GENETIC MECHANISMS BEHIND FLOWER COLOR VARIATION IN *CAULANTHUS AMPLEXICAULIS* VAR. *AMPLEXICAULIS* (CAA) AND *CAULANTHUS AMPLEXICAULIS* VAR. *BARBARAE* (CAB)

An Undergraduate Research Scholars Thesis

by

REBECCA LANE

Submitted to Honors and Undergraduate Research
Texas A&M University
in partial fulfillment of the requirements for the designation as

UNDERGRADUATE RESEARCH SCHOLAR

Approved by
Research Advisor: Dr. Alan Pepper

May 2013

Major: Molecular and Cell Biology

TABLE OF CONTENTS

Page

ABSTRACT

Genetic Mechanisms Behind Flower Color Variation in *Caulanthus amplexicaulis* var. *amplexicaulis* (CAA) and *Caulanthus amplexicaulis* var. *barbarae* (CAB). (May 2013)

Rebecca Lane
Department of Biology
Texas A&M University

Research Advisor: Dr. Alan Pepper
Department of Biology

Genetic research opens doors to applied science. The more that is understood about the genetics of something, the more that organism can be used to achieve some desired outcome, an example being increase in crop yield or pest resistance. This study seeks to locate the gene for flower color in two species of the flower *C. amplexicaulis*, CAA and CAB, by looking for sequence similarity between the *Caulanthus* genome and known transcription factors involved in pigment production in *Arabidopsis thaliana* using plants from the family Brassicaceae as intermediates. Once a high degree of similarity was found in a region of the CAB and CAA genomes, a gene model was developed to achieve an understanding of why CAA is a different color than CAB. The results of this study will then aid in understanding other differences between the flowers.

NOMENCLATURE

BLAST	Basic Local Alignment Search Tool
CAA	*Caulanthus amplexicaulis* var. *amplexicaulis*
CAB	*Caulanthus amplexicaulis* var. *barbarae*
EB	Elution Buffer
Coding DNA Sequence	CDS
Single Nucleotide Polymorphism	SNP
Open Reading Frame	ORF

CHAPTER I

INTRODUCTION

C. amplexicaulis is a member of the Streptanthoid complex, which is known for being extremely diverse (Burrell et al. 2011). As a prime example, the two *Caulanthus* subspecies CAA and CAB are very ecologically different; while they both grow in California, *barbarae* is found in serpentine soils containing heavy metals, but *amplexicaulis* grows in nonserpentine soil. Tolerance to nickel, which is toxic to most plants, is known to be a major adaptation among plants that grow in serpentine soils, and is characterized by reduced nickel uptake or some other mechanism of coping with the high levels. A marked difference in the accumulation of nickel in CAA and CAB is present though the mechanism has yet to be uncovered (Burrell et al. 2012). Despite this environmental variation, the only morphological difference is color; the sepals of *amplexicaulis* are purple and those of *barbarae* are white. It is unknown whether this difference is due to selection or just genetic drift. It is considerably important to evolution however considering the color of a flower affects how it is pollinated, and subsequently the evolutionary trajectory of the species. *Caulanthus* is a close relative of the well-studied plant *Arabidopsis thaliana* within the family Brassicaceae. Initial genetic linkage mapping indicated that the underlying color gene is between two markers corresponding to the *A. thaliana* genes AT1G64990 and AT1G66740 (Burrell et al. 2011). In the *Arabidopsis* genome, MYB transcription factors MYB113 (AT1G66370), MYB114 (AT1G66380) and PAP2／MYB90 (AT1G66390) lie between these two markers. These genes are known to be regulators of anthocyanin biosynthetic enzymes (The Arabidopsis Information Resource), and therefore key starting points in the search for the cause of different pigmentation in *amplexicaulis* and

barbarae. Orthologs of these genes were first found in *Brassica* given that the *Brassica* lineage is between those of *Arabidopsis* and *Caulanthus*, and therefore should match *Caulanthus* more closely. With the use of bioinformatics software, searches of both *Caulanthus* genomes were made to find genes closely related to these *Arabidopsis* and transcription factors and their *Brassica* orthologs, and markers were developed to test co-segregation with the flower phenotype. The focus of this paper will be on characterization of the genetic basis for the purple flower color in the CAA subspecies.

This project