MCDB 201L: Molecular Biology Laboratory Professor Maria Moreno Teaching Fellow: Christopher Bartley

By submitting this essay, I attest that it is my own work, completed in accordance with University regulations. -Kevin Hochstrasser

Visualization of localization and expression of Arabidopsis thaliana gene AT1G52340, an ortholog of Tasselseed2

May 8, 2012

Kevin Hochstrasser1, Maria Moreno2, Christopher Bartley3, Rebecca Delventhal2, Shannon Stewart2 & Skye Zeller4

1Yale College 2Department of Molecular, Cellular and Developmental Biology 3MSTP Program, Department of Neurobiology 4Department of Microbiology

Abstract

The goal of our research was to create a synthetic gene construct and insert it into the *Arabidopsis thaliana* genome to visualize the localization and expression of the *AT1G52340* gene. *AT1G52340* is a known ortholog of a gene in maize (*GRMZM2G332976*) that is known to function in sex determination. By studying the *AT1G52340* gene in *Arabidopsis*, we hoped to understand how maize regulates the cell death-signaling pathway in sex determination. The reporter construct was assembled from the promoter and terminator regions of the *AT1G52340* gene and a reporter gene called GUS (beta glucuronidase) which, when expressed in the presence of its substrate, will create a blue color. Though attempts to assemble, amplify, and transform the synthetic construct into *E. coli* cells seemed successful based on diagnostic gels of the transformed bacteria, sequencing results showed that what was transformed was not the expected synthetic gene construct. More work must be done in order to amplify and clone the synthetic gene before it can be infected into *Arabidopsis* and studied.

Introduction

The impending food crisis will make food unavailable to vast amounts of the world's population unless crop yields can be markedly increased in the coming century. Maize hybrids and the hybrids of similar plant species demonstrate a trait, known as heterosis, defined as the increased performance of a hybrid relative to its parents, that if harnessed could help combat this crisis (Springer and Stupar, 2007). Maize hybrids tend to produce stronger plants, larger seed-bearing ears, and higher crop yields. How and why hybridization of plant species leads to increased vigor in offspring is not fully understood, but because heterosis in crops leads to significant improvements in agricultural yields, it would be advantageous if selective breeding could produce heterosis in other staple crops like rice, wheat, and barely. These organisms have perfect flowers (both sex organs expressed within the same flower). Only plants with imperfect flowers can be crossed (or hybridized) because they do not exhibit self-fertilization. In this experiment, genes in Arabidopsis known to be orthogonal to known sex determination genes in maize will be investigated to see if the orthologs might have a similar function in a plant with perfect flowers. Investigations of these orthologs may one day allow for the engineering of crops that express perfect flowering to express imperfect flowering and heterosis.

Maize has separated, unisexual flowers. The *tasselseed* and *silkless1* family of genes largely controls sex determination in maize (Wu X, Knapp S, Stamp A, 2007). *Tasselseed* genes are epistatic to *silkless1* and are responsible for pistil cell death (Wu X, Knapp S, Stamp A, 2007). If the *silkless1* gene is not expressed, tasselseed 2 will cause the death of pistil cells and the division of stamen cells. Silkless1 is responsible for stamen cell arrest (Calderon-Urrea A, Dellaporta SL, 1999). If *silkless1* is expressed, it inhibits cell death

signaling of the pistil cells and the stamen cell cycle is arrested. The biochemical mechanisms behind these observed behaviors are not yet fully understood. Jasmonic acid signals cell death, but the mechanisms behind this process and jasmonic acid production in cells has yet to be described (Acosta, Laparra, Romero, 2009). By replacing the orthologs of *silkless1* and *tasselseed2* in *Arabidopsis* with a reporter gene that can be visualized, it is possible to observe where the orthologs are expressed in these plants and possibly even phenotypic variances resulting from the replacement. Research into the expression and localization of these proteins could uncover more information about how the closely related sequences in maize regulate the cell death-signaling pathway.

