

# Uncovering imprinting of PAV genes and transposable elements using whole genome assemblies

Kaitlin Higgins<sup>1</sup>, Peng Zhou<sup>2</sup>, Yaniv Brandvain<sup>2</sup>, Nathan Springer<sup>2</sup>, Sarah Anderson<sup>1</sup>

<sup>1</sup> Department of Genetics, Development, and Cell Biology; Iowa State University; Ames, IA

<sup>2</sup> Department of Plant and Microbial Biology; University of Minnesota; St. Paul, MN

## An overview of Genomic Imprinting

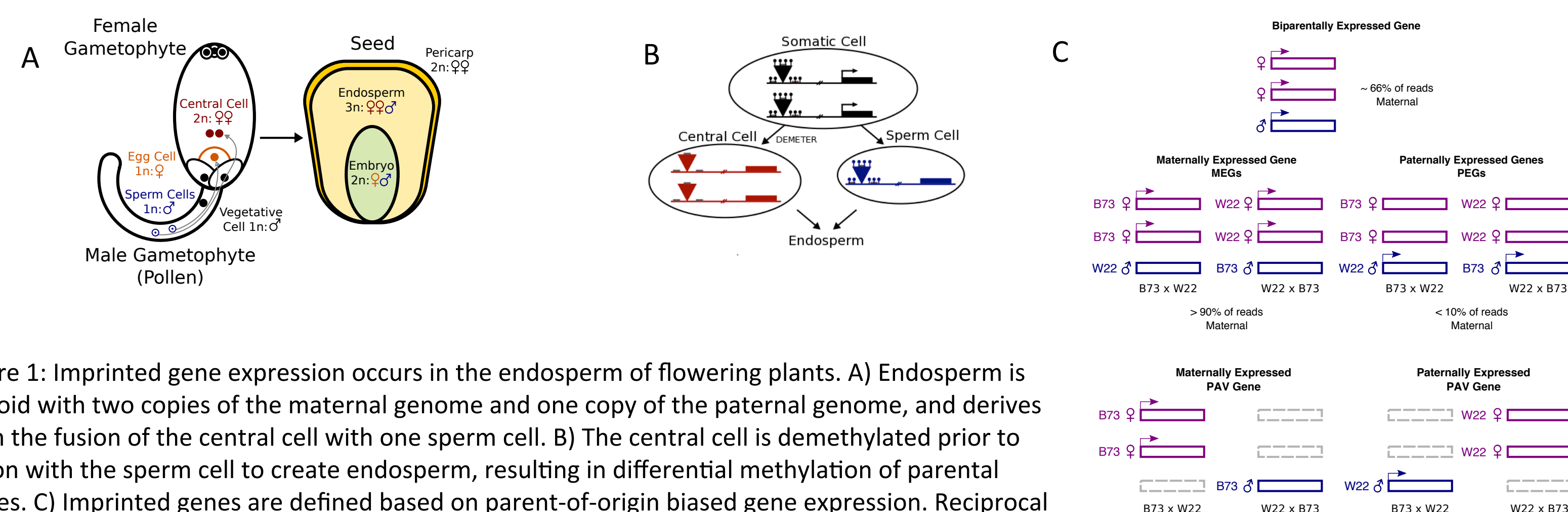


Figure 1: Imprinted gene expression occurs in the endosperm of flowering plants. A) Endosperm is triploid with two copies of the maternal genome and one copy of the paternal genome, and derives from the fusion of the central cell with one sperm cell. B) The central cell is demethylated prior to fusion with the sperm cell to create endosperm, resulting in differential methylation of parental alleles. C) Imprinted genes are defined based on parent-of-origin biased gene expression. Reciprocal crosses are required to distinguish genotype effects from parent-of-origin effects, and alleles from each genome can be tested individually for these patterns.

