

# *AHL9, AHL11, and AHL12: gene overexpression and phenotypic analysis in Arabidopsis*

Paul S. Froese

Spring 2012

Advised by:

Dr. Michael M. Neff

Department of Crop and Soil Sciences

College of Agricultural, Human, and Natural Resource Sciences

## Précis

Genes that influence how plants adjust growth and development in response to changes in light quality are key points of interest with respect to potential improvement of crop plant traits. The *AHL* plant gene family, which comprises 29 distinct genes in the genome of the model plant species *Arabidopsis*, has been implicated in plant growth and developmental responses to light. Each gene in this family codes for similar protein structural domains, suggesting that they may all have some function in common; yet it would be illogical to propose that a plant has 29 genes that all do the same thing. Therefore the multiplicity of these *AHL* genes strongly intimates that each has some unique purpose. Previous experiments with genes in this family have indeed demonstrated both overlapping and distinct functions of the *AHL* genes studied. For example, both *AHL27* and *AHL29* are responsible for suppressing growth of seedling stems (hypocotyls) in *Arabidopsis*. Abolishing the function of different *AHL* genes in *Arabidopsis* has produced different plant sizes, indicating that not all *AHL* genes act in the same way or with the same intensity.

With this in mind, we decided to study two previously uncharacterized and highly similar *AHL* genes in *Arabidopsis*, *AHL9* and *AHL11*, in order to determine how much their genetic similarities and differences translate into differences in plant growth and development, and to gain insight into how we might exploit these genes for crop improvement. To study these genes, we separately transformed *Arabidopsis* plants with a highly promoted copy of *AHL9* and *AHL11* genes, generating plants overexpressing these genes. Because of time limitations, we were unable to take measurements from *AHL9* overexpressing plants, but were able to use *AHL12* overexpressors to compare with those for *AHL11*. Plants overexpressing genes *AHL11* and

*AHL12* were studied by measuring their hypocotyl lengths and by making qualitative visual observations of adults.

We noticed changes in plant growth and development of *AHL11* and *AHL12* overexpression lines when compared to wild type plants. We can link these changes to overexpression of the respective genes, since we also observed higher expression of *AHL11* and *AHL12* at the molecular level. Seedlings overexpressing either gene were shorter than wild type plants, and adult overexpressing plants were smaller and weaker than their wild type peers. These observations imply that *AHL11* and *AHL12* may control plant stature and vigor. Although the observable characteristic, or phenotype, of a short hypocotyl in plants overexpressing *AHL11* and *AHL12* echoes that seen for plants overexpressing *AHL27* and *AHL29*, the dwarf adult phenotype of *AHL11* and *AHL12* overexpressors is opposite that observed for *AHL27* and *AHL29* overexpressors. Thus we conclude that *AHL11* and *AHL12* may have functions both redundant and non-redundant with *AHL27* and *AHL29* in regulating plant growth in response to light.

Future directions for this project will entail characterization of *AHL9*, as well as continued study of *AHL11* and *AHL12* via *elimination* of gene function (as opposed to gene *overexpression*). This step will allow for greater accuracy in determining what these genes really do, since gene function is more robustly determined when the gene is absent as opposed to when it is in overabundance. Not only should we continue studying *AHL9*, *AHL11*, and *AHL12*, but we should also pursue knowledge of all other genes in the *AHL* family to gain a complete picture of this family's functionality. Finally, any present or future results amenable to crop improvement should be applied to that end.

## **I. Introduction**

Humans use plants for food, fuel, fiber, medicine, and building materials. As human population increases and arable land area remains constant, a pressing need for more productive and sustainable agriculture arises (Gee, 2002). A basic understanding of how plants grow and develop helps immensely in the effort to improve crops. Researching plant genes which have direct implications for increased agricultural productivity in factors such as yield, food nutrition, and environmental impact, improves the potential for future enhancement of these important traits and should thus be a priority in the discipline of crop genetics (Kishmore and Shewmaker, 1999).

Crop plants have extremely sensitive signal receptor and internal communication systems to allow for swift adaptation to changes in their surroundings. Light quality is a fundamental environmental signal that plants sense and to which they developmentally respond; for example, a plant's photoreceptors detect neighboring individuals, and use perceived light quality status to activate a shade avoidance mechanism, growing above or away from nearby competition (Smith and Whitelam, 1997). The genetic factors that contribute to the many ways in which plants respond to light quality – such as adjustment of total biomass, modification of organ size and quantity, and increase or decrease in chlorophyll concentration – are key points of interest with respect to potential improvement of these relevant crop plant traits. With this broad goal in mind, the Neff lab at WSU studies how light affects seedling growth and development. Genes of interest are characterized in the model species *Arabidopsis thaliana*, and further studied and applied in the crops camelina and wheat.

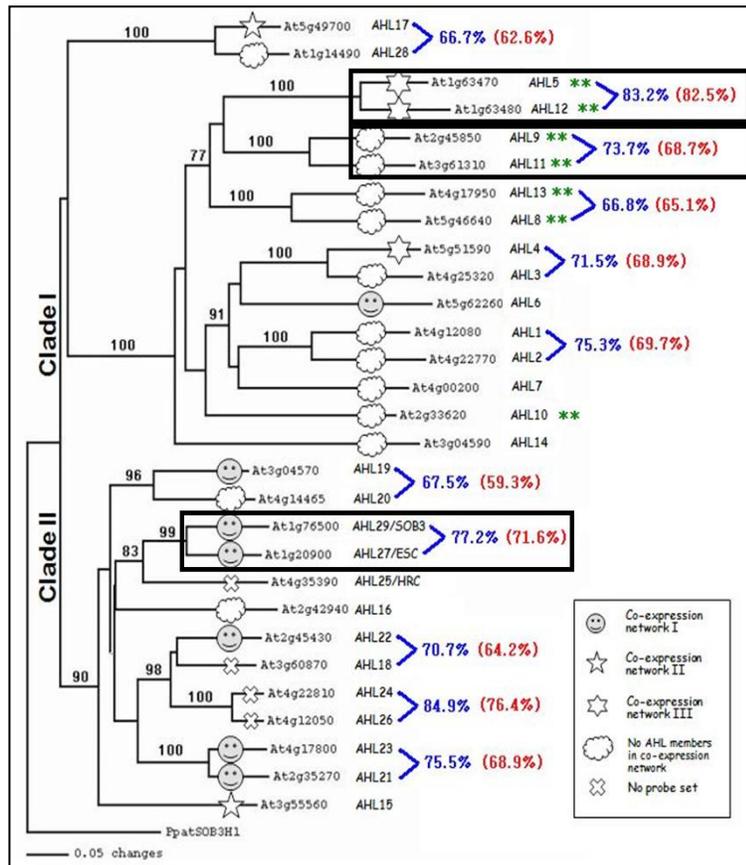
Genes involved in plant photomorphogenesis may have redundant or semi-redundant functions (Smith and Whitelam, 1997). Functionally redundant genes are difficult to identify by

conventional single-gene null-mutations, since abolishing the function of one redundant gene does not produce an aberrant phenotype; therefore, one scientific approach is to first demonstrate the function of redundant genes through gain-of-function analysis before conducting more informative and reliable loss-of-function studies (Weigel *et al.*, 2000). Activation tagging, a gain-of-function mutagenesis tool developed by Weigel *et al.* (2000), is used to uncover functionally repetitive genes by randomly inserting DNA enhancer elements into an organism's genome, with the hope of amplifying the native expression pattern of a gene of interest. Phenotypic differences observed in activation tagged organisms are caused by overexpression of a gene near the inserted enhancer elements. The resulting plant phenotype suggests possible functions of the gene with enhanced expression (Weigel *et al.*, 2000). If, for example, an activation-tagged *Arabidopsis* plant grows extra hairs on its leaves, the enhancer element may have landed in the genome next to a gene that controls leaf hair production.

