

WHAT ARE THE GENE REGULATORY PROCESSES

REQUIRED TO MAKE A SOYBEAN SEED?

ROBERT B. GOLDBERG (PI) [UCLA]

MATTEO PELLEGRINI (CO-PI) [UCLA] JOHN J. HARADA (CO-PI) [UC DAVIS]



RESULTS INTRODUCTION Soybean Seed Development Major Questions Being Addressed In Soybean Seed Development Maturation & Preparation For Dormancy Dormancy Germination Morphogenesis L Target Genes LAZ Target Genes LAZA Target Genes LA Target Genes mm - Stages Studied - Im pd1 pd2 dry sdlg Photosynthesis What are the targets of seed region- and Photosynthesis COT-- Gibberellin Biosynthesis & Gibberellin Biosynthesis & - Embryo Morphogenesis Únifoliate subregion-specific transcription factors? Leaf Seed Maturation • What processes do seed region- and subregion-LA LAZA LAZ specific transcription factors regulate? mature seedlina maturation • What cis-regulatory motifs do seed region- and LAZ LAZA Fertilization CRMs CRMs Endosperm Development subregion-specific transcription factors interact Free Cellular Absorption Reserve with? Mobilization Seed Protein And Lipid Accumulation Differentiation Decrease in of Embryo Axis And Cotyledon **RNA And Protein** Synthesis

Are seed region- and subregion-specific transcription factor targets and interacting cisCharacterizing Regulatory Network That Operates During the Maturation Stage of Seed Development

Figure 2. Cis-regulatory modules that mediate the LEC1 gene regulatory network. (A) Target genes directly regulated by different combinations of seed transcription factors (TFs). L (LEC1), LA (LEC1 and ABI3), LAZ (LEC1, ABI3, and bZIP67), and LAZA (LEC1, ABI3, BZIP67, and AREB3) bind to genes directing different biological processes as represented by enriched Gene Ontology (GO) categories. Closely clustered TF binding sites define cis-regulatory modules (CRMs). Horizontal bars represent CRMs in the 5' regions of different target genes. (B) The L, LA, LAZ and LAZA CRMs are located in accessible chromatin regions as indicated by the average distribution of ATAC-Seq reads over the CRMs (top **panel**), the read density heat map over representative CRMs in

Cell Division Cell Expansion Loss of Water	
--	--

DNA Endoduplication

Figure 1. Major Events in Soybean Seed Development. glob, hrt, cot, em, mm, lm, pd, and dry refer to globular, heart, cotyledon, early maturation, late maturation, pre-dormancy, and dormant seed stages, respectively. COT and sdlg refer to cotyledon and seedling, respectively.

regulatory modules conserved between different plant species?

• What are the gene networks required to program seed differentiation and maturation?

