Uncovering the Mystery of *Ufo1* in Maize

Jin Cui[1], Kameron Wittmeyer[1], Debamalya Chatterjee[1], Pearl Chang[2], Jui-Hsien Rita Lu[2], Ming-Ren Yen[2], Pao-Yang Chen[2], Tzuu-fen Lee[3], Blake Meyers[4], Surinder Chopra[1]

[1]Department of Plant Science, Pennsylvania State University, University Park, PA 16802
[2]Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan
[3]DuPont Pioneer, Des Moines, IA 50131
[4]The Donald Danforth Plant Science Center, St. Louis, MO 63132

This work is supported by the Natural Science Foundation (Grant Number: 1051654)

Introduction

The maize *Unstable factor for orange1 (Ufo1)* is a dominant mutation which gives rise to ectopic phlobaphene accumulation in various tissues. The mutant *Ufo1-1* exhibits poor penetrance and variable expressivity, which is associated with variable degree of expression and DNA methylation of *pericarp color1 (p1)*, a Myb transcription factor. Fine mapping has narrowed *Ufo1* gene to a 40Mb region in chromosome 10 near the centromere.

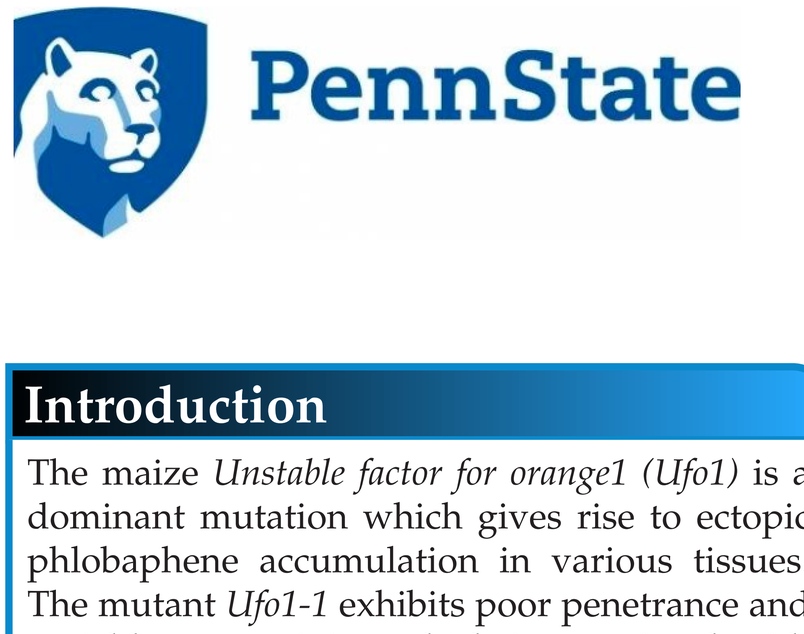

Figure 1: A) The ear and plant body phenotype of wild-type (WT) and *Ufo1-1*. B) variable expressivity of *Ufo1-1*. C) fine mapping of *Ufo1-1* on chromosome 10.

Transcriptomic analysis of *Ufo1-1* has helped identify a candidate gene (GRMZM2G053177) within the mapping region. Transgenic overexpression of GRMZM2G053177 phenocopies *Ufo1-1* mutant.

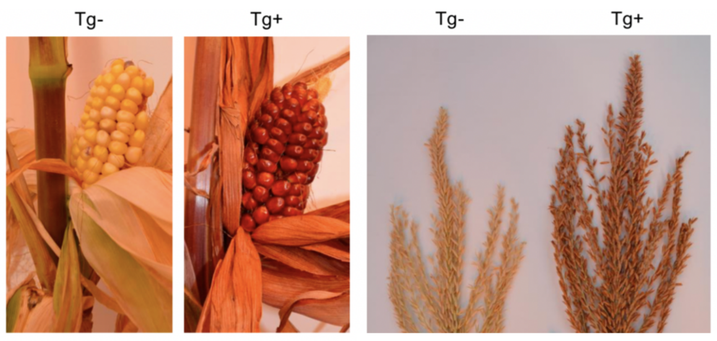

Figure 2: Ear (left) and tassel (right) phenotype of non-transgenic (Tg-) and transgenic (Tg+) plants.

Meanwhile, we are analyzing small RNA and whole-genome bisulfite sequencing data. We are interested in answering these **questions**:

1. What is the gene structure and possible function of GRMZM2G53177?
2. Could the variable expressivity of the mutant be related to epigenetic mechanisms?
3. What is the small RNA and DNA methylation profile in the mutant?