Phytochrome B (PHYB) is a photoreceptor involved in the shade avoidance phenomena in plants (Smith and Whitelam, 1997). *phyB* mutants grow as if constantly sensing shade, producing taller seedling stems (hypocotyls) than wild type seedlings raised under equivalent light conditions in an attempt to outpace whatever they perceive to be causing the nonexistent shade (Street *et al.*, 2008). Street *et al.* (2008), using activation tagging to find genes that suppress the long hypocotyl phenotype of *phyB* *Arabidopsis* mutants, identified a gene they called SUPPRESSOR OF THE PHYB-4 LONG HYPOCOYTL #3 (*SOB3*), also known as *AHL29*, which can restore *phyB* mutant hypocotyls to wild type size when overexpressed in the *phyB* background. *SOB3/AHL29* overexpression in a wild type background suppresses hypocotyl growth in light grown seedlings, and additionally confers delayed development, larger wavy leaves, and larger flowers in adult plants (Street *et al.*, 2008). When first developing the

technology, Weigel *et al.* (2000) used activation tagging to search for flowering time and leaf growth genes in Arabidopsis. They identified a gene called *ESCAROLA* (*ESC*), also known as *AHL27*, and observed that overexpression of *ESC/AHL27* produces late-flowering plants with more, larger, darker-green crinkled leaves, similar to overexpression of *SOB3/AHL29* (Weigel *et al.*, 2000). Street *et al.* (2008) also found that overexpressing *ESC/AHL27*, like overexpression of *SOB3/AHL29*, suppresses the *phyB* phenotype.

Genes *AHL27* and *AHL29* code for AT-Hook Containing Nuclear Localized (AHL) proteins that have an AT-hook DNA-binding motif as well as a plant and prokaryote conserved/domain of unknown function #296 (PPC/DUF296) (Fujimoto *et al.*, 2004). The PPC/DUF296 domain plays a role in directing AHL proteins to the cell nucleus where they are thought to regulate gene expression by binding Adenine/Thymine-rich DNA with their AT-Hook



domain (Fujimoto *et al.*, 2004;

Street *et al.*, 2008; Zhao,

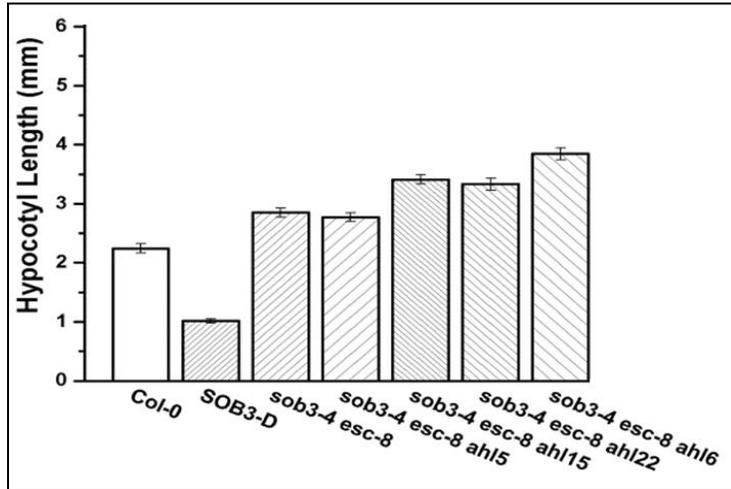
Unpublished).

**Figure 1: Phylogenetic tree of AHL proteins, adapted from Street *et al.* (2008).** Star, cloud, and smiley face symbols indicate co-expression networks based on publicly available gene expression data. Two asterisks following the gene name indicate the presence of two AT-hooks. Blue percentages indicate protein similarity, red percentages indicate nucleotide similarity. SOB3/AHL29 and ESC/AHL27 are boxed together near the middle; AHL9 and AHL11 are boxed together near the top; AHL5 and AHL12 are boxed together above AHL9.

The protein-similarity tree in Figure 1 illustrates relatedness of all 29 proteins in the AHL family in Arabidopsis (Street *et al.*, 2008). Notice that AHL27 and AHL29 are immediate neighbors on this tree. Furthermore, AHL27 and AHL29 demonstrate similar expression patterns, placing them together in co-expression network I. These initial observations suggest that highly similar AHL proteins may play redundant roles in plant growth and development.

In contrast to gene *overexpression*, loss-of-function genetic modification abolishes gene expression, demonstrating more conclusively a gene's role in plant signaling, growth, or development. As noted previously, however, no phenotype appears in Arabidopsis when only knocking out one of multiple potentially redundant genes, such as *AHL27* and *AHL29*, due to functional compensation by one or more partner genes with overlapping function (Weigel *et al.*, 2000; Street *et al.*, 2008). After first evaluating the functions of *AHL27* and *AHL29* with overexpression studies, Street *et al.* (2008) proceeded to generate double mutants with non-functional versions of both genes, in an attempt to overcome the issue of functional redundancy. These double knockout light-grown seedlings have longer hypocotyls than wild type, in direct contrast to suppression of the long *phyB* hypocotyl phenotype conferred by overexpression of either gene (Street *et al.*, 2008). Mutant Arabidopsis triple null for *AHL27-AHL29-AHL15* or *AHL27-AHL29-AHL22* produce hypocotyls longer than the *AHL27-AHL29* double null, suggesting that these genes also act redundantly with *AHL27* and *AHL29* to negatively modulate hypocotyl elongation (Figure 2). Hypocotyls of seedlings triple null for *AHL27-AHL29-AHL5*, however, are no longer than those of the *AHL27-AHL29* double null, suggesting that *AHL5* has a different function from these genes. Conversely, the triple null *AHL27-AHL29-AHL6* is yet taller than the aforementioned triple nulls made with *AHL15* or *AHL22*, suggesting that *AHL6* may

have a more significant effect than either *AHL15* or *AHL22* on regulation of hypocotyl growth in response to light (Zhao and Favero, Unpublished).



**Figure 2: Hypocotyl lengths of various *AHL* null-mutant lines, adapted from Zhao and Favero, Unpublished.** Col-0 wild type control hypocotyl lengths compared to SOB3-D (overexpressor of *AHL29*), and various combinations of *AHL* knockout lines. Measurements were taken from 6 day old seedlings grown under  $22 \mu\text{mol m}^{-2} \text{sec}^{-1}$  white light, at  $25^\circ\text{C}$ .