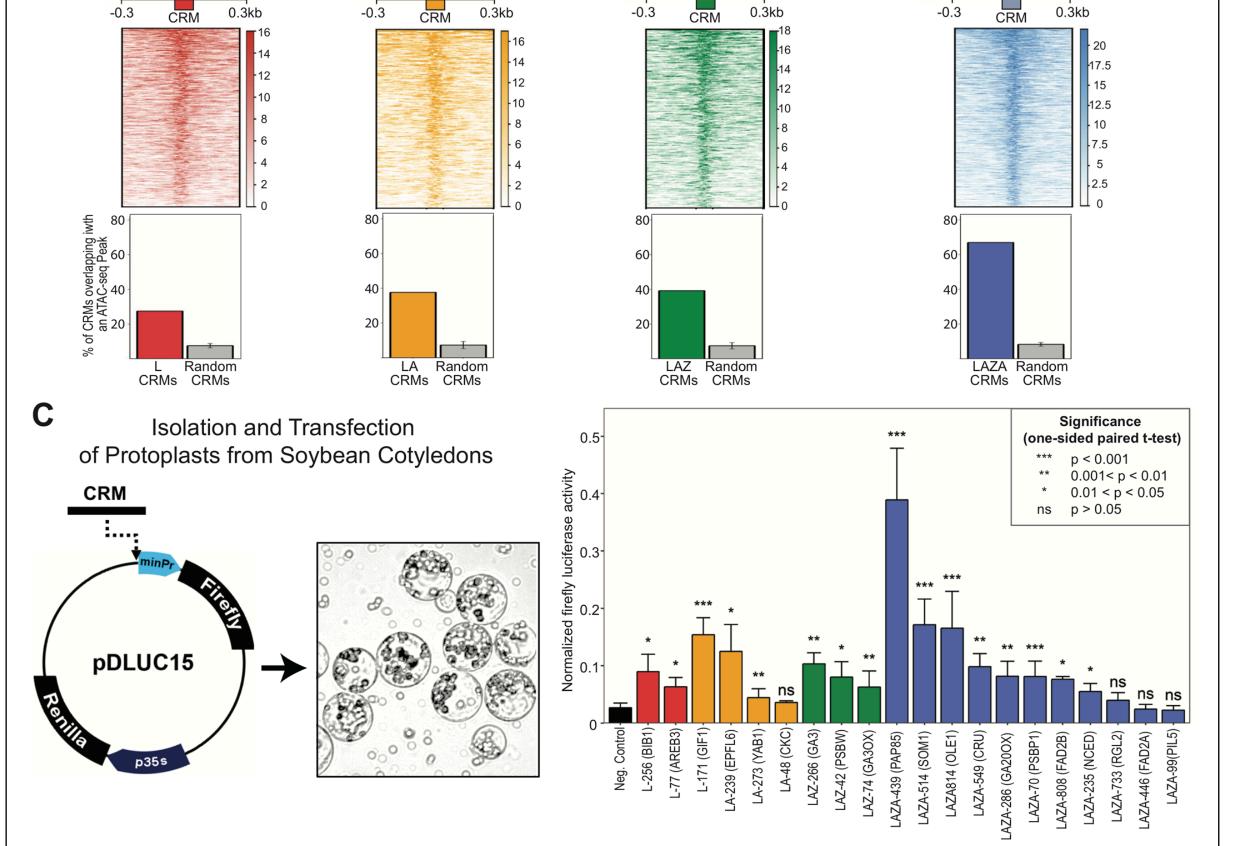
RESEARCH GOALS

Cell Division/ Expansion

• The long-term goal of this project is to obtain insight into the regulatory processes required to control gene activity during seed development that are required to "make a seed". We are investigating soybean seeds because they are: (1) a major source of food and fuel worldwide and, in 2018, it was the largest crop in the United States in area harvested (90M acres) with a value of \$41B, (2) an excellent system for studying the basic processes controlling seed development, and (3) excellent genomic resources are available that can be used to gain insights into the processes required for seed differentiation.

• Specific objectives are to identify the gene regulatory networks that are responsible for controlling the differentiation and function of major seed regions (e.g., embryo) and subregions (e.g., seed coat layers) throughout development. Functional genomic technologies and bioinformatics are being used to identify the downstream target genes and cognate DNA control elements of transcription factors that control the differentiation of major soybean seed regions from fertilization through maturation. Comparative approaches will be used to determine whether orthologous transcription factors in the seeds of other plants target similar genes and cis-control elements.

