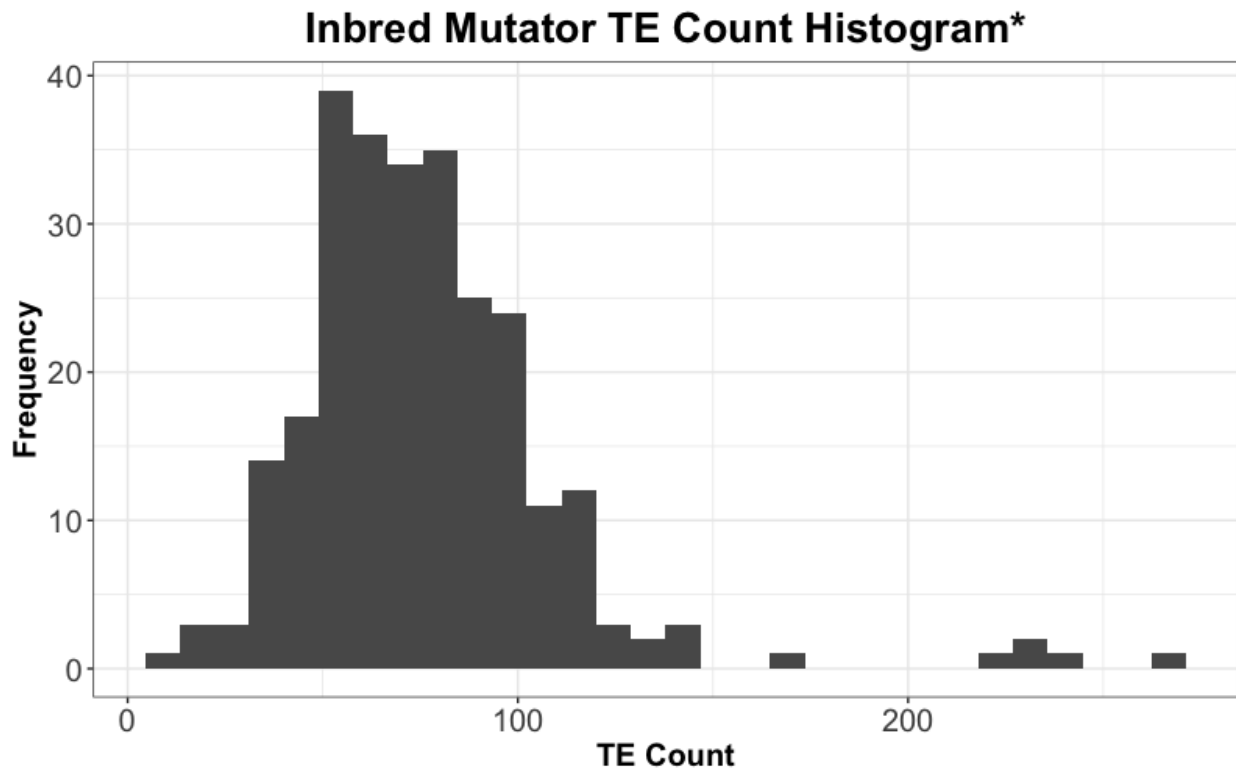TE count histogram

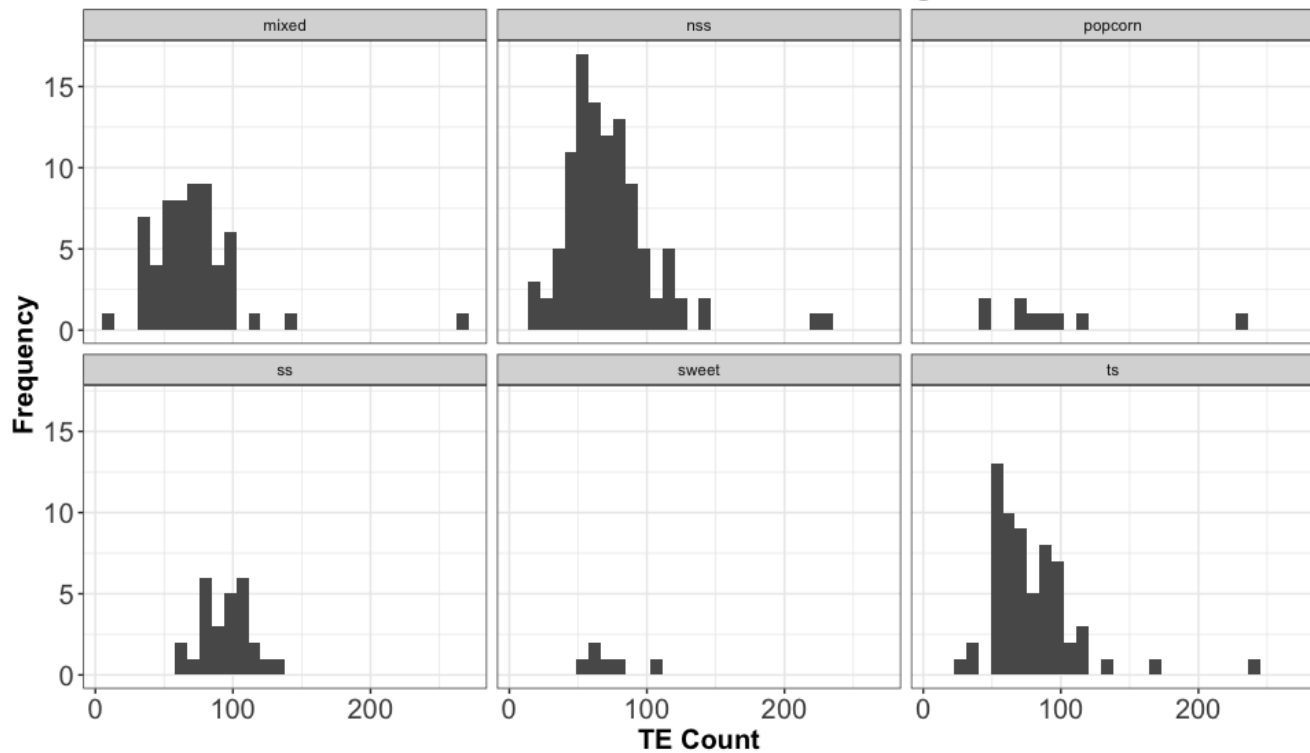**Inbred Mutator TE Count Histogram***

*histogram does not include Inbreds with insertions >300, as they had 500-1000 insertions
* "A6" "B57" "VaW6" excluded from barchart