## Preliminary Data

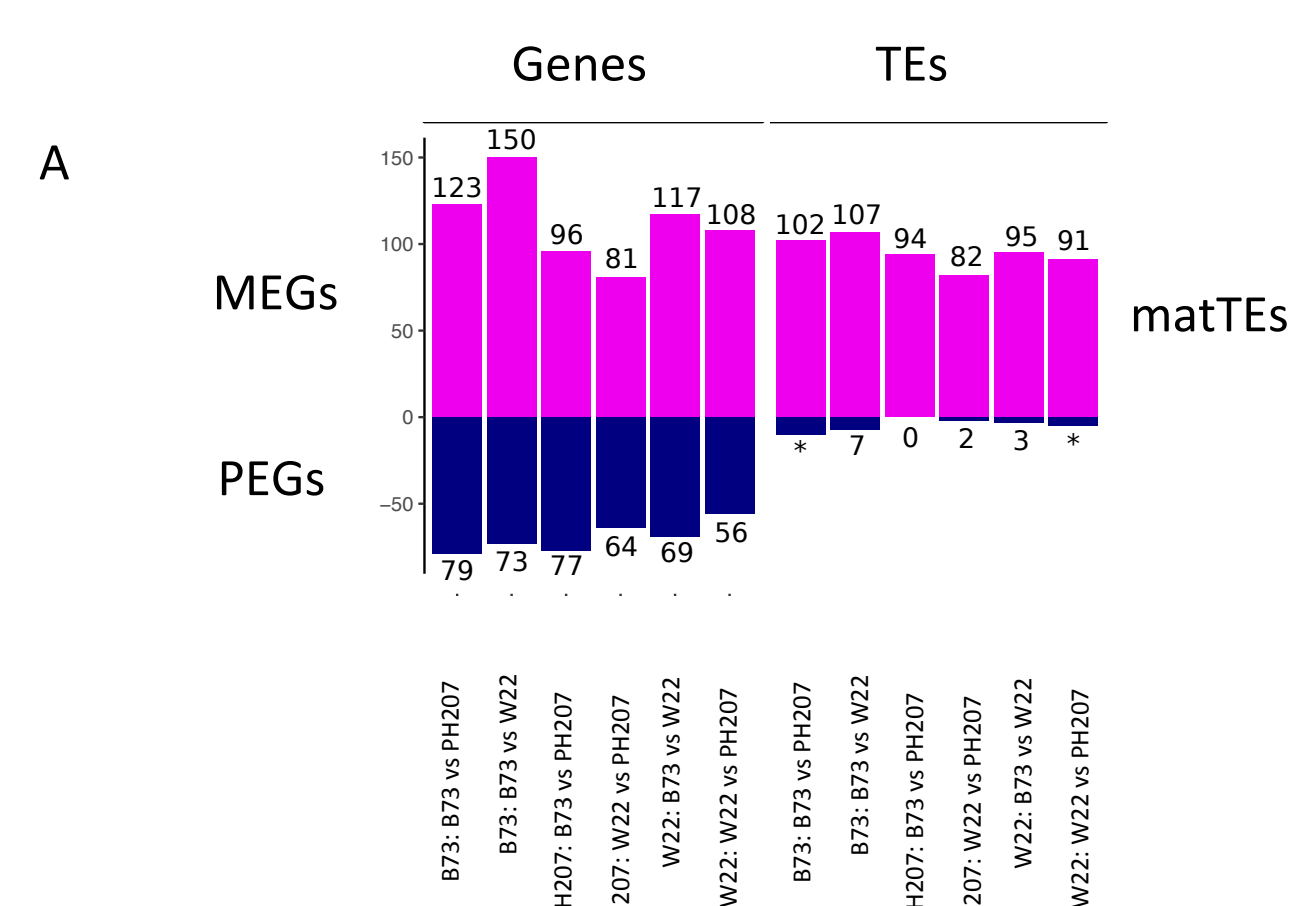
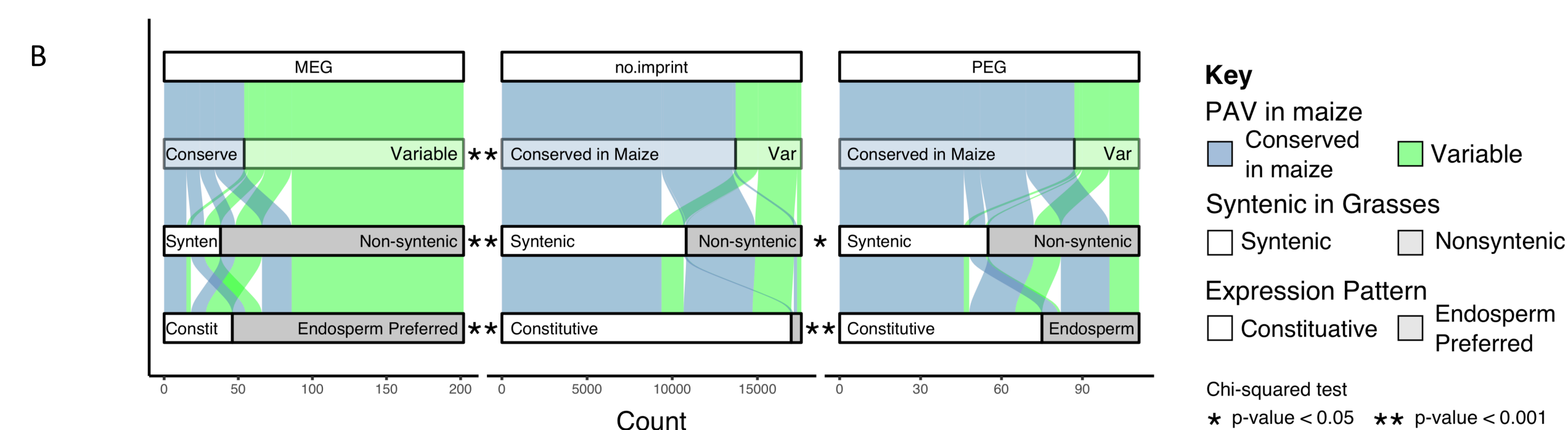