GRMZM2G053177 and CACTA

Our PacBio genome assembly of *Ufo1-1* reveals a novel CACTA transposon is inserted in the 5′-UTR region of GRMZM2G053177, giving rise to a novel transcript in the mutant.

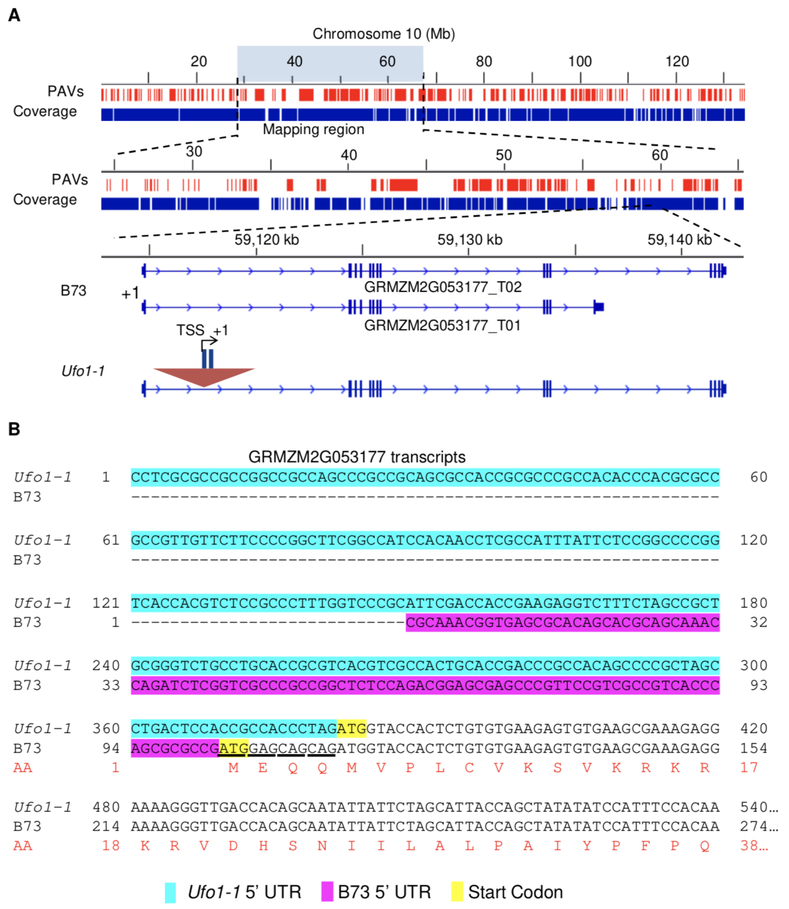

Figure 4: A CACTA transposon is inserted in GRMZM2G053177 in *Ufo1-1*. A) Alignment of *Ufo1-1* PacBio genome assembly with the reference genome (B73 RefGen v4) to identify presence absence variations (PAV, shown in red). CACTA insertion is indicated in the red triangle. B) Amino acid alignment shows CACTA insertion in the 5′-UTR of GRMZM2G053177.

Ufo1-1 expressivity is associated with DNA methylation as well as smRNA abundance in the CACTA element.

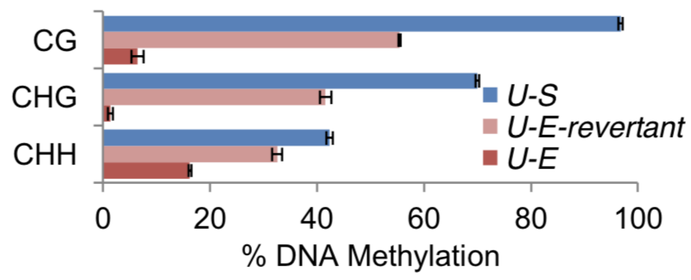

Figure 5: Bisulfite sequencing of CACTA transposon in various samples. *U-E* revertant plants are progeny of *U-S* which regained pigmentation.

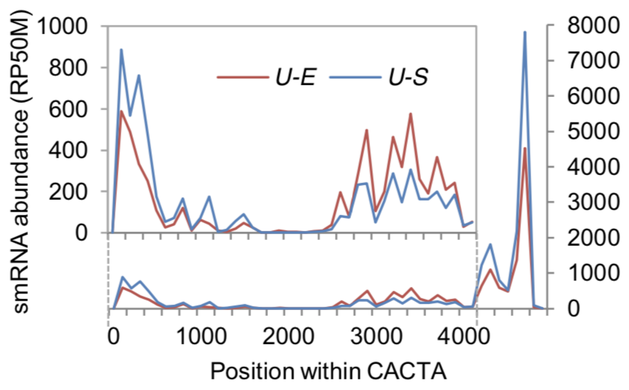

Figure 6: smRNA level of CACTA in *U-E* and *U-S*.

smRNA and DNA methylation

Further analysis on DNA methylation reveals that changes in 24-nt smRNA is mostly associated with changes in CHH methylation.