This paper describes the cloning and characterization of gene *AT1G52340* in *Arabidopsis*. Previous research has found that overexpression of *AT1G52340* in *Arabidopsis* leads to larger plant size (Lin, Hwang, Endo, 2007) and may also be involved in the biosynthesis of other enzymes as a part of a more complex metabolic pathway (Xie, Williamson, Holroyd, 2006). In order to determine the location and temporal expression of the *AT1G52340* gene product, a reporter construct was created from the promoter and terminator regions *AT1G52340* and a reporter gene (uidA, GUS) coding for an enzyme which, when expressed in the presence of its substrate, will create a blue color. The experiments conducted in this paper produced a successful transformation of the constructed synthetic gene described into *E. coli* cells, but subsequent sequencing results showed that the transformed synthetic gene did not have the desired sequence.

Results

Genomic DNA was isolated from *Arabidopsis thaliana* and then used as template DNA for PCR amplifications of the promoter and terminator of the gene *AT1G52340*. The

AT1G52340, gene was chosen for study because of its homolog in maize (*Zea mays*) *GRMZM2G332976*. Figure 1 shows the protein domains of the AT1G52340 protein. Many of these domains are also present in the GRMZM2G332976 protein, including the DH_sc/Rdtase_SDR, DHB_DH, and Glc/ribitol_DH of the Prints domain, the 51735 Superfamily domain, as well as the DH_sc/Rdtase_SDR of the Pfam domain. Both the AT1G52340 and GRMZM2G332976 proteins are members of the superfamily SDR, recognized to function in hormonal biosynthesis and sex determination in maize and *Arabidopsis* (Cheng, 2002). These commonalities, as well as the fact that both genes contain two identically proportioned exons, make the *AT1G52340* gene a likely homolog of the sex determination gene *GRMZM2G332976*, making it ideal for study.

The TEV GUS 35sa (Tobacco etch viral plasmid, β -Glucuronidase gene) was used as a template in order to extract the fluorescent reporter uidA gene (gene encoding the GUS enzyme) by PCR amplification of the gene segment (figure 2). The SLIC ends of around 20 base pairs each that flanked the amplified reporter would later function in the ligation independent cloning (LIC) reaction detailed in figure 4A.

The amplified uidA reporter and the promoter and terminator regions of the *AT1G52340* gene were gel extracted from a preparatory agarose gel electrophoresis (AGE) and run on a diagnostic AGE to check for accurate and high-yield amplification of the fragments. Three clear banding products of around 1000 base pairs, around 2000 base pairs, and around 500 base pairs were observed when compared against a 1 kb ladder that was run in the first lane (figure 3).

The fragments were combined in an equimolar mix and run through the SLIC cloning procedure in order to assemble the synthetic gene with 5' and 3' LIC sites (figure

4A). The assembled gene was then inserted into the tDNA vector

(pYU2588_pPZP200b_T-DNA_FRT35SBarTr7_LIC) to a form a recombinant plasmid with a negative selection (sacB absence, ability to digest sucrose) and a positive selection (spectinomycin resistance, spectinomycin survival) in order to confirm the transfer of the plasmid DNA after PCR amplification of the colonies (figure 4B-D).

Promoter and terminator fragments were also inserted into the plasmid vectors containing a positive selection mechanism of resistance to kanamycin (not shown) in order to later sequence the regions.

Bacterial (*E. coli*) colonies containing the inserted synthetic gene as well as those containing the inserted promoter and terminator genes grew prolifically indicating an efficient transfer and uptake of the plasmid, which was necessary to the survival of the cells. Figure 5 shows PCR reactions carried out on the bacterial colonies carrying the promoter, terminator, and full-length synthetic gene inserts. Promoter banding is around 1000 base pairs, terminator banding is around 500 base pairs, and full length synthetic gene banding is around 3500 base pairs.

Sequencing data (not shown) was obtained for the promoter, terminator, and fulllength synthetic gene plasmids. The terminator region showed a perfectly matched sequence from both directions of sequencing, and 100% fidelity with the expected sequence product. No discrepancies were observed. The promoter region and full-length synthetic gene sequencing data showed matches of 60% and 57% respectively to their expected sequence products, below the threshold fidelity for construction of readable chromatographs.



С

MSTNTESSSYSSLPSQRLLGKVALITGGATGIGESIVRLFHKHGAKVCIVDLQDDLGGEV CKSLLRGESKETAFFIHGDVRVEDDISNAVDFAVKNFGTLDILINNAGLCGAPCPDIRNY SLSEFEMTFDVNVKGAFLSMKHAARVMIPEKKGSIVSLCSVGGVVGGVGPHSYVGSKHAV LGLTRSVAAELGQHGIRVNCVSPYAVATKLALAHLPEEERTEDAFVGFRNFAAANANLKG VELTVDDVANAVLFLASDDSRYISGDNLMIDGGFTCTNHSFKVFR

Figure 1. Gene AT1G52340 structure, domains, and protein sequence A.)