Figure 2: Defining imprinted genes in the endosperm. A) MEGs, PEGs, and maternally expressed TEs (matTEs) were identified across all contrasts and genome assemblies. B) Features of imprinted and non-imprinted B73 genes. Imprinted genes, particularly MEGs, are enriched for variability within maize and other grass genomes as well as expression specific to the endosperm, suggesting de-repression specifically of the maternal allele in this tissue.



Key  
PAV in maize  
Conserved in maize  
Variable  
Syntenic in Grasses  
Syntenic  
Non-syntenic  
Expression Pattern  
Constitutive  
Endosperm Preferred  
Chi-squared test  
\* p-value < 0.05 \*\* p-value < 0.001

## Maternally expressed TEs enriched near maternally expressed genes

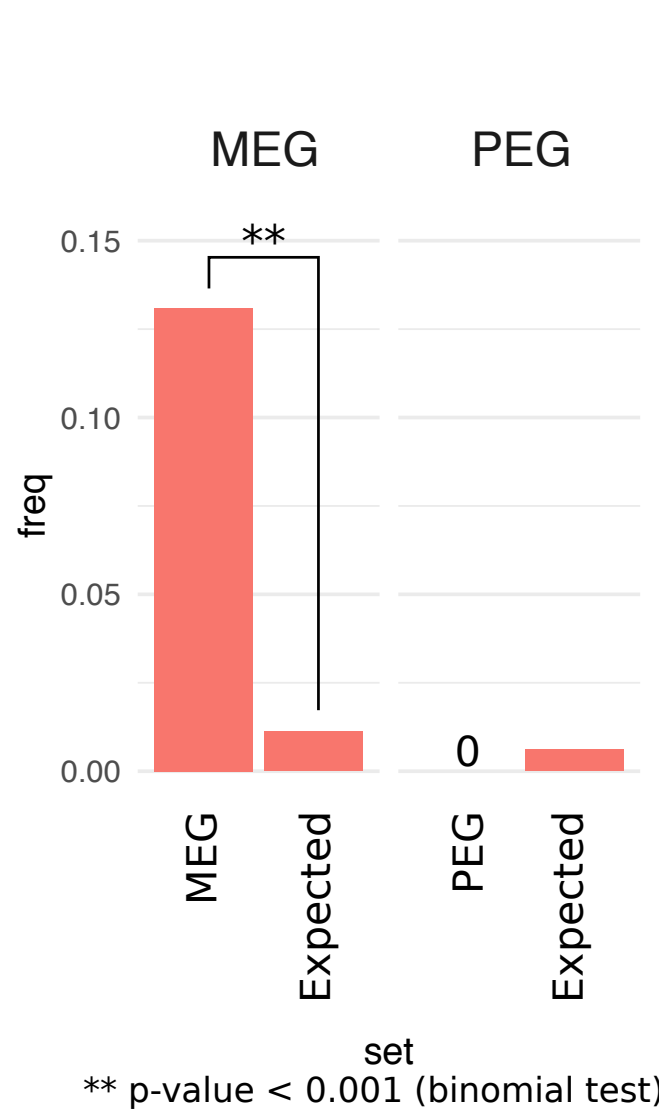


Figure 3  
The proportion of maternally expressed TEs where the nearest gene is a MEG or PEG is shown compared to the expectation based on the number of MEGs or PEGs among genes assessed for imprinting. The nearest gene to maternal TEs is a MEG 11.6 x more frequently expected.