These observations elicit the following propositions: the fact that each of the 29 *AHL* proteins in *Arabidopsis* carry at least one AT-hook domain and one PPC/DUF296 domain suggests that they may all have some function in common – that protein similarity is positively correlated with protein functional redundancy; yet, the plurality of *AHL*s strongly suggests that each has some unique purpose. Overexpression data in the literature further demonstrates this theme of overlapping yet discrete functions: akin to *AHL27* and *AHL29*, the overexpression of *AHL22* delays flowering and inhibits hypocotyl elongation (Xiao *et al.*, 2009). Diversely, *AHL20* overexpression makes plants susceptible to pathogen attack, which may indicate a role for *AHL20* in plant pathogen response mechanisms (Lu *et al.*, 2010). Bearing in mind this diversity, and having observed in *AHL27* and *AHL29* a positive correlation of protein homology with functional redundancy, the following questions arise: To what extent do other *AHL*s' functions overlap? In what ways are they distinct?

## II. Thesis Activity

To begin answering these questions, we separately overexpressed two phylogenetic *AHL* neighbors, *AHL9* and *AHL11* (Figure 1), in Arabidopsis, expecting that at least one of the two overexpression products would generate useful results for this project. We then compared the phenotype of Arabidopsis *AHL11* overexpression lines (the only one of these two overexpressed genes to reach a testable T2 generation at the time of writing) to that of T2 generation *AHL12* overexpression lines which were generated, and kindly provided for this project, by David Favero, a Ph.D. student in the Neff lab.

## III. Materials and Methods

**a) Bioinformatic analysis:** The Bio-Array Resource for Plant Biology internet database (University of Toronto, Toronto, Canada) was used to generate tissue and developmental stage gene expression maps for *AHL9* (At2g45850), *AHL11* (At3g61310), and *AHL12* (At1g63480).

**b) Expression vectors:** The *AHL9* and *AHL11* genes of interest were acquired pre-inserted in a Gateway-compatible entry vector (Arabidopsis Biological Resource Center, Ohio State University, Columbus). Via the Gateway LR clonease reaction (Invitrogen), each gene of interest was swapped out of their pENTER223.1 entry vectors and into pED15 binary expression vectors to make the new constructs named pAHL9 and pAHL11. (Binary vectors are so named because they can replicate in both *E. coli* and *Agrobacterium*.) Virtual plasmids were constructed using Gene Construction Kit software (Textco Inc.). To verify clone identity, plasmids were digested with the restriction enzyme SacI (New England Biolabs, Inc.), which cuts on the pED15 backbone and within either *AHL* gene of interest. Resulting plasmid fragments were electrophoretically analyzed and compared to expected band sizes.

**c) *Escherichia coli*:** Z-competent *E. coli* strain DH5 $\alpha$  (Zymo Research) were transformed with the expression vectors (from step (b)) carrying our genes of interest and carbenicillin antibiotic resistance. The transformed *E. coli* were grown in lysogeny broth media (Fisher Scientific) containing carbenicillin as a control to select for correctly transformed cells. Endogenous plasmid replication mechanisms of *E. coli* multiplied our vectors, which were extracted from the bacteria using the Zyppy Plasmid Miniprep Kit (Zymo Research).

**d) *Agrobacterium tumefaciens*:** Competent *Agrobacterium* strain GV 3101 resistant to gentamicin antibiotic was transformed with the expression vectors extracted from *E. coli* in the previous step. Large cultures of *Agrobacterium* carrying the vector with the gene of interest were grown in carbenicillin and gentamicin selective YEP media in preparation for plant dipping (step (e)). Carbenicillin and gentamicin antibiotics in the growth media selected for *Agrobacterium* which had been correctly transformed with the vector of interest.

**e) *Arabidopsis thaliana*:** Using the floral dip protocol described by Clough and Bent (1998), we induced *Agrobacterium* carrying the expression vectors to transform wild type *Arabidopsis* strain Columbia-0 with the *AHL* genes of interest. In brief: the large cultures of *Agrobacterium* from the previous step were centrifuged to precipitate the bacteria; cells were re-suspended in 5.0% sucrose solution with a surfactant and 6-benzylaminopurine, a plant hormone; seed pods (siliques) were removed from *Arabidopsis* inflorescences, leaving only fresh flowers on the plant; inflorescences were submerged in the *Agrobacterium* sucrose solution for fifteen minutes. Dipped plants were then grown to maturity, at which time seeds were harvested.

**f) Discovering single transgene insertion lines:** To screen for successful transformants, the first generation (designated generation T1) seeds harvested from plants transformed with *AHL9* and *AHL11* overexpression constructs in the previous step were sprinkled on Basta herbicide (Bayer

CropScience) selection plates made with Linsmaier and Skoog (1965) nutrient media (LS) (Phytotechnology Laboratories) and 0.8% Phytoblend agar (Caisson Laboratories, Inc.). All Basta selection plates used in our analyses were made with this formula. Seedlings surviving the Basta herbicide selection, thus putatively carrying the transgene, were transplanted to Sunshine Mix #4 soil (Sun Gro Horticulture Canada Ltd.). Seeds harvested from the lines which putatively overexpress *AHL11* (now T2 generation) were again plated on Basta selection media, this time to identify lines that behaved according to the Mendelian law of segregation: if a single transgene was indeed inserted into the T1 genome, T2 progeny of that transgenic event should segregate phenotypically in such a way that three individuals survive on selection (i.e. carry the transgene that confers resistance to Basta) for every individual that dies (Acquaah, 2007). Observed ratios were counted and compared to the expected 3:1 ratio using chi-square statistical analysis,  $p=0.05$ .