Structure of the *AT1G52340* gene B.) Structure of cDNA with motifs and domains labeled C.) Protein amino acid sequence for AT1G52340



Figure 2. Structure and Reporter Gene PCR Isolate A.) TEV_GUS_35Sa plasmid including GUS gene used as a reporter with forward and reverse PCR primers annotated as green and red arrows B.) Expected PCR product with 5' and 3' SLIC overhangs indicated



Figure 3. AGE banding photograph of synthetic gene components AGE of the gel extracted promoter, reporter, and terminator sequences. 1 kb ladder is annotated.



Figure 4. Synthetic Gene and Plasmid Assembly A.) Sue assembly procedure of the promoter, reporter, and terminator extracted PeR fragments. B.) The full length assembly gene after Sue assembly, consisting of promoter, reporter, and terminator sequences and Ue sites added by PeR C.) tDNA vector including, sacS gene with promoter and terminator. Bsal restriction sites, and kanamycin resistance gene with promoter and terminator. D.) Assembled recombinant SLIC plasmid vector. Contains GUS reporter gene and kanamycin resistance gene.



Promoter 2036 1636 1018 506

В





С

Full-Length Synthetic Gene





Discussion

The goal of our research was to create a synthetic gene construct and insert it into the *Arabidopsis thaliana* genome. The synthetic gene consisted of the *uidA* (GUS) reporter gene placed between the promoter and terminator regions of *AT1G52340*. The promoter and terminator regions were amplified via PCR then checked with agarose gel electrophoresis (AGE) to see if the amplified products were of the expected sizes. Figure 3 shows that the amplified segments were of the correct lengths, and the presence of single bands indicates that DNA was amplified at a single length. These lengths of about 1000, about 500, and about 2000 base pairs for the promoter region, terminator region, and reporter gene respectively were the expected DNA segment lengths.

Initially, synthetic gene assembly and subsequent transformation of the GUS SLIC complex (which included the promoter, uidA, and terminator regions) was not successful, and no bacterial colonies resulted. Due to this setback, the promoter and terminator regions were cloned individually by blunt-TOPO cloning. This transformation was successful (figure 6) and the plasmids were sent to be sequenced to see if they had been amplified with 100% fidelity.

After modifying the LIC reaction procedure (lengthening extension time) and PCR amplification procedure (using the HF high fidelity method without DMSO), the assembled full-length synthetic gene was successfully amplified and transformed into the SLIC vector. The results of the PCR, shown in figure 6, revealed that three colonies showed bands of around 3500 base pairs. This was the expected length of the full-length gene. This result offers strong evidence that the full-length synthetic gene had been assembled, amplified, and transformed correctly. In order to verify this, the plasmid was isolated and sent for sequencing.

Sequencing data from the promoter, terminator, and full–length synthetic gene shows that only the terminator was successfully cloned and transformed. It was a 100% match in both the forward and reverse directions (p264, p265) and with the expected sequence. The promoter and full-length synthetic gene were not amplified or cloned with high fidelity, as they were both less than 60% matches to the expected gene products and were below the threshold to create chromatogram comparison graphs.

Sequencing yielded unexpected results. This was most likely caused by incorrect amplification of the promoter sequence since the terminator sequence was a 100% match. Primers may not have bound correctly to the gDNA, binding at unexpected locations. Proceeding forward, the next step that must be taken will be to redesign the PCR primers for the promoter region and make it more specific perhaps by expanding it in length. Once the full-length synthetic gene is transformed successfully, we will next attempt to create more transformed bacterial colonies using the same method and then attempt to infect *Arabidopsis thaliana* samples with a bacterial carrier capable of insertion of the full-length synthetic gene.