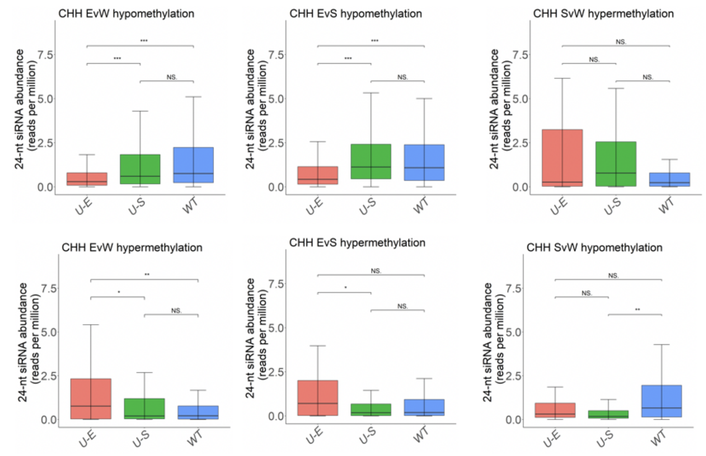

Figure 8: Correlation between 24-nt siRNA and CHH methylation. EvW: *U-E* vs. WT, EvS: *U-E* vs. *U-S*, SvW: *U-S* vs. WT.

Gene body methylation

Our whole-genome bisulfite sequencing data shows that *Ufo1-1* promotes CG methylation in gene body region.

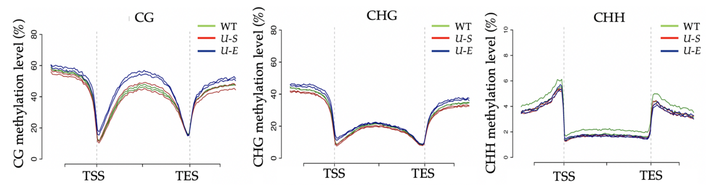

Figure 9: Metagene plot showing CG (left), CHG (middle) and CHH (right) DNA methylation level around gene body region. TSS: translation start site, TES: translation termination site.

Dysregulated exon usage

Manual inspection found that changes in gene body methylation could be associated with exon skipping/intron retention.

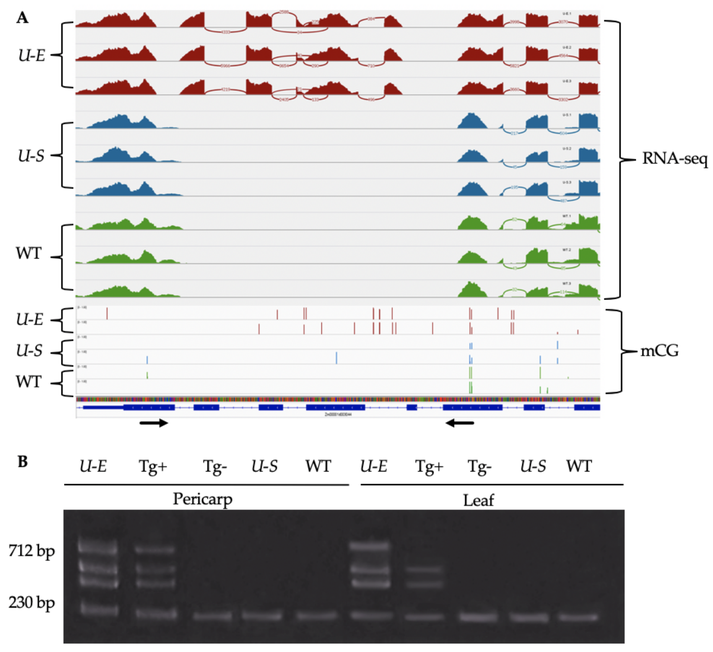

Figure 10: Intron retention of Zm00001d003044 (putative beta-glucosidase) is associated with elevated CG methylation. A) IGV plot displaying RNA-seq exon coverage and CG methylation (mCG), RT-PCR primer positions are indicated in the black arrow. B) RT-PCR validation of intron retention in pericarp and leaf samples.

Meanwhile, we found widespread dysregulation of exon usage in multiple tissues, with nucleotide binding genes being most affected.