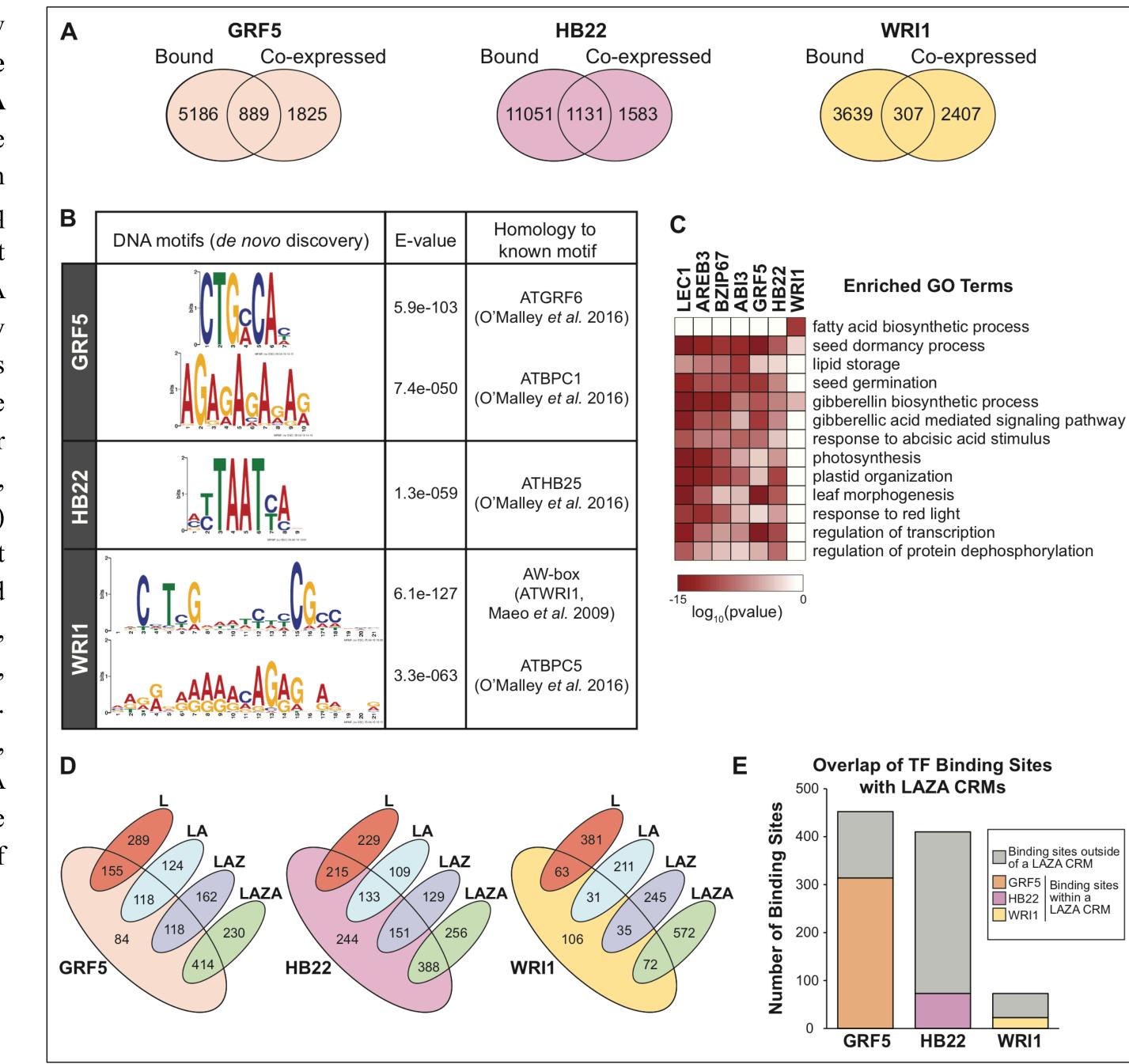
• The significance of this project is that important new information is being obtained about the transcriptional mechanisms controlling gene activity during soybean seed development, which, in the long term, should aid in the improvement of



each group (middle panel), and the percentage of the CRMs in each group that overlap with ATAC-Seq peaks when compared to random sequences of equal size and distribution (bottom panel). (C) Gene constructs used to assess CRM function in transient assays with soybean embryo cotyledon protoplasts (left). Normalized firefly luciferase activity that indicates the ability of L, LA, LAZ, and LAZA CRMs to promote transcription from the CaMV 35S minimal promoter (right). Specific gene (e.g., OLE1 – Oleosin) and 5' upstream region containing the CRM that drives the 35S minimal promoter are

defined below each bar.