```
> summary(adj_gene_region_summary$te_count_assigned)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   8.00   55.75   72.00   76.77   92.00  266.00
>
```
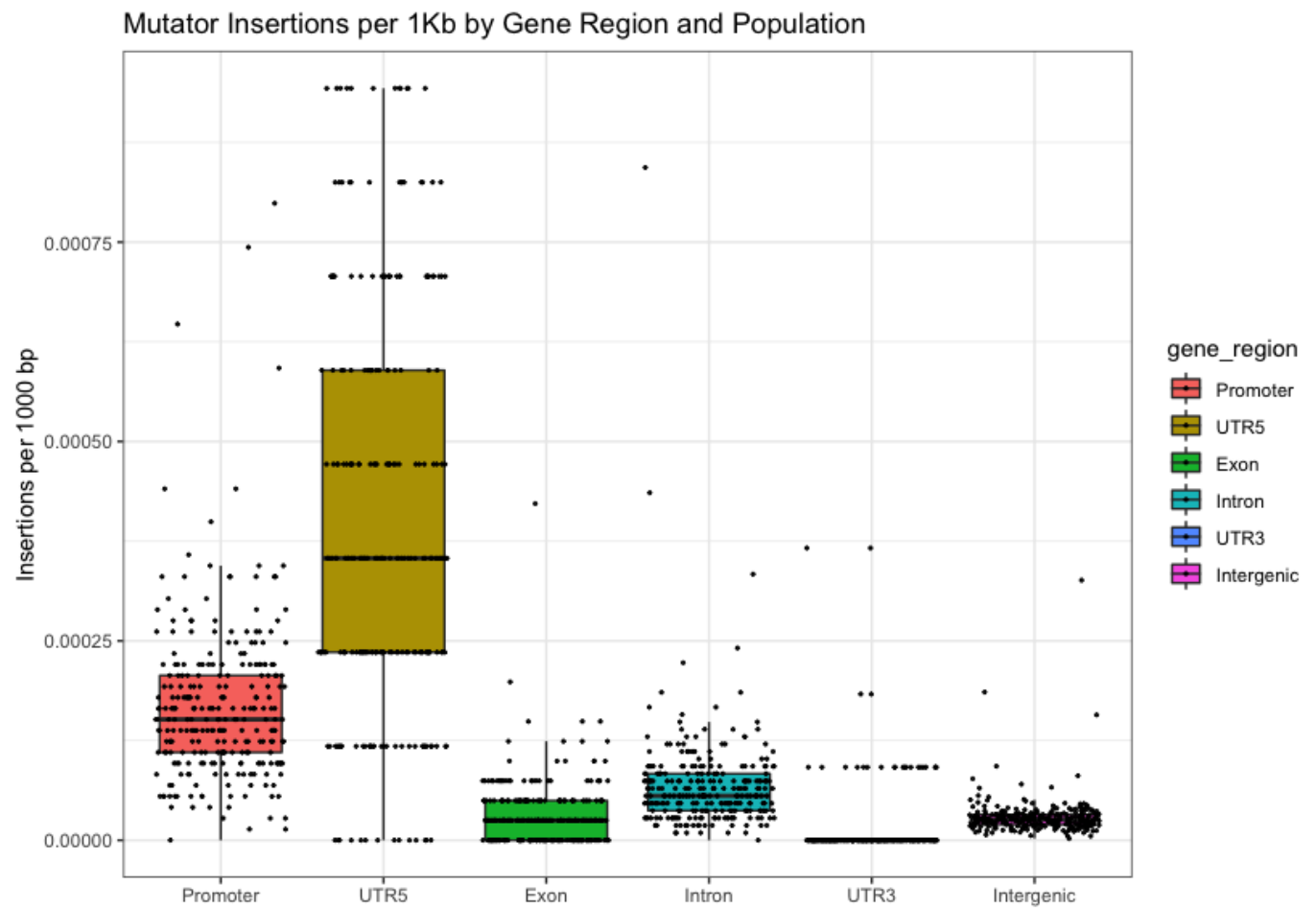
TE Count Histogram by Population