Gene ID	TE superfamily	Gene Variability	TE Imprint	Gene Imprint	distance
Zm00001d004147	DHH	variable	matTE	MEG	0
Zm00001d005712	RLC	variable	matTE	MEG	-29185
Zm00001d005712	RLG	variable	matTE	MEG	-3087
Zm00001d007488	RLC	variable	matTE	MEG	-852
Zm00001d017481	DHH	variable	matTE	MEG	1078
Zm00001d023985	RLX	variable	matTE	MEG	0
Zm00001d026623	RLC	conserved.maize	matTE	MEG	-797
Zm00001d029042	DHH	variable	matTE	MEG	0
Zm00001d031633	RLG	variable	matTE	MEG	-925
Zm00001d038034	RLG	variable	matTE	MEG	-2662
Zm00001d041755	RLX	variable	matTE	MEG	-1453
Zm00001d041887	DHH	variable	matTE	MEG	0
Zm00001d043716	RLX	conserved.maize	matTE	MEG	0
Zm00001d043716	RLX	conserved.maize	matTE	MEG	1670
Zm00001d046395	RLG	variable	matTE	MEG	-47897
Zm00001d046395	RLC	variable	matTE	MEG	-10170
Zm00001d046395	RLG	variable	matTE	MEG	-3398
Zm00001d048646	DHH	variable	matTE	MEG	0
Zm00001d050068	DHH	variable	matTE	MEG	0

Table 1: Cases where nearest gene to maternally expressed TE is a MEG. Genes with more than one nearby imprinted TE are boxed in black. The distance relative to the transcription start site of the gene is noted on the right with upstream TEs in green, and downstream TEs in red. Zeros denote TEs that either overlapped or were contained within genes, the ones within TEs (Also present in Table 2) are boxed in red.