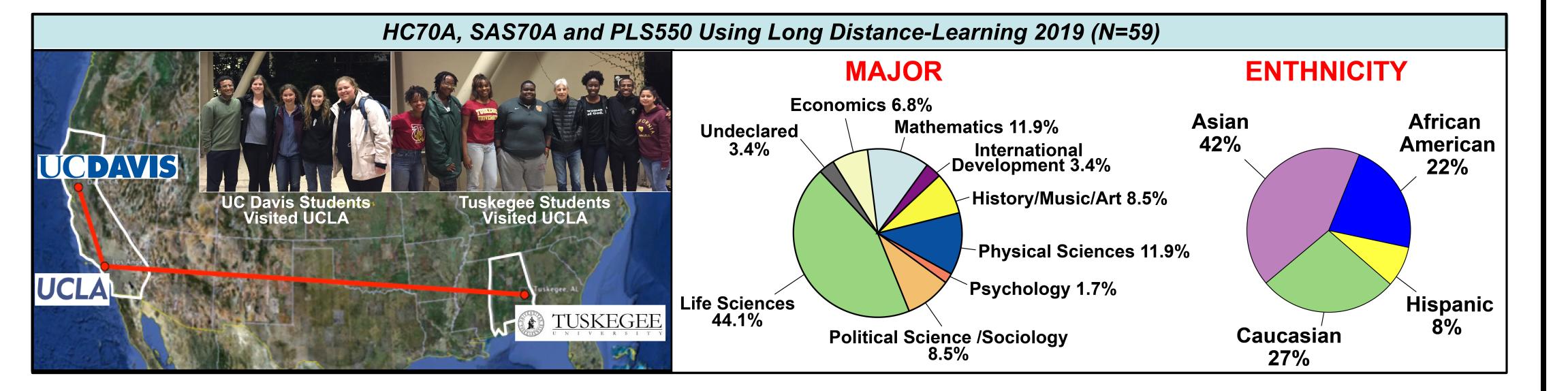
Figure 3. Expanding the LEC1 gene regulatory network by identifying the target genes of three TFs that are directly regulated by the LAZA **TFs.** (A) Target genes directly regulated by the GRF5, HB22 and WRI1 TFs. The intersection between genes bound by the TFs using Chip-Seq and those expressed using RNA-Seq represent candidate genes regulated by the TFs. (B) DNA motifs identified by de novo motif discovery analyses that are overrepresented in DNA regions bound by GRF5, HB22, and WRI1 TFs. (C) Gene Ontology (GO) biological processes enriched for target genes regulated by GRF5, HB22, WRI1, LEC1, AREB3, bZIP67, and ABI3 TFs. (D) Comparison of GRF5, HB22, and WRI1 target genes with those regulated by the L, LA, LAZ, and LAZA TF combinations. (E) Overlap of of GRF5, HB22, and WRI1 DNA binding sites with LEC1, ABI3, bZIP67, and AREB3 (LAZA) CRMs. Orange, purple, and yellow bars represent GR5, HB22, and WRI1 TF binding sites within LAZA CRMs, respectively, whereas grey bars indicate GR5, HB22, and WRI1 TF binding sites outside of LAZA CRMs.



seeds of soybean and other legume crops, using both marker-assisted breeding and genetic engineering approaches.

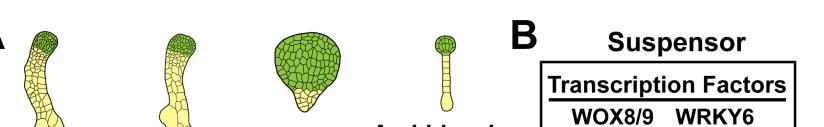
EDUCATION AND OUTREACH

Broader Impacts - Using Long-Distance Learning To Teach Students Simultaneously At UCLA, UC Davis, And Tuskegee University



"Genetic Engineering in Medicine, Agriculture, and the Law" was taught using a novel long-distance teaching approach that linked ~183 UCLA (HC70A), UC Davis (SAS70A) and Tuskegee (PLS550) students over the high speed internet-based video conferencing system into one "inter-campus" classroom. During the quarter UC Davis and Tuskegee students visited UCLA allowing in-person interactions with their UCLA counterparts.