Inbred Mutator TE Count Histogram*
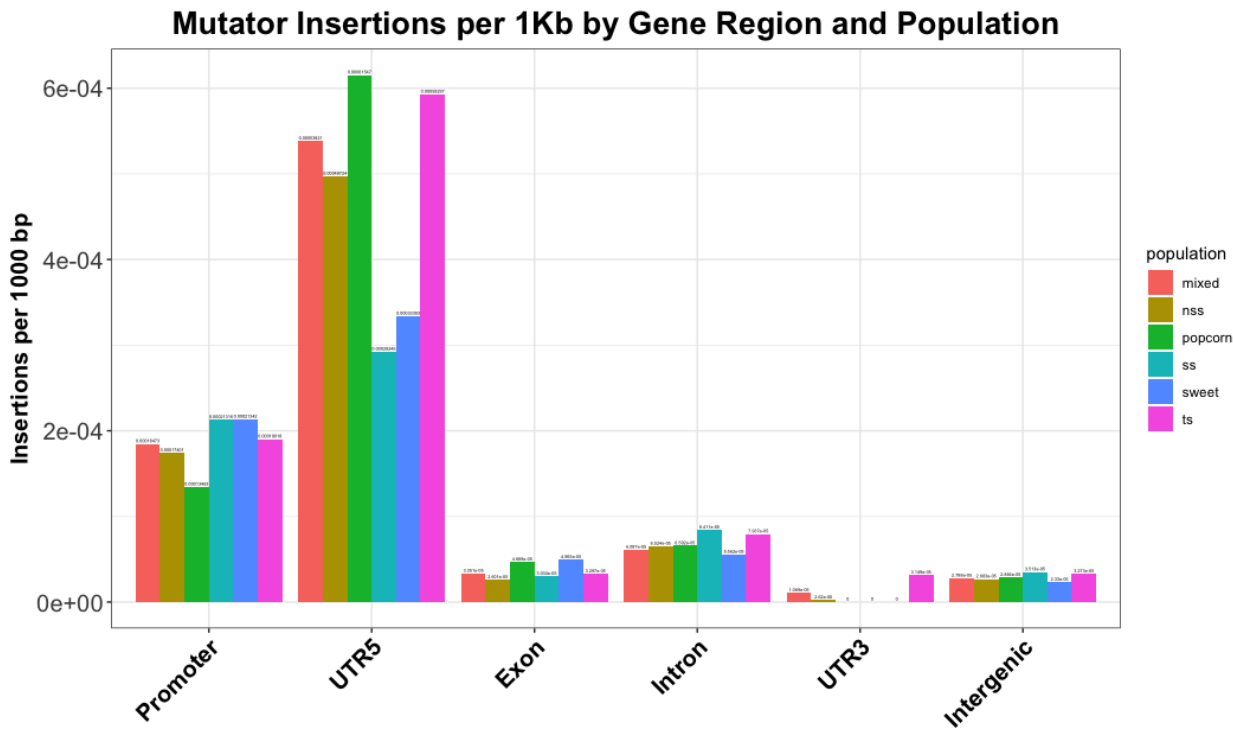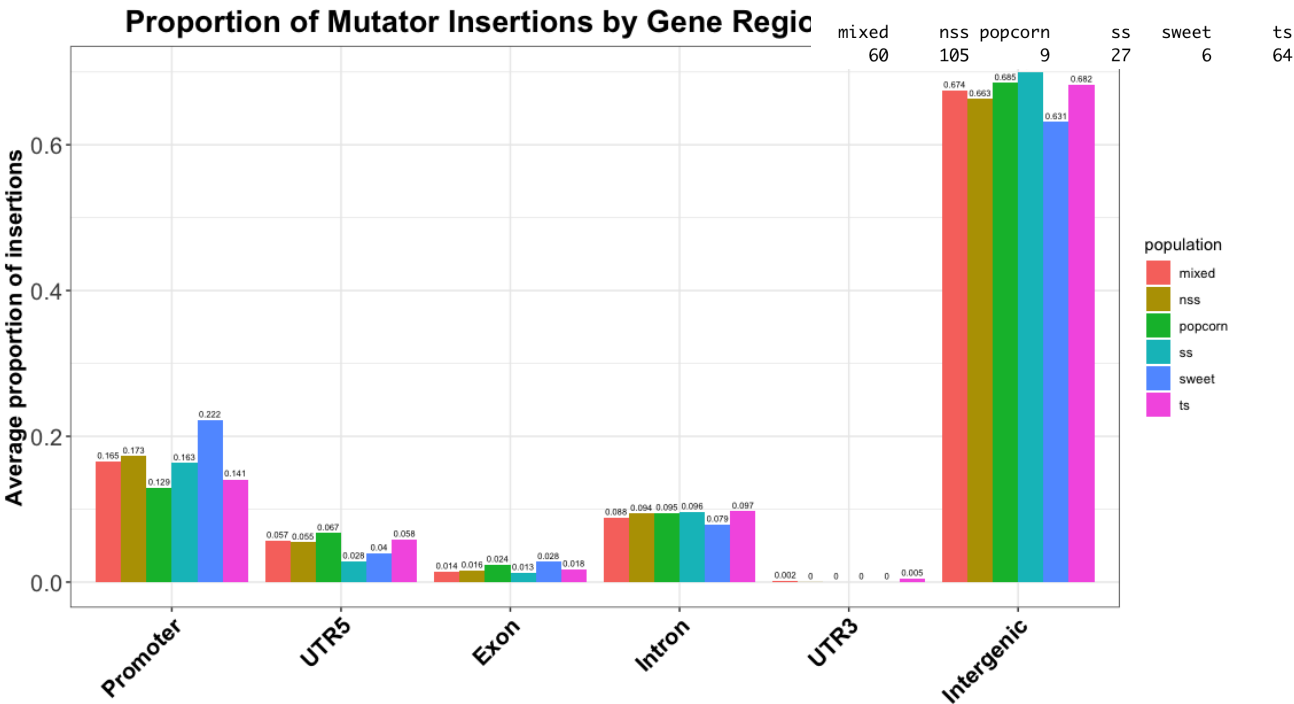


```
> quantile(gpi_gene_region_summary[which(gpi_gene_region_summary$Subpopulation == "mixed"),]$te_count_assigned)
    0%    25%    50%    75%   100%
  8.00  53.75  67.50  83.25 570.00
> quantile(gpi_gene_region_summary[which(gpi_gene_region_summary$Subpopulation == "nss"),]$te_count_assigned)
  0%  25%  50%  75% 100%
  15   52   67   85  406
> quantile(gpi_gene_region_summary[which(gpi_gene_region_summary$Subpopulation == "popcorn"),]$te_count_assigned)
  0%  25%  50%  75% 100%
  42   70   80   97  232
> quantile(gpi_gene_region_summary[which(gpi_gene_region_summary$Subpopulation == "ss"),]$te_count_assigned)
    0%    25%    50%    75%   100%
  61.0   82.5   95.0  108.5  133.0
> quantile(gpi_gene_region_summary[which(gpi_gene_region_summary$Subpopulation == "sweet"),]$te_count_assigned)
    0%    25%    50%    75%   100%
 53.00  61.00  65.50  76.75 105.00
> quantile(gpi_gene_region_summary[which(gpi_gene_region_summary$Subpopulation == "ts"),]$te_count_assigned)
    0%    25%    50%    75%   100%
 25.00  57.75  72.50  93.25 937.00
```