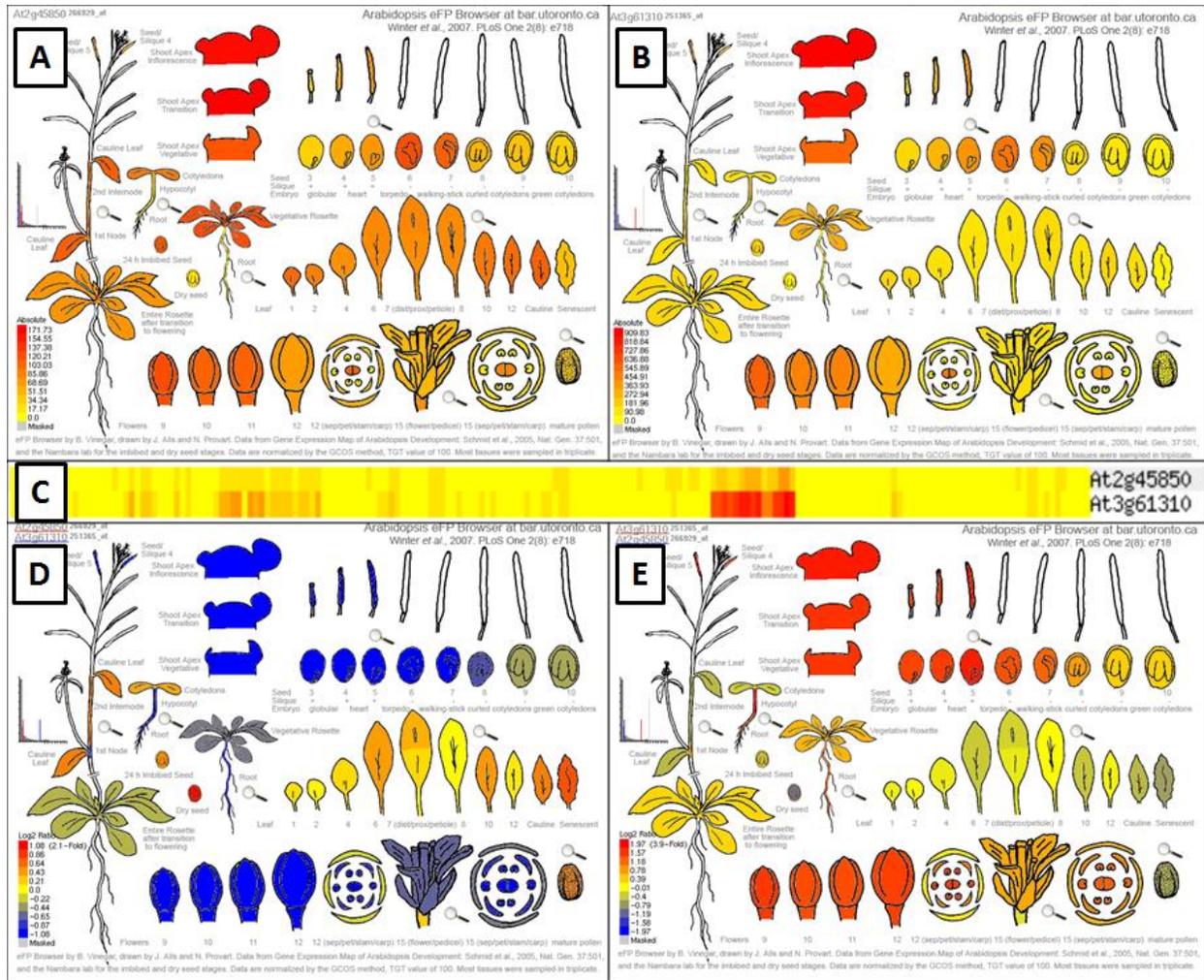
**g) Gene expression analysis:** pAHL11.14 and pAHL11.15, two of the many independent, single insertion, T2 *AHL11* overexpression lines discovered in the previous step, were chosen for further analysis. Two other lines generated by David Favero to overexpress *AHL12* were evaluated as well. These T2 lines, pDF34 4/2-12 and pDF34 4/2-3, were also thought to carry single locus insertions. T2 seeds of all four lines were seeded onto Basta selection plates alongside Basta resistant and Basta sensitive Columbia-0 wild type controls. Plates were grown in sub-saturating white light (approximately  $20\text{-}23 \mu\text{mol m}^{-2}\text{s}^{-1}$ ) at  $25^{\circ}\text{C}$  for seven days. Samples of about five surviving seedlings per line were collected and flash frozen in liquid nitrogen. RNA was then extracted from these tissue samples using Qiagen's RNA extraction kit and protocol. Extracted RNA was immediately reverse transcribed to cDNA using Colntech's EcoDry Premix for reverse transcriptase polymerase chain reaction (RT-PCR). Gene expression levels (i.e.

cDNA abundance) of *AHL11* and *AHL12* in their two respective overexpression lines, and in the two wild type controls, were measured semi-quantitatively via PCR with *AHL11* and *AHL12* gene-specific primers. As an internal control, expression levels of these two genes were normalized against expression level of the housekeeping gene actin. *AHL11* primer sequences: 5' – CATATGGGGGTGATGCATATC – 3' and 5' – TATCGTTCTCGCATCAAGGTC – 3'. *AHL12* primer sequences: 5' – TCTGTGAATTCGACGGAAGAGAAGCAATGG – 3' and 5' – TAGAAGCTTTCATCCACGAGTCAAATCAA – 3'. Actin primer sequences: 5' – TCGGTGGTTCCATTCTTGCT – 3' and 5' – GCTTTTTAAGCCTTTGATCTTGAGAG – 3'.

**h) Phenotypic analysis:** Seeds from the same four T2 lines that were evaluated in the previous step, along with two sectors of Columbia-0 wild type controls, were grown on non-selection LS nutrient media with a 1.0% Phytigel substrate (Sigma-Aldrich Co. LLC.) at the same time, in the same growth chamber, and under the same light and temperature conditions as the gene expression seedling plates. More than 30 seedlings from each of the four lines, along with the two wild type controls, were up-rooted and photographed with an Epson Perfection 1200U scanner on the same day that samples were collected from Basta selection plates for gene expression analysis. Hypocotyl lengths were measured using ImageJ software (National Institutes of Health). These data were then graphed in histograms to demonstrate height frequencies and segregation ratios of the putative overexpression lines and their wild type siblings.

#### **IV. Results and Discussion**

**a) Bioinformatic analysis:** *AHL9* and *AHL11* are paired neighbors on the AHL phylogenetic tree, meaning they are structurally similar (Figure 1). Based on publicly available microarray data, *AHL9* and *AHL11* have very similar expression *patterns* as seen in Figure 3A, B, and C.

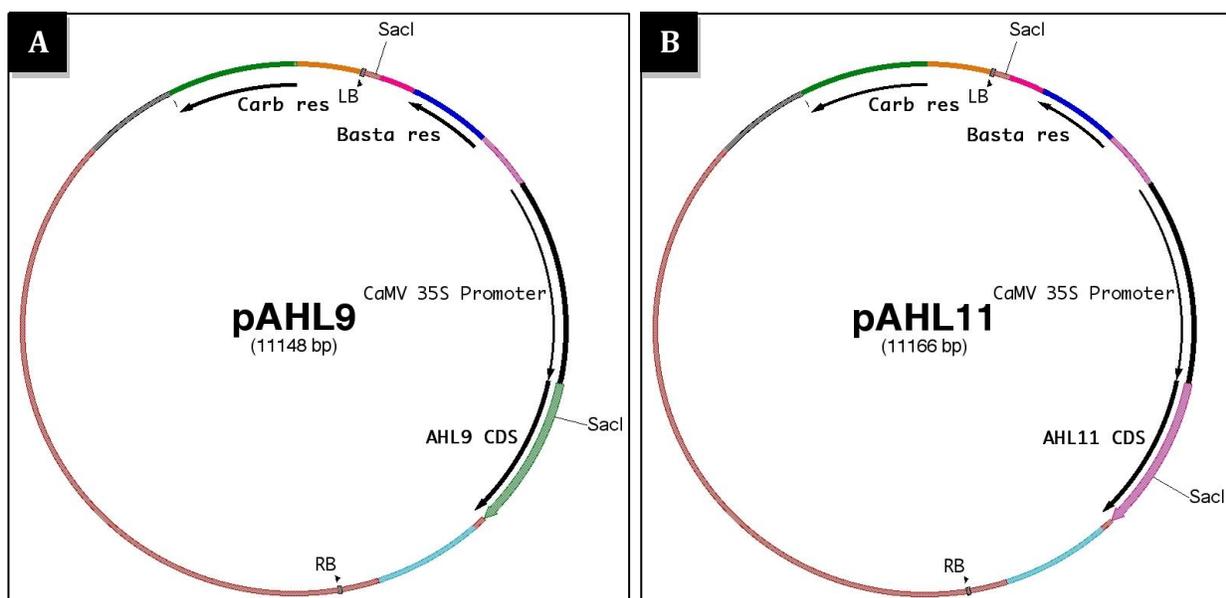


**Figure 3: Microarray gene expression data for *AHL9* and *AHL11* derived from the Bio-Array Resource, University of Toronto, Ontario, Canada.** Qualitative representation of gene expression levels in different tissues and developmental stages of (A) *AHL11/At3g61310* and (B) *AHL9/At2g45850*. (C) *AHL9/At2g45850* (top line) expression compared to that of *AHL11/At3g61310* (bottom line) in multiple tissues at multiple points in development. (D) *AHL9* expression with respect to *AHL11*. (E) *AHL11* expression with respect to *AHL9*. Note: Images are intended to render general, qualitative expression level comparisons and are not meant for numerical evaluation.