	AGL62 YAB1 WOX9				
Stage	Globular Heart				
	Endosperm	Embryo Proper	Susp	ensor	



PROGRESS

• The cis-regulatory modules (CRMs) of different TF combinations of four soybean transcription factor (TF) genes – LEC1, AREB3, bZIP67, and ABI3 were identified. Using transposase-accessible chromatin (ATAC)-Seq, we confirmed the cisregulatory modules (CRMs) represent regions are accessible to TF binding.

• To expand LEC1 gene regulatory networks, we conducted ChIP-Seq experiment of three downstream seed TFs - GRF5, HB22, and WRI1 that have been demonstrated to play roles in seed development. Comparison of GRF5, H22, and WRI1 target genes with those targeted by L, LA, LAZ, and LAZA TFs shows the greatest overlap with the LAZA target genes.

• Target genes of soybean seed region- and sub-region-specific transcription factors are being identified to expand the soybean seed gene regulatory networks. These include AGL62, YAB1, and WOX9 TF that are specific for the endosperm, embryo, and suspensor, respectively, during the globular and heart stages of development.

• To identify suspensor-specific transcription factors that are conserved between four plant species, we used RNA-Seq and laser capture micro dissection (LCM) to profile the mRNAs present in the giant suspensors of scarlet runner bean (Phaseolus coccineus) and common bean (Phaseolus vulgaris) and compared the mRNAs of two giant bean suspensors to those we obtained from much-smaller and less specialized suspensors of soybean and Arabidopsis. To date, we have uncovered six TF mRNAs that are specific for globular-stage suspensors of scarlet runner bean, common bean, soybean, and Arabidopsis.

Region & Sub-Region Expression Pattern					Phaseolus coccineus ~8
Target Candidate Genes (TFs)	Bound Co-expressed	Bound Co-expressed 8913 (98) 451	Bound Co-expressed	Bound Co-expressed 6930 (334 (33) 1625	Embry C ARF3
Enriched Biological Process GO Terms	 * Regulation of transcription * Nitrogen compound metabolic process 	* Oxidation reduction	* Regulation of transcription	* Regulation of transcription	TMOS CIB4 C2C2
DNA Motifs (<i>de novo</i> discovery)					Figure 5. 7 regulated > and suspe
Homology to Known Motif	ATAGL15 (O'Malley <i>et al</i> . 2016)	ATbZIP911 (Martinez-Garcia <i>et al</i> . 2002)	ATBPC1 (O'Malley <i>et al</i> . 2016)	ATCDF5 (O'Malley <i>et al</i> . 2016)	Phylogenetic and Arabide

Figure 4. Identifying target genes of soybean seed region- and sub-regionspecific transcription factors during early seed development. Numbers in red indicate target candidates for each TF.

	seolus vu	seolus Igaris	Glycine max	Arabid thalia	opsis ana	HDG11 ARF16	C2C2 TF ANAC9
C	cineus					Transpor	t Related
	~8 MYA		YA ~12 uspensor	20 MYA		YSL3 YSL7 AAP6	INT2 PHT4;2
			Em	bryo Pro	oper		
Transcription Factors			Hormon	e Related			
	ARF3/ETT TMO5L1 CIB4 C2C2 TF	PAN LRL1 SAP HAN	OBP1 YAB5 SHI ZFP4	KAN1 KAN2 LRP1 HAT14	GRF4 GRF5 CUC2 CUC3	PIN1 AUX1 NPY1	LOG7 YUC4 ACL5

Transcription factors that are conserved and up->five-fold in the globular-stage embryo proper ensor of four different plant species. (A) tic tree of beans with giant suspensors and soybean dopsis with small suspensors. Green and yellow colors indicate the embryo proper and suspensor, respectively. Embryo cartoons are not drawn to scale. > Five-fold upregulated genes in the suspensor (B) and embryo proper (C) of all four plant species: scarlet runner bean (Phaseolus coccineus), common bean (Phaseolus vulgaris), soybean (*Glycine max*) and *Arabidopsis (Arabidopsis thaliana*).