Insertion of the reporter gene (*uidA*) into *Arabidopsis thaliana* genome will allow us to visualize the localization and expression patterns of the *AT1G52340* gene. The *AT1G52340* gene was chosen for this study because it is an ortholog of the *GRMZM2G332976* gene in *Zea mays*. The two genes share many of the same domains and motifs, including the *superfamily* domain, *prints* domains, and a *pfam* domain. They are similar to one another in structure as well, containing two exons each with a single intron. The *GRM7Jv12G332976* gene has been determined to have a role in the *tasselseed-2* sex determination pathway (Cheng, 2002) New information regarding the role of this gene in the sex determination pathway in *Arabidopsis thaliana* may lead to discoveries allowing us to induce imperfect flowering similar to that of

Zea mays in plants that naturally express perfect flowers. This feat could increase the competence of unisexual crops (heterosis) and lead to higher crop yields.

Materials and Methods

Genomic DNA Extraction

Genomic DNA (gDNA) was isolated from a single leaf collected from *Arabidopsis thaliana* Col-O wild type. gDNA was extracted using the QIAgen DNeasy Plant Minikit following the manufacturer's protocol of Purification of total DNA from plant tissue, with the following modification: A single leaf was ground by mortar and pestle under continuous application of liquid nitrogen to inhibit DNase activity.

Yield and purity (260/280) of the DNeasy extraction was quantified using the NanoVue plus spectrophotometer. Agarose gel electrophoresis (AGE) was used to verify that the products were present in high concentration and purity. Suitability of isolated gDNA for PCR was verified by PCR amplification of the control gene (*AT5G60390*).

Gel Electrophoresis

Agarose gel electrophoresis was used to verify successful PCR amplification and for DNA separation prior to gel extraction. Samples were run through 0.8% SYBR green dye (Invitrogen) agarose gels (USB Corporation). Each sample consisted of 25µL of solution comprised of 1X RapidRun loading dye (Affymetrix), approximately 100 ng of DNA sample and distilled water. Gels were visualized under UV light adjacent to a 1 kb ladder (Invitrogen).

PCR Amplification

The control gene (*AT5G60390*), the promoter (*AT1G52340*), the terminator (*AT1G52340*), the reporter gene (*uidA*), and the full-length synthetic gene were amplified by

PCR. For each sequence of interest, four distinct PCR reactions with a final volume of 25µL containing 1x Phusion Master Mix (Finnzymes) (with either HF Buffer or GC Buffer), 0.52 µM forward primer, 0.52 µM reverse primer, 70-100 ng of template DNA with or without 2% dimethyl sulfoxide were amplified. PCR reactions were run using general cycling methods with slight modifications for each individual sample: Initial denaturation for 30 seconds at 98°C followed by 30 cycles of denaturation for 10 seconds at 98°C, annealing for 30 seconds at 3°C higher than the lowest Tm of the two annealing primers, extension lasted 1 kb per 30 seconds at 72°C, and final extension for 5 minutes

at 72°C. The samples were stored at 4°C.

A control PCR reaction was carried out for gene *AT5G60390* from purified gDNA to verify that the gDNA was suitable for PCR. The primers used were MP36 (5'-GTGAGCACGCTCTTCTTGCT-3') and MP37 (5'-TCCTTGACAGCAACATTATT-3'). The annealing temperature was 55°C and the extension time was 30 seconds. The expected size of the amplified gene (2361 base pairs) was verified by AGE.

A PCR reaction was carried out to isolate and amplify the reporter gene (β -Glucuronidase or GUS) from the plasmid pRTL2-GUS using the primers MB12RF12 (5'-GATCTAAGGAGCTAAAGCTATTatggtccgtcctgtagaaac - 3') and MB12RR12 (3'-ACATTCTTTAGCAAAATGCAtcattgtttgcctccctgct - 5'). Capitalized letters are homologous to the targeted gene, while lowercase letters are extensions (SLIC ends) that are homologous to the promoter and terminator of *AT1G52340*. The annealing temperature was 64°C and the extension time 1 minute.