ins/1kb by gene region

Mutator Insertions per 1Kb by Gene Region and Population

ins/kb by pop + gene region

> table(gpi_gene_region_summary$Subpopulation)

mixed   nss popcorn   ss   sweet   ts
   60   105   9   27   6   64

Proportion of Mutator Insertions by Gene Region

Mutator Insertions per 1Kb by Gene Region and Population

linear model summaries

```
Call:
glm(formula = promoter_ins_prop ~ te_count_assigned + Subpopulation,
    data = gpi_gene_region_summary)

Deviance Residuals:
     Min          1Q      Median          3Q         Max
-0.139129   -0.037436   -0.004487    0.033055    0.240547

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)             1.635e-01  8.798e-03  18.579   <2e-16 ***
te_count_assigned       2.299e-05  5.157e-05   0.446   0.6562
Subpopulationnss        7.313e-03  9.749e-03   0.750   0.4538
Subpopulationpopcorn   -3.630e-02  2.154e-02  -1.685   0.0931 .
Subpopulationss        -2.519e-03  1.398e-02  -0.180   0.8571
Subpopulationsweet      5.723e-02  2.579e-02   2.219   0.0273 *
Subpopulationts        -2.491e-02  1.084e-02  -2.298   0.0224 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.003626671)

    Null deviance: 1.03101  on 270  degrees of freedom