## Conserved Imprinted Genes

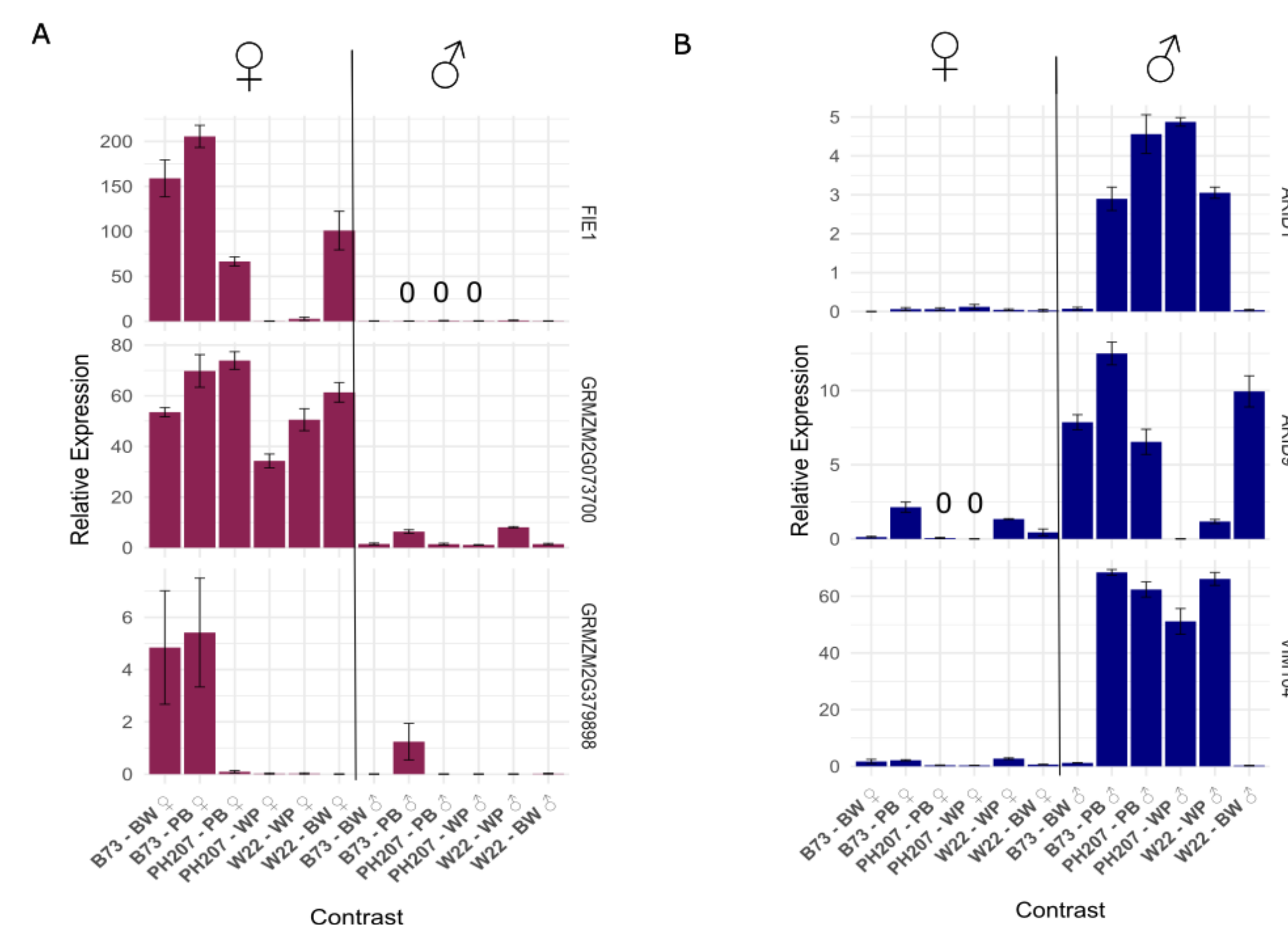


Figure 4: Relative expression levels of maternally expressed imprinted genes (A) and paternally expressed imprinted genes (B) with conserved imprinting status in maize, rice, and Arabidopsis (Waters et al. 2011). Expression for each syntelog is shown, and contrasts are labeled by genome, then genotype of the hybrid. Female and male symbols denote the parental source of alleles from specified genomes. True zeros are labeled, other low readings can be in part due to non-informative reads or few reads.

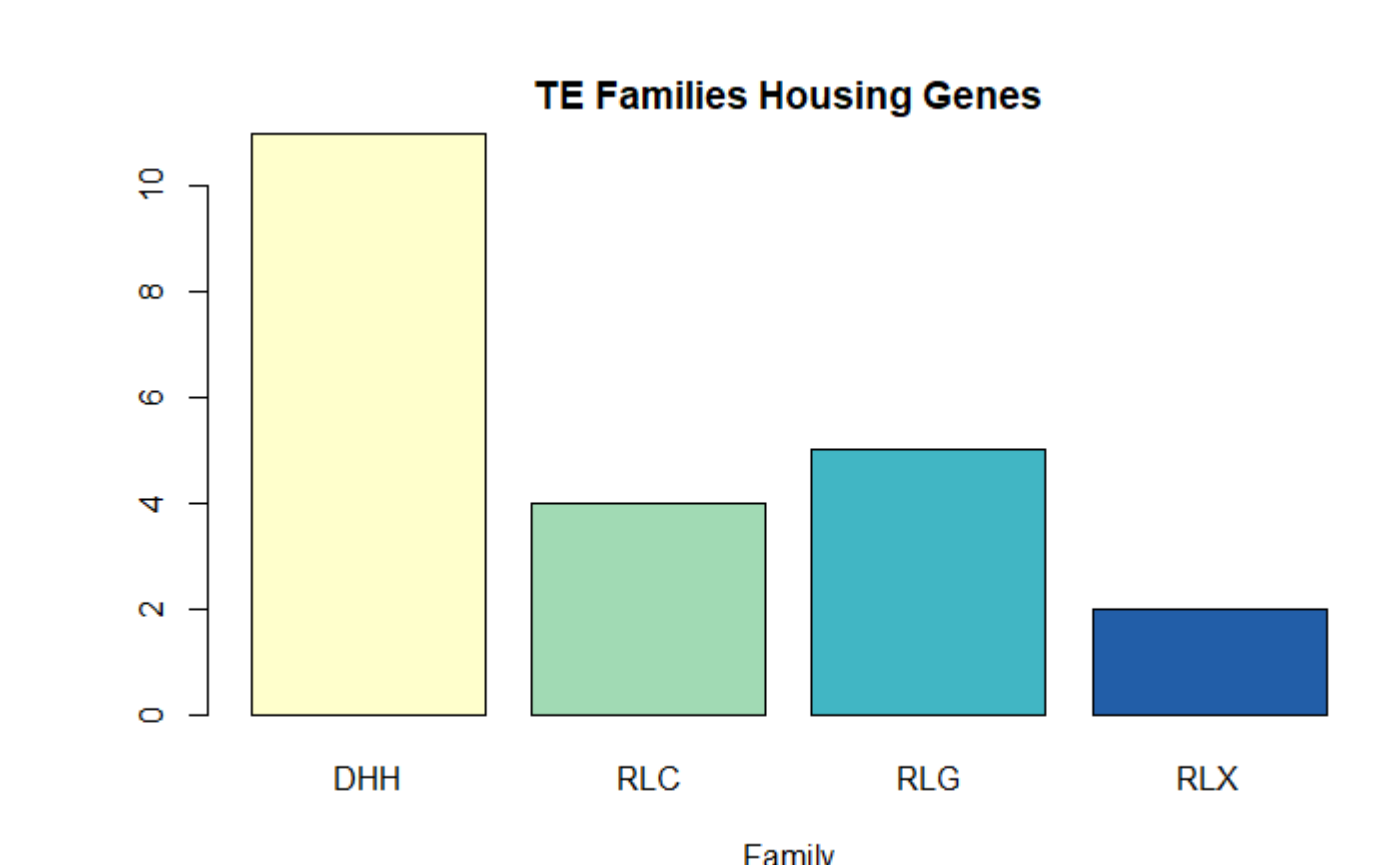
## Imprinted Genes can be Located within TEs