GO term	Ontology	Description	Number in input	Number in BG	FDR
GO:0000166	F	nucleotide binding	154	4586	0.019
GO:0017076	F	purine nucleotide binding	128	3780	0.019
GO:0032559	F	adenyl ribonucleotide binding	113	3266	0.019
GO:0032555	F	purine ribonucleotide binding	123	3607	0.019
GO:0032553	F	ribonucleotide binding	123	3607	0.019
GO:0001882	F	nucleoside binding	118	3443	0.019
GO:0001883	F	purine nucleoside binding	118	3434	0.019
GO:0005524	F	ATP binding	110	3105	0.019
GO:0030554	F	adenyl nucleotide binding	118	3432	0.019
GO:0042803	F	protein homodimerization activity	9	82	0.024
GO:0044424	C	intracellular part	295	8943	1.10E-06
GO:0005737	C	cytoplasm	226	6663	1.60E-05
GO:0043227	C	membrane-bounded organelle	230	7016	0.00012
GO:0043229	C	intracellular organelle	247	7690	0.00015

Table 1: Top GO terms enriched in DEU-containing genes. F: molecular function, C: cellular component.

Summary of sequencing data

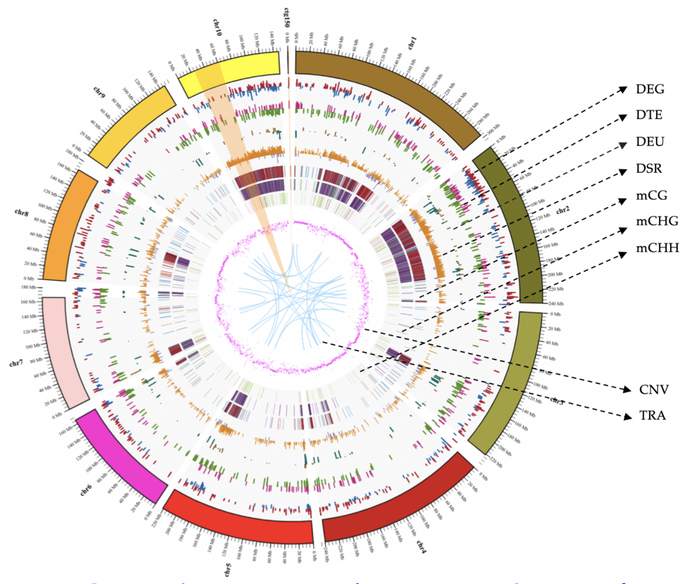

Figure 3: Circos plot summary of sequencing data in *Ufo1-1*. DEG: differential gene expression, DTE: differential transcript expression, DEU: differential exon usage, DSR: differential small RNA (smRNA) expression, mCG: differential CG methylation, mCHG: differential CHG methylation (H=A, C or T), mCHH: differential CHH methylation, CNV: copy number variation, TRA: translocation.

smRNA in TE

Genome-wide smRNA sequencing data shows that *U-S* accumulates higher level of both 24-nt small interfering RNAs (siRNAs) and non 24-nt siRNAs (e.g. 20, 21, 22, 23-nt siRNAs and micro RNAs) in transposable element (TE).

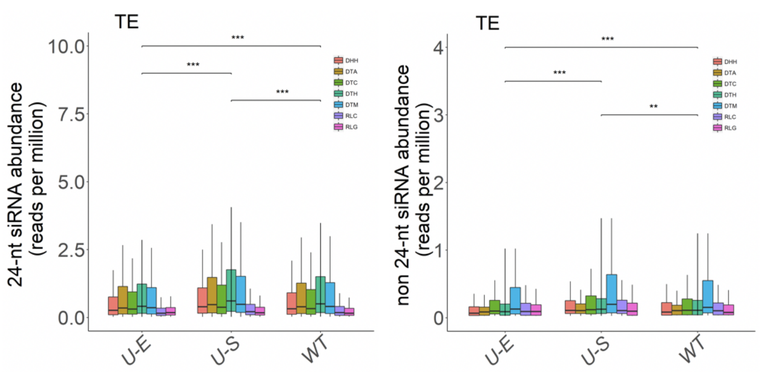

Figure 7: smRNA level in different types of TE. DHH: Helitron, DTA: *hAT*; DTC: *CACTA*; DTH: *PIF/Harbinger*, DTM: *Mutator*, DTT: *Tc1/Mariner*, RLC: *Copia*, RLG: *Gypsy*. *, **, ***, indicates sigificant difference $P < 0.05$, $P < 0.01$ and $P < 0.001$, respectively. NS indicates $P > 0.05$, Mann-Whitney *U* test.

Ongoing experiments

1. GRMZM2G053177 subcellular localization in *Nicotiana benthamiana*.
2. GFP-Trap immunoprecipitation to identify potential binding partners for GRMZM2G053177.