Residual deviance: 0.95744  on 264  degrees of freedom
AIC: -744.9

Number of Fisher Scoring iterations: 2


Call:
glm(formula = utr5_ins_prop ~ te_count_assigned + Subpopulation,
    data = gpi_gene_region_summary)

Deviance Residuals:
     Min          1Q      Median          3Q         Max
-0.058020   -0.022797   -0.003975    0.019066    0.082943

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)             5.683e-02  4.637e-03  12.256  < 2e-16 ***
te_count_assigned       1.328e-06  2.718e-05   0.049 0.961059
Subpopulationnss       -2.394e-03  5.138e-03  -0.466 0.641586
Subpopulationpopcorn    1.005e-02  1.135e-02   0.885 0.376912
Subpopulationss        -2.888e-02  7.368e-03  -3.920 0.000113 ***
Subpopulationsweet     -1.726e-02  1.359e-02  -1.270 0.205354
Subpopulationts         1.097e-03  5.713e-03   0.192 0.847840
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.001007347)

    Null deviance: 0.28832  on 270  degrees of freedom
Residual deviance: 0.26594  on 264  degrees of freedom
AIC: -1092

Number of Fisher Scoring iterations: 2
```
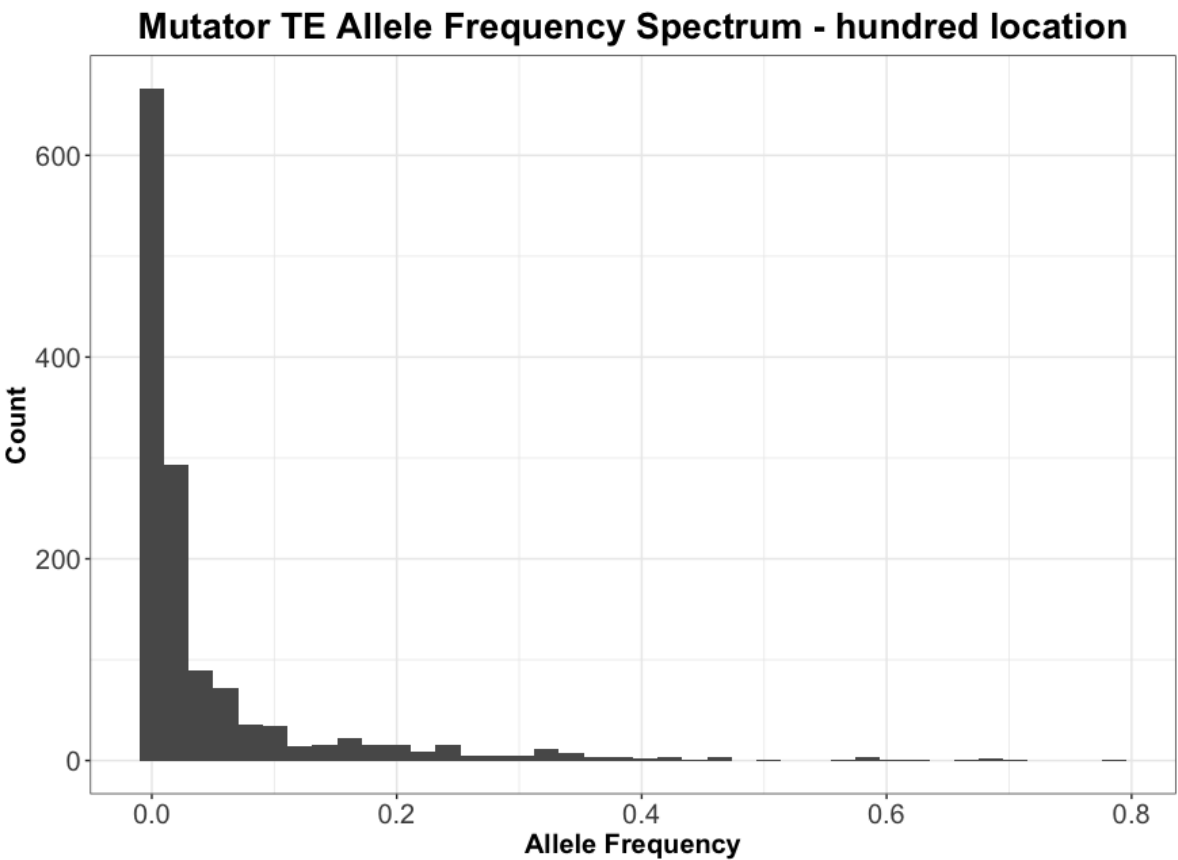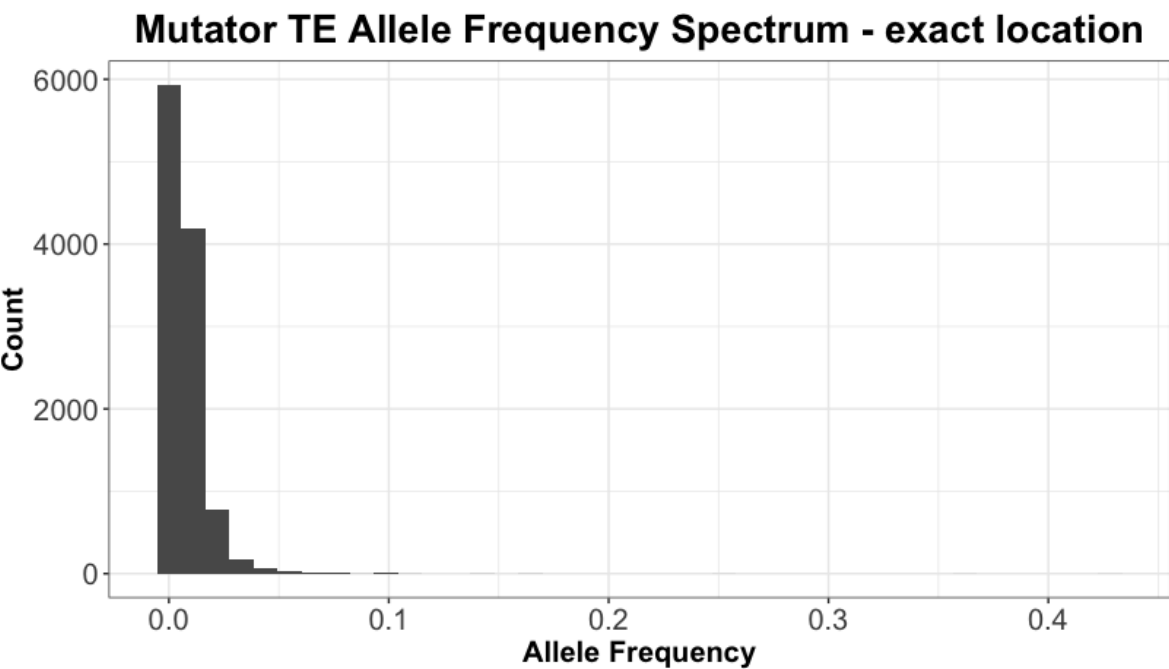
Ins matrix PCAs