Of note, both genes are most highly expressed in the shoot apex of the inflorescence. When *AHL11* tissue-specific microarray gene expression level was compared to that of *AHL9* and vice versa (Figure 3D and E), however, *AHL11* is seen to be more highly expressed than *AHL9* in flowers, roots, siliques, and embryos, whereas *AHL9* is more highly expressed than

*AHL11* in the inflorescence, seeds, cotyledons, pollen, and some stages of rosette leaves. The molecular similarity and parallel expression patterns of these two genes, and yet their obviously discrete expression levels in different tissues, highlighted them as candidates for overexpression study to begin to describe the amount of functional redundancy and novelty displayed by highly similar genes in the *AHL* family.

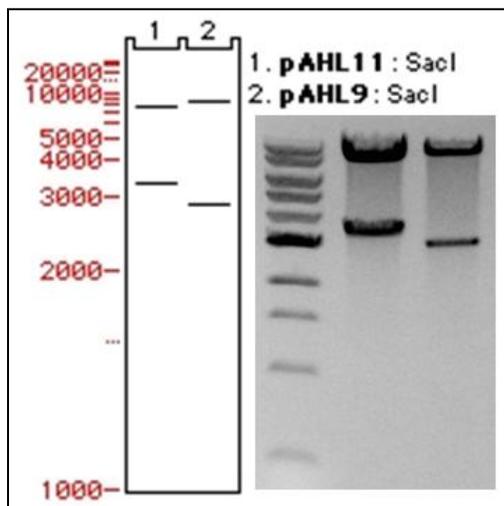
**b) Overexpression vectors:** Hoekema *et al.* (1983), in a letter to *Nature*, first described *Agrobacterium tumefaciens*' potential for guided plant transformation. Soil-dwelling *Agrobacterium* insert tumor-causing genes into injured plant cells, genetically modifying their host plant to produce food and shelter structures for them. These tumor genes are found between specific sequences designated left border (LB) and right border (RB) on the natural expression vector – the so-called “tumor inducing” or Ti-plasmid. The Ti-plasmid can be disarmed by replacing the tumor-causing sequence with different genes of interest. New DNA sequences falling between LB and RB will then be cut out of the vector and pasted into the plant genome in floral cells via *Agrobacterium*'s natural plant genetic transformation mechanism (Hoekema, *et al.*, 1983).



**Figure 4: Virtual representation of expression vectors pAHL9 and pAHL11.** (A) pAHL9 and (B) pAHL11. Clockwise from top: Carb res: carbenicillin resistance gene; LB: T-DNA left border; SacI: restriction endonuclease SacI cut site; Basta res: Basta resistance gene; CaMV 35S Promoter: CaMV 35S constitutive promoter for gene of interest; AHL9/AHL11 CDS: AHL9/AHL11 coding sequence; RB: T-DNA right border. (Gene Construction Kit, Textco Inc.)

The expression vector pED15, our Ti-plasmid equivalent, carries a carbenicillin antibiotic resistance gene, a Basta herbicide resistance gene, and the cauliflower mosaic virus (CaMV) 35S promoter, which induces constitutive gene expression (Odell *et al.*, 1985). The plasmids pAHL9 and pAHL11 (Figure 4A and B, respectively) were made by removing the respective *AHL* coding sequences (CDS) from the pENTER223.1 entry vector, and inserting them into pED15 directly following the CaMV 35S promoter sequence. This swap was accomplished by using Gateway cloning reagents and protocol (Invitrogen). T-DNA LB and RB sequences on the expression vector flank the Basta resistance gene, the CaMV 35S promoter, and the *AHL* gene of interest.

pAHL9 and pAHL11 vectors were extracted from *E. coli* and digested with the restriction



endonuclease SacI. Resulting fragments matched the predicted band sizes when separated electrophoretically on a gel (Figure 5), verifying plasmid identity.

**Figure 5: Plasmid digestions.** Computer-generated gel with predicted banding pattern. Inlay is photograph of experimental results with the following loading order (left to right): 1kb ladder; pAHL11; pAHL9. Dark band in the 1kb ladder is 3.0 kb long. (Gene Construction Kit, Textco Inc.)

**c) Chi-square analysis:** Due to the random nature of *Agrobacterium* mediated transformation, some overexpression lines generated using the floral dip procedure may carry multiple

transgenes. Single locus insertion lines are the most desirable and scientifically tenable transgenic plants to work with since they tend to behave with more stability than lines with multiple insertions, and can thus be better characterized over multiple generations (Dr. Michael Neff, personal communication). Of the 25 lines that survived the screen for primary transformants, eighteen behaved statistically equivalent to T2 single locus insertion lines at a  $p=0.05$  cutoff chi-square value (Figure 6). At the first screening, lines pAHL11.13, 11.18, 11.21, and 11.35 all gave borderline significant results. Of these, lines pAHL11.13, 11.21, and 11.35 were screened again to gather more data points. Chi-square analysis was repeated on additional data gathered from this second round of screening. After re-screening and re-analysis, pAHL11.13 and 11.21 were promoted to putative single locus insertion status, while pAHL11.35 was rejected.

AHL11 OX Lines: Chi-Square Analysis									
Row #	Line #	Seed #	Germ	Germ Rate	Survived	Died	Ratio	Chi-square test	
1	3	39	37	0.95	28	9	3.11	P=0.9244; ACCEPT	
2	4	44	44	1.00	32	12	2.67	P=0.7277; ACCEPT	
3	5	47	46	0.98	33	13	2.54	P=0.6095; ACCEPT	
4	6	42	42	1.00	33	9	3.67	P=0.5930; ACCEPT	
5	7	41	37	0.90	33	4	8.25	P=0.0462; REJECT	
6	8	42	41	0.98	34	7	4.86	P=0.2411; ACCEPT	
7	9	46	44	0.96	30	14	2.14	P=0.2963; ACCEPT	
8	10	59	58	0.98	46	12	3.83	P=0.4484; ACCEPT	
9	12	46	35	0.76	17	18	0.94	P=0.0003; REJECT	
10	13	49	48	0.98	41	7	5.86	P=0.0956; Doubtful	
11	13'	92	88	0.96	64	24	2.67	P'=0.6225; ACCEPT	
12	14	51	49	0.96	39	10	3.90	P=0.4579; ACCEPT	
13	15	48	48	1.00	33	15	2.20	P=0.3173; ACCEPT	
14	17	38	37	0.97	26	11	2.36	P=0.5064; ACCEPT	
15	18	39	36	0.92	22	14	1.57	P=0.0543; Doubtful	
16	19	54	52	0.96	37	15	2.47	P=0.5218; ACCEPT	
17	20	50	50	1.00	37	13	2.85	P=0.8703; ACCEPT	
18	21	46	46	1.00	26	16	1.63	P=0.0500; Doubtful	
19	21'	88	84	0.95	58	22	2.64	P=0.6056; ACCEPT	
20	22	54	52	0.96	38	14	2.71	P=0.7488; ACCEPT	
21	23	48	38	0.79	35	3	11.67	P=0.0149; REJECT	
22	24	40	38	0.95	28	10	2.80	P=0.8514; ACCEPT	
23	25	45	34	0.76	21	13	1.62	P=0.0745; Doubtful	
24	25'	84	61	0.73	38	23	1.65	P=0.0219; REJECT	
25	26	43	27	0.63	10	17	0.59	P=0.0001; REJECT	
26	27	44	43	0.98	30	13	2.31	P=0.4281; ACCEPT	
27	28	41	34	0.83	19	15	1.27	P=0.0100; REJECT	
28	29	39	31	0.79	22	9	2.44	P=0.6041; ACCEPT	