PCR reactions were carried out for the promoter and terminator regions of the gene being studied, *AT1G52340*. The promoter region of the gene was amplified with forward primer MB12PF12 (5'- tacttccaatccatg ATACTCCGGTGGCAGAAGTGACGG - 3') and reverse primer MB12PR12 (3'-

gtttetacaggacggaccatAATAGCTTTAGCTCCTTAGATC-5'). The annealing temperature for the promoter sequence was 58°C and extension time was 1 minute. Amplicon size was expected to be 1141 base pairs for the promoter and 529 for the terminator. The terminator region of *AT1G52340*was amplified with forward primer MB12TF12 (5'agcagggaggcaaacaatgaTGCATTTTGCTAAAGAATGT- 3') and reverse primer MB12TR12 (3'- tatccacctttactg ATTTGACACCGTTACGGAGACTAACG- 5'). The annealing temperature was 61°C and extension time was 1 minute. The lowercase letters at the 5' end of the forward promoter primer and the 3' end of the reverse terminator primer correspond to LIC sequences homologous to those of the vector it was intended to be inserted into. Lowercase letters of the amplified promoter, terminator, and reporter gene sequences were SLIC ends, designed to be homologous to one another for SLIC assembly.

The full-length synthetic gene was amplified with the forward primer MB12PF12 (5'tacttccaatccatg ATACTCCGGTGGCAGAAGTGACGG - 3') and the reverse primer MB12TR12 (3'- tatccacctttactg ATTTGACACCGTTACGGAGACTAACG- 5'). HF PCR conditions without the addition of DMSO and "Touchdown" cycling conditions were employed. Touchdown conditions included two cycles of denaturation, annealing, and extension in addition to initial denaturation (98°C, 30 seconds) and final extension (72°C, 5 minutes). The annealing temperatures were 70°C and 69°C respectively and lasted 30 seconds each. The extension times were 4 minutes each at 72°C. The expected size of the amplicon was 3482 base pairs.

In order to screen the pCR Blunt II-TOPO recombinant plasmids with promoter and

terminator sequences inserted, forward primer p264 (5'-

GCTATGACCATGATTACGCCAAG-3') and reverse primer p265 (5'-

GACGTTGTAAAACGACGGCCAG-3') were used. Cycling conditions were modified: Initial denaturation occurred at 95°C for 8 minutes then at 98°C for 2 minutes, denaturation and annealing lasted 15 seconds, annealing temperature was 60°C and extension time was 1 minute for both the promoter and terminator. The expected size of the amplicon for the promoter was again 1141 base pairs and again 529 base pairs for the terminator. Diagnostic AGE was used to determine that the PCR amplification was successful, and the NanoVue plus spectrophotometer was used to measure concentration and purity (260/280).

Gel Extraction and Purification

DNA fragments were excised from a preparative gel and purified using the QIAquick gel extraction kit protocol. Preparative gel samples for the promoter, terminator, reporter and full-length synthetic gene fragments consisted of 50μ L of solution comprised of 1X RapidRun loading dye (Affymetrix), 40μ L of combined PCR product and distilled water before being excised. Two elutions were collected after purification as backup samples. Concentration and purity of extracted products was quantified using the NanoVue plus spectrophotometer.

Enzymatic Assembly of Synthetic Gene

The promoter, terminator, and reporter gene sequences were combined in a 5μ L equimolar mix calculated from the length of each segment and the concentrations obtained from the NanoVue Plus spectrophotometer. This was added to 15μ L of SLIC reaction mix mixed on ice and incubated for 50°C for one hour in a BioRad MyCycler thermocycler. The

assembled gene was then amplified by PCR.

Ligation Independent Cloning of Synthetic Gene

A 100µL sample was prepared consisting of 1x NEB4 buffer and Bovine Serum Albumin (NEB), 5µg of plasmid vector (pYU2588), 3µL BsaI restriction enzyme (NEB) and distilled water. The mixture was incubated for 3 hours at 50°C followed by 65°C for 15 minutes in order to inactivate the BsaI enzyme. The plasmid vector was purified using the QIAquick PCR purification procedure (gel extraction).

 50μ L of the digested BsaI vector was combined with 1X T4 Buffer (NEB 2), 2.5mM dGTP, 1X BSA (NEB), 5 mM DTT, 5μ L of T4 DNA polymerase, and distilled water, then incubated for 30 minutes at 22°C. It was then incubated at 70°C for 20 minutes using the BioRad MyCycler thermocycler to heat inactivate the T4 polymerase and stored at - 20° C.