[Folder with all PCAs](#) - each PCA plot has a version without inbred label and with inbred label
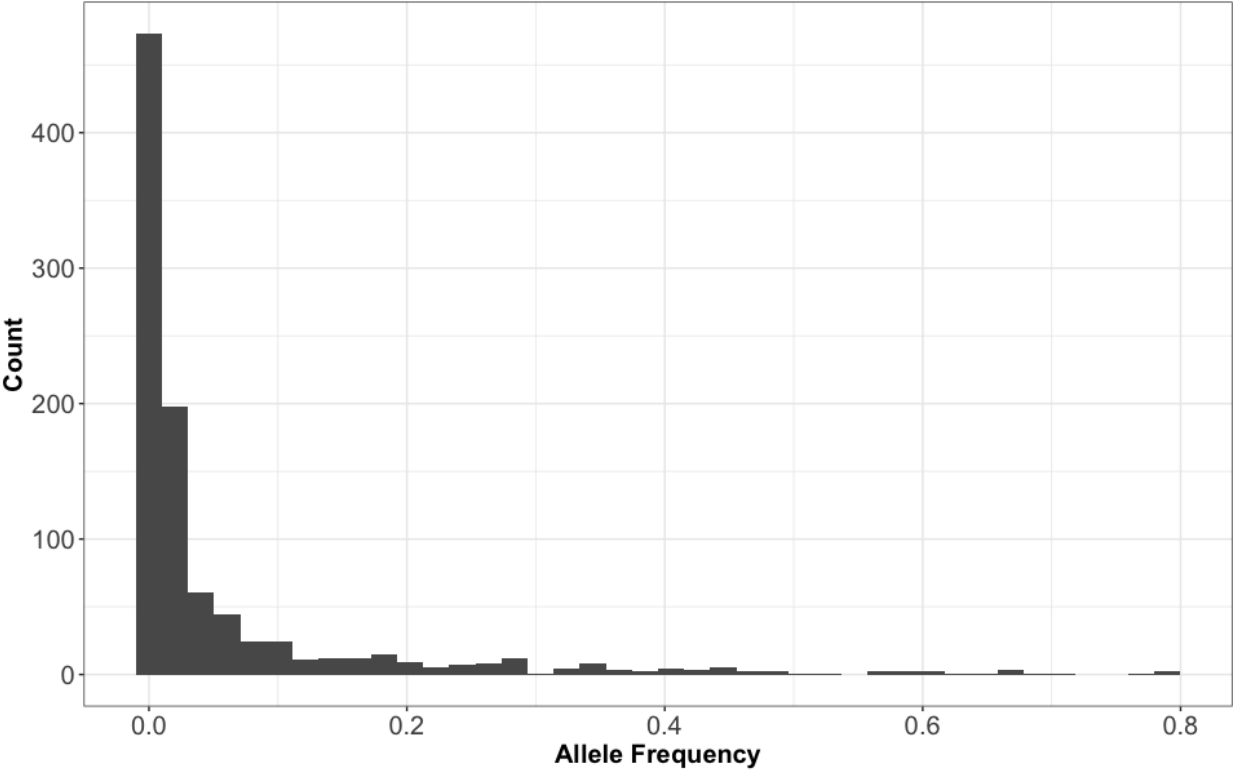Exact - ins matrix based on ins within 18 bp
Hundred - ins matrix based on ins rounded to nearest 100
Thousand - ins matrix based on ins rounded to nearest 1000
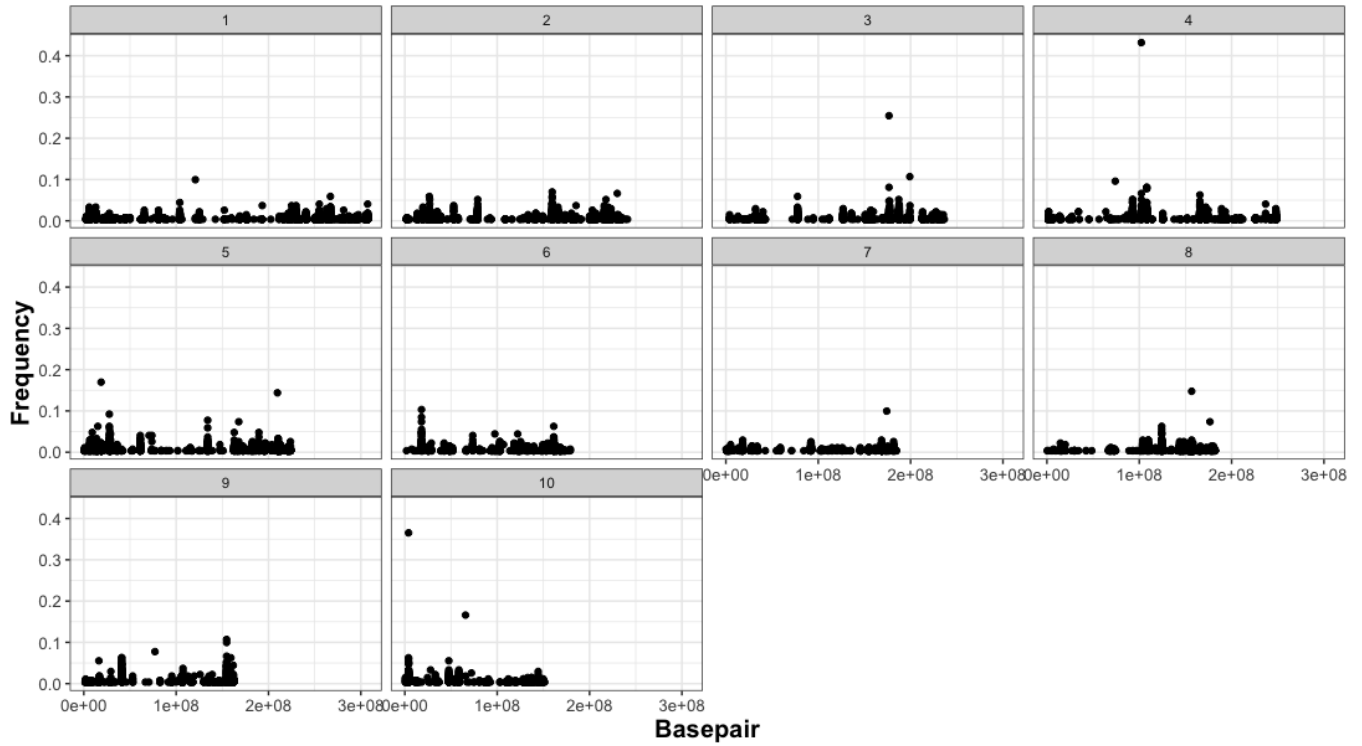
Allele freq spectrum hist