**Figure 6: Chi-square analysis to identify T2 generation *AHL11* overexpression lines with single locus transgene insertions. Line#: e.g. 3 = pAHL11.3. Chi-square cutoff value set at  $p=0.05$ .**

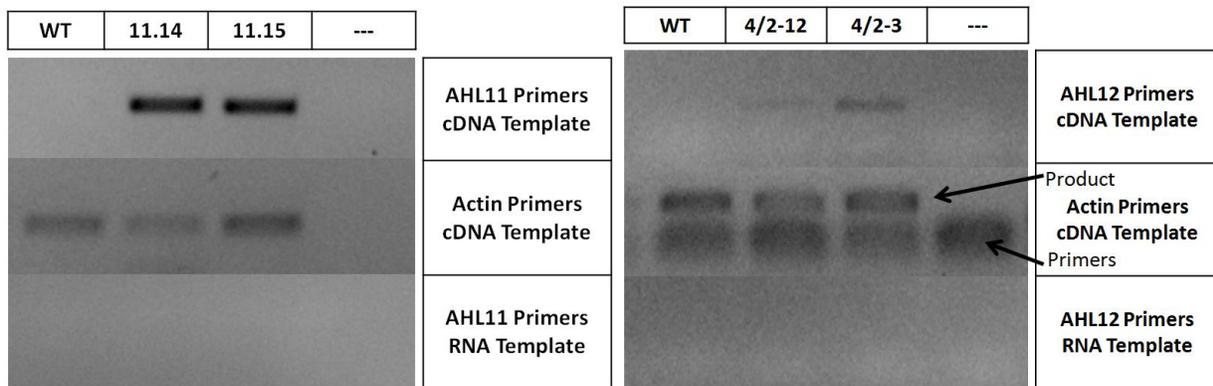
Although it was relatively simple to identify and count those seedlings that did and did not survive the Basta herbicide treatment, surviving T2 siblings of the same T1 parentage were not all phenotypically identical – some were taller than others, and some had curled

cotyledons while others appeared unperturbed. Perhaps this phenotypic variability, witnessed even in surviving T2 lines with single locus insertion segregation behavior, was due to microclimate variability on the growth plates or to unstable gene expression. pAHL11.14 and 11.15 were chosen for further phenotypic and gene expression study because of their obviously perturbed phenotypes as seedlings and adults, and because these phenotypes were homogenous among all surviving lines. At the time of writing, T2 *AHL9* overexpression seeds were dormant, thus not competent for germination and growth studies. David Favero, however, had generated T2 lines overexpressing *AHL12* and had already identified those with single locus insertions. Thus we decided to compare seedling phenotype and gene expression data of our pAHL11.14 and 11.15 lines to David's lines pDF34 4/2-12 and 4/2-3.

**d) Gene expression data:** When we first extracted RNA from our six samples (pAHL11.14, 11.15, wild type, pDF34 4/2-12, 4/2-3, wild type), we ran an electrophoresis gel of the RNA extraction reaction mixture to determine if we had successfully extracted any non-degraded RNA. Each sample's lane on the gel showed two large bands corresponding to 28S and 18S ribosomal subunits, indicating successful RNA extraction (data not shown). Reverse transcribed RNA extraction product, cDNA, (see Materials and Methods, step (g)) was amplified by PCR with actin primers to determine relative quantities of total RNA extracted from each line (data not shown). Actin is a housekeeping gene, and should theoretically have equal relative transcription abundance in all six samples tested, whether from wild type or gene overexpression lines. Some researchers have taken issue with using actin as a gene expression normalization standard, however, citing variability in actin expression in different tissues and under different growth conditions as grounds for mistrust of actin-normalized data (Czechowski *et al.*, 2005). Nevertheless, since we sampled essentially the same tissues for all lines, and all seedlings were

grown under identical circumstances, we feel confident that actin expression can be accurately used as an indicator of total cDNA abundance. The only limitation in this assumption is if *AHL11* or *AHL12* overexpression somehow perturbs actin expression.

From actin amplification results, we normalized cDNA template input for each line. Subsequent semi-quantitative PCR demonstrated that indeed *AHL11* was overexpressed in pAHL11.14 and pAHL11.15 and that *AHL12* was overexpressed in pDF34 4/2-12 and 4/2-3 (Figure 7). No product is visible when using gene-specific primers and wild type template for either *AHL11* or *AHL12*. However, microchip data for tissue-specific expression of *AHL11* (Figure 3A) demonstrate that *AHL11* is not highly expressed in seedlings, if at all; therefore, we cannot expect high levels of wild type PCR product for this gene. The same is true of *AHL12* (data not shown). Furthermore, amplification with actin primers produced a very visible product for wild type cDNA template, indicating presence of RT-PCR cDNA template for both wild type samples. That no product appears in the lanes using extracted RNA as a template assures us that there has been no confounding genomic DNA contamination of our cDNA samples.