 $5.0 \ \mu$ L of the extracted and purified PCR insert of the full-length synthetic gene was then combined with 1X T4 buffer (NEB 2), 2.5 mM dCTP, 50 mM DTT, 5X BSA (NEB), .5 μ L of T4 DNA polymerase, and distilled water and incubated for 22°C for 30 minutes in a BioRad MyCycler thermocycler. It was then incubated at 70°C for 20 minutes to inactivate the T4 polymerase.

 1μ L BsaI-T4 treated vector and 2μ L of T4 treated PCR insert were combined and incubated for 10 minutes at 22° C and then placed on ice.

Transformation of the cells was then carried out by adding 50μ L of competent cells (Invitrogen TOP 10) and incubated on ice for 15 minutes. The cells were then heat shocked in a water bath for 42°C for 30 seconds and immediately placed on ice for 2 minutes. 100µL of SOC medium (Invitrogen) was added and the cells were shaken for 1 hour at 37°C, after

which the cells were ready for plating onto LB medium with 5% sucrose and spectinomycin (see section on plating).

TOPO Cloning

TOPO cloning was used to transform both the purified promoter and terminator fragments. 4μ L of purified product was combined with 1μ L of salt solution and 1μ L of pCR II-Blunt-TOPO solution (Invitrogen) and incubated for 15 minutes at room temperature. 3μ L of the solution was added to a vial of One Shot Chemically Competent *E. coli* (Invitrogen). Aside from this modification, the TOPO Cloning reaction protocol (Invitrogen) was then followed for the transformation of cells.

The cells were plated on LB medium with kanamycin (see section on plating).

<u>Plating</u>

Transformed E. coli were plated on LB (liquid broth) plates containing appropriate antibiotic (kanamycin for TOPO vector, spectinomycin for pYU2588) and sucrose conditions described in sections on transformation. Two plates of bacteria were grown for each cloning procedure by spreading 150μ L on one plate and 100μ L on the second plate using glass beads.

Isolation of Recombinant Plasmid

Recombinant plasmids containing the blunt-TOPO cloned promoter and terminator and the plasmid containing the full-length synthetic gene were isolated in order to be sequenced. The QIAGEN plasmid midi kit and protocol were employed for each isolation.

Sequence Analysis

Primers were annotated using the software Lasergene. The Gene sequence information

was obtained from GenBank. DNA sequence information was obtained from the W.M. Keck

DNA sequencing facility at Yale University. Isolated DNA plasmids (promoter and terminator

TOPO vector plasmids and full-length synthetic gene pYU2588 vector plasmid) were

combined with 4µM primer stock (p264 or p265 for the promoter and terminator, p3640 or

p3639 for full-length synthetic gene) and distilled water and packaged for transport.

References

- Acosta IF, Laparra H, Romero SP, Schmelz E, Hamberg M, et al. tasselseed1 is a lipoxygenase affecting jasmonic acid signaling in sex determination of maize. *Science*. 2009;323:262–265. doi: 10.1126/science.1164645.
- Calderon-Urrea A, Dellaporta SL. Cell death and cell protection genes determine the fate of pistils in maize. *Development*. 1999;**126**:435–441.
- Cheng WH, Endo A, Zhou L, Penney J, Chen HC, Arroyo A, León P, Nambara E, Asami T, Seo M, et al. (2002) A unique short-chain dehydrogenase/reductase in *Arabidopsis* glucose signaling and abscisic acid biosynthesis and functions. Plant Cell **14**: 2723–2743.
- Lin PC, Hwang SG, Endo A, Okamoto M, Koshiba T, et al. Ectopic expression of ABSCISIC ACID 2/GLUCOSE INSENSITIVE 1 in Arabidopsis promotes seed dormancy and stress tolerance. *Plant Physiol*. 2007;143:745–758.
- Springer NM, Stupar RM. (2007) Allele-specific expression patterns reveal biases and embryo-specific parent-of-origin effects in hybrid maize. Plant Cell **19**: 2391–2402.
- Wu X, Knapp S, Stamp A, Stammers DK, Jornvall H, Dellaporta SL, Oppermann U (2007) Biochemical characterization of TASSELSEED 2, an essential plant short- chain dehydrogenase/reductase with broad spectrum activities. FEBS J 274 1172–1182.
- Xie X, Wang Y, Williamson L, Holroyd GH, Tagliavia C, et al. The identification of genes involved in the stomatal response to reduced atmospheric relative humidity. *Curr Biol.* 2006;16:882–887.