Mutator TE Allele Frequency Spectrum - exact location

Mutator TE Allele Frequency Spectrum - hundred location

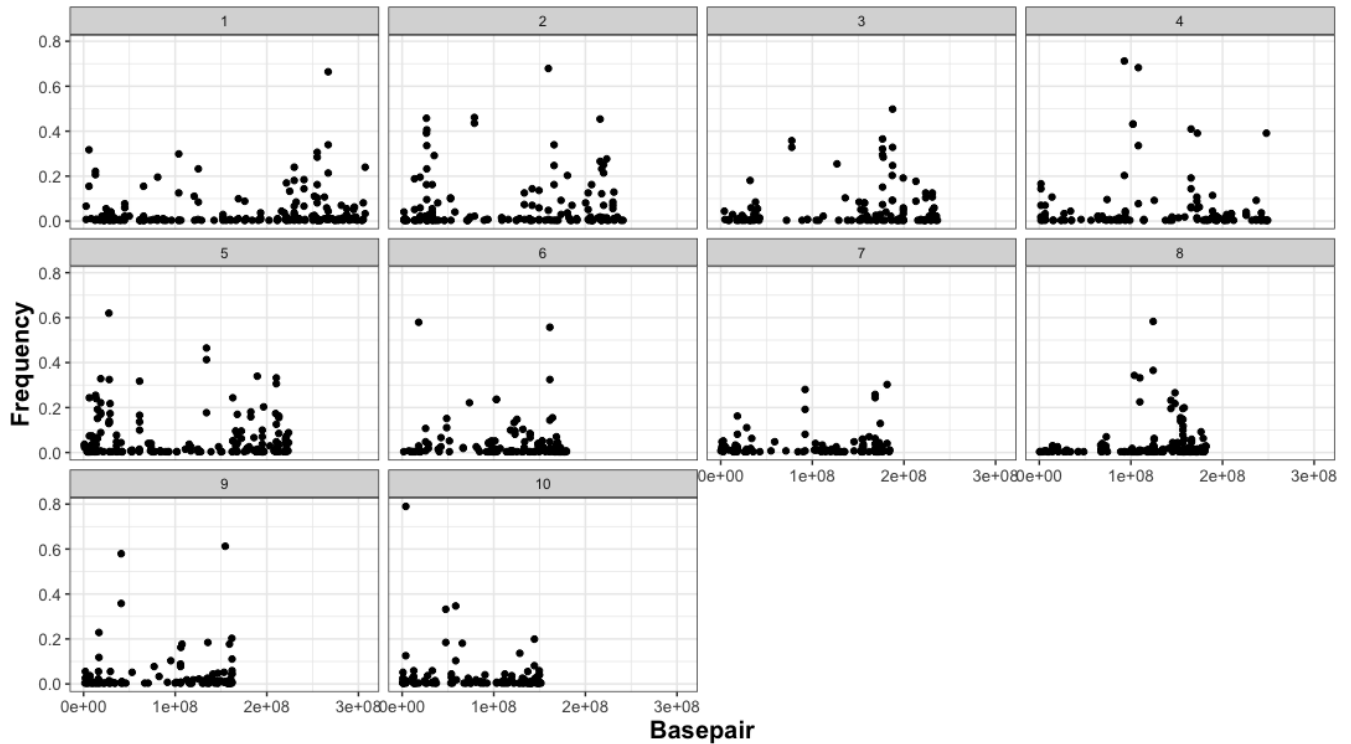**Mutator TE Allele Frequency Spectrum - thousand location**

Allele freq by chrom hist

Mutator TE Allele Frequency by chromosome and bp location - exact location

Mutator TE Allele Frequency by chromosome and bp location - hundred location

Mutator TE Allele Frequency by chromosome and bp location - thousand location