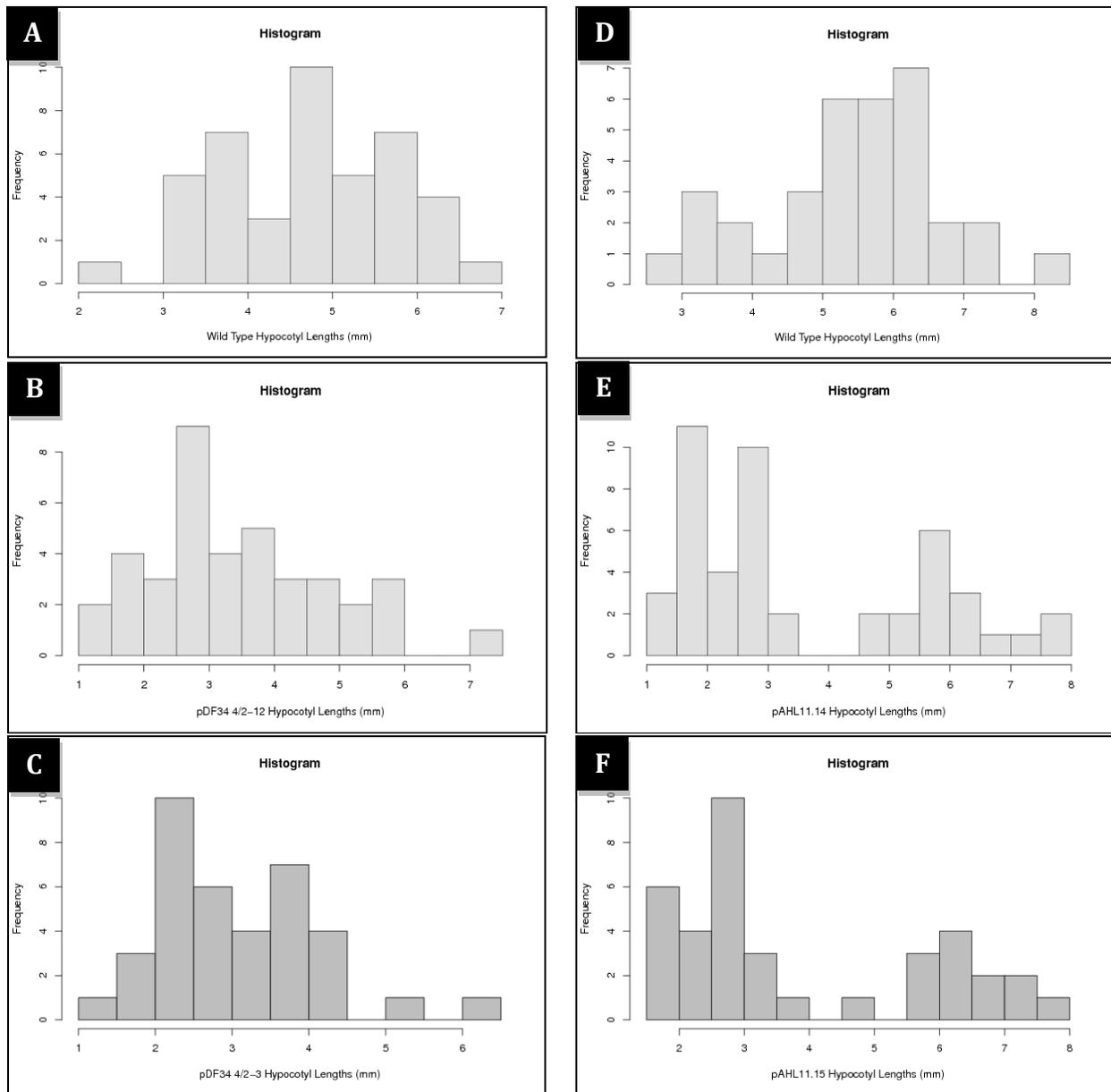


**Figure 7: Semi-quantitative RT-PCR.** WT: wild type; ---: no-template negative control.

**e) Hypocotyl phenotypes:** Given that published data for characterization of *AHL29* and *AHL27* (Street *et al.*, 2008), and *AHL22* (Xiao *et al.*, 2009) demonstrates that overexpression of these

three genes from the AHL family suppresses hypocotyl elongation in Arabidopsis, we decided to measure seedling hypocotyls of our four selected overexpression lines and two wild type controls. These measurements were plotted in histograms to demonstrate population tendencies as well as hypocotyl length phenotypic segregation (Figure 8). Since seeds were from T2 lines with single locus insertions, we expected to witness a segregation ration of 3:1 short:tall, corroborating our observations made when these seeds were placed on selection media, and segregated 3:1 live:dead. This segregation pattern is somewhat detectable in the histograms, though not as clearly in the case of the two pDF34 lines, as for the pAHL11 lines.

The wide range in wild type hypocotyl lengths helps explain why the segregation ratio is not so clearly seen in the T2 overexpression lines. Unlike the binary outcome of survival or death on Basta selection media, hypocotyl length is a quantitative trait demonstrating variability among individuals of identical genotypes (Acquaah, 2007). Replication in triplicate would help to establish stronger trends. Regardless, in each histogram of lines overexpressing *AHL11* or *AHL12*, a distinct enrichment in short-hypocotyl individuals can be seen, concurrent with a decrease in tall seedling counts. Our preliminary quantitative observations of early-generation Arabidopsis transformants therefore imply that overexpression of *AHL11* and *AHL12* confers dwarf seedling stature.



**Figure 8: Histogram representation of hypocotyl lengths.** (A) Wild type control from pDF34 seedling plate; (B) pDF34 4/2-12; (C) pDF34 4/2-3; (D) wild type control from pAHL11 seedling plate; (E) pAHL11.14; and (F) pAHL11.15. Plants were grown on 1.0% Phytigel non-selection LS nutrient media plates in sub-saturating white light at approximately  $20\text{-}23 \mu\text{mol m}^{-2} \text{s}^{-1}$  at  $25^\circ\text{C}$  for seven days.

**f) Adult phenotypes:** As adults, most Arabidopsis lines overexpressing *AHL11* are smaller than wild type, with reduced rosette area and spindly, weak inflorescences (Figure 9). The line pAHL11.20, however, was as large as – if not larger than – the wild type control. This is likely

due to epigenetic effects. For example, the transgene may have landed in a heterochromatic region of the genome from which it is not expressed, or it may be silencing endogenous *AHL11* expression through the RNAi pathway. This phenomenon of gene silencing by overexpressing plant-endogenous genes was first observed and reported by Napoli *et al.* (1990) who were trying to overexpress a gene for purple color in petunias and came up with white flowered plants. The prevailing adult phenotype observed for both *AHL11* and *AHL12* (*AHL12*: David Favero, personal communication; data not shown) overexpression, however, is diminished plant size and vigor. As opposed to the dwarf seedling phenotype, which agrees with published data for overexpression of other *AHL* genes, the dwarf adult phenotype witnessed in pDF34 and pAHL11 lines contrasts with the larger adult leaf phenotype recorded for overexpression of *AHL27* and *AHL29* (Street *et al.*, 2008).



11.20	WT	11.8	11.24	11.10	11.15	11.14	11.9
-------	----	------	-------	-------	-------	-------	------

**Figure 9: Dwarf adult phenotypes of plants overexpressing *AHL11*.** Wild type, pAHL11.15, and 11.14 lines are highlighted. About six plants per pot; each vertical set of two pots contains plants from the same line. All plants are the same age, seeded and transplanted simultaneously.

## V. Conclusions

In completing this project, I have successfully (a) created new plasmid constructs, (b) used those constructs to transform *Arabidopsis*, (c) identified transgenic *Arabidopsis* carrying a

single copy of the inserted gene of interest, (d) extracted RNA from two of my single locus insertion lines and two of David Favero's, (e) analyzed that RNA to demonstrate that the gene of interest is truly being overexpressed, and (f) observed both seedling phenotypes of plants overexpressing *AHL11* and *AHL12* and adult phenotypes of plants overexpressing *AHL11*. Through these steps, I have demonstrated that overexpression of *AHL11* and *AHL12* confers a dwarf seedling phenotype in Arabidopsis. From this observation, we can conclude that *AHL11* and *AHL12* may share some functionality with other genes in the *AHL* family that have already been characterized. Yet we can also agree that since *AHL27* and *AHL29* adults are larger than wild type (Street *et al.*, 2008), and *AHL11* and *AHL12* adults are dwarf, these two sets of genes may have some distinct functions.