Gene ID	TE superfamily	Gene Imprint	TE imprint	Gene Conservation
Zm00001d029114	DHH	MEG	NA	variable
Zm00001d029237	DHH	MEG	NA	variable
Zm00001d004147	DHH	MEG	matTE	variable
Zm00001d007685	DHH	MEG	NA	variable
Zm00001d048646	DHH	MEG	matTE	variable
Zm00001d049284	DHH	MEG	NA	variable
Zm00001d053206	DHH	MEG	no.imprint	variable
Zm00001d024363	DHH	MEG	no.imprint	conserved.maize
Zm00001d029042	DHH	MEG	matTE	variable
Zm00001d031558	DHH	MEG	NA	variable
Zm00001d041887	DHH	MEG	matTE	variable
Zm00001d051953	DHH	MEG	NA	variable
Zm00001d052773	DHH	MEG	NA	variable
Zm00001d013760	DHH	MEG	NA	variable
Zm00001d014389	DHH	MEG	NA	variable
Zm00001d047301	DHH	MEG	NA	variable
Zm00001d042952	RLC	MEG	NA	variable
Zm00001d042399	RLC	MEG	NA	variable
Zm00001d030984	RLC	MEG	NA	variable
Zm00001d043716	RLC	MEG	no.imprint	conserved.maize
Zm00001d036112	RLC	PEG	NA	variable
Zm00001d040499	RLG	MEG	NA	variable
Zm00001d004147	RLG	MEG	no.imprint	variable
Zm00001d011714	RLG	MEG	no.imprint	variable
Zm00001d040499	RLG	MEG	NA	variable
Zm00001d013760	RLX	MEG	NA	variable
Zm00001d025207	RLX	MEG	NA	variable
Zm00001d029114	RLX	MEG	NA	variable
Zm00001d023985	RLX	MEG	matTE	variable
Zm00001d050344	RLX	MEG	no.imprint	variable
Zm00001d016728	RLX	MEG	NA	variable

Table 2: Imprinted genes located within TEs, noting which superfamily of TE they are located in, the imprint type, whether the TE they are in is imprinted and whether they are conserved. Red boxed Gene IDs are in maternally expressed TEs (also in Table 1).

Through this we see that 26 of 27 are maternally imprinted, and 25 of 27 are non conserved. We can also see that some TEs may house imprinted genes without themselves being imprinted.

Figure 5: Barchart of TE superfamilies containing or carrying imprinted genes. (DHH = helitron, RLC = copia LTR, RLG = gypsy LTR, and RLX = unknown LTR)



## Future Directions

- Characterize promoters of conserved imprinted genes
- Characterize maternally expressed TEs contained within genes
- Investigate relationship between TEs containing or carrying imprinted genes and imprinting status of genes
- Investigate relationship between TEs and MEGs and TEs and PEGs separately