## **VI. Pitfalls**

Multiple pitfalls may impact the experiments and results reported here. Perhaps most obvious is that using the CaMV 35S constitutive promoter to overexpress a gene can cause mis-expression of the gene of interest where and when the gene would not normally be expressed. As a case in point: from microarray and RT-PCR data, we gather that neither *AHL11* nor *AHL12* are highly expressed in wild type seedlings; yet we witnessed a seedling phenotype when overexpressing these genes. This pitfall is further addressed in the following section, Future Directions.

The main hazard underlying the RT-PCR experiment is the assumption that actin expression is constant in all samples, and is thus an accurate normalization control. The use of multiple other known reference genes for normalizing cDNA concentrations would help to address this concern. A more general caveat lies in the fact that few of the foregoing experiments have been done in triplicate. This lack of replication may have allowed misleading data to be

reported. We therefore lack a measure of confidence in our results; that we observed somewhat of what we hypothesized and expected regarding distinct but overlapping gene function lends much emotional, but little additional scientific/statistical credence to our work.

## **VII. Future Directions**

As stated in the Introduction and discussed in the Conclusion and Pitfalls, gene overexpression is not always a reliable way to characterize gene function. Thus, following the experimental pattern established by Street *et al.* (2008), after observing overexpression phenotypes we must proceed to acquire and characterize gene loss-of-function lines for *AHL11* and *AHL12*. If our observations herein have been correct, loss-of-function seedling and adult phenotypes should be opposite from our observed overexpression phenotypes: Arabidopsis null for *AHL11* and *AHL12* should have taller seedling hypocotyls and produce larger, stronger adult plants with respect to wild type. This may not be entirely the case, however, since we have already noticed that neither *AHL11* nor *AHL12* are expressed in normal seedlings (*AHL11*: Figure 3; *AHL12*: data not shown) and therefore knocking these genes out may not produce any phenotype in a tissue in which neither gene is normally expressed.

In the first half of this project, I also overexpressed *AHL9* in Arabidopsis. *AHL9* overexpression lines should be screened and characterized, just as were those of *AHL11*, thereby adding to the burden of data that is beginning to describe functionality of the 29 genes in the *AHL* family in Arabidopsis. Any promising traits uncovered in these genetic studies should be considered for crop enhancement. Some have already begun exploiting the potential of this *AHL* gene family for future crop improvements; for example, Mendel biotechnology holds a patent on *AHL25* (a.k.a. *HERCULES*), describing it as a “method for modifying plant biomass” when overexpressed (Mendel Biotechnology). Also, manipulating *AHL29* in camelina, a low input,

dryland oil seed crop of interest for biofuel production in Washington State, has induced better seedling emergence under dryland farming conditions (Koirala *et al.*, Unpublished). Using this technology to stimulate thicker crop stands would increase the viability and efficiency of local dryland camelina farming. Since overexpression of *AHL11* makes adults dwarf, why not knock the gene out – along with its redundant partners – to make a super-cabbage with 25lb heads, or an alfalfa plant that grows 50% larger between cuttings? These ideas are just the beginnings of what may be accomplished when the *AHL* gene family is fully characterized and applied to improve crops.

**VII. Acknowledgements:** I must thank each member of the Neff lab, who all to greater or lesser extent sacrificed time, space, reagents, and equipment to help me learn what I needed to know to complete this project. Special thanks to: Dr. Neff for agreeing to take me on in his lab, and for help with how to be a scientist; Jianfei Zhao, who was my closest mentor throughout the many stages of the project, and who spent the most time and energy leading me through the entire process; Pushpa Koirala for her help with all the many practical lab details involved; David Favero for seeds, advice, and revision of this document; Reuben Tayengwa for introducing me to molecular plant science; Kulbir Sandhu for guidance in seedling measurements.

### **VIII. References Cited**

**Acquaah, G.** 2007. Principles of plant genetics and breeding. Blackwell Publishing.

**Clough, S.J., and Bent, A.F.** 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant Journal*. **16**: 735 – 743.

**Czechowski, T., Stitt, M., Altmann, T., Udvardi, M.K., and Scheible, W.** 2005. Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiology*. **139**: 5-17.

**Fujimoto, S., Matsunaga, S., Yonemura, M., Uchiyama, S., Azuma, T., and Fukui, K.** 2004. Identification of a novel plant MAR DNA binding protein localized on chromosomal surfaces. *Plant Molecular Biology*. **56**: 225-239.

- Gee, H.** 2002. Food & the future. *Nature*. **418**: 667.
- Hoekema, A., Hirsch, P.R., Hooykaas, P.J.J., and Schilperoort, R.A.** 1983. A binary plant vector strategy based on separation of *vir*- and T-region of the *Agrobacterium tumefaciens* Ti – plasmid. *Nature*. **303**: 179-180.
- Kishmore, G.M., and Shewmaker, C.** 1999. Biotechnology: Enhancing human nutrition in developing and developed worlds. *Proceedings of the National Academy of Sciences*. **96**: 5968-5972.
- Linsmaier, E.M., and Skoog, F.** 1965. Organic growth factor requirements of tobacco tissue culture. *Physiologia Plantarum*. **18**: 100-127.
- Lu, H., Zou, Y., and Feng, N.** 2010. Overexpression of *AHL20* negatively regulates defenses in *Arabidopsis*. *Journal of Integrative Plant Biology*. **52**: 801-808.
- Napoli, C., Lemieux, C., and Jorgensen, R.** 1990. Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. *The Plant Cell*. **2**: 279-289.
- Odell, J.T., Nagy, F., and Chua, N.** 1985. Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. *Nature*. **313**: 810-812.
- Smith, H., and Whitelam, G.C.** 1997. The shade avoidance syndrome: multiple responses mediated by multiple phytochromes. *Plant, Cell and Environment*. **20**: 840-844.
- Street, I.H., Shah, P.K., Smith, A.M., Avery, N., and Neff, M.M.** 2008. The AT-hook-containing proteins SOB3/AHL29 and ESC/AHL27 are negative modulators of hypocotyl growth in *Arabidopsis*. *Plant Journal*. **54**: 1-14.
- Weigel, D., Ahn, H.J., Blázquez, M.A., Borevitz, J.O., Christensen, S.K., Fankhauser, C., Ferrándiz, C., Kardailsky, I., Malancharuvil, E.J., Neff, M.M., Nguyen, J.T., Sato, S., Wang, Z.Y., Xia, Y., Dixon, R.A., Harrison, M.J., Lamb, C.J., Yanofsky, M.F., and Chory, J.** 2000. Activation tagging in *Arabidopsis*. *Plant Physiology*. **122**: 1003-1013.
- Xiao, C., Chen, F., Yu, X., Lin, C., and Fu, Y.F.** 2009. Over-expression of an AT-hook gene, *AHL22*, delays flowering and inhibits the elongation of the hypocotyl in *Arabidopsis thaliana*. *Plant Molecular Biology*. **71**: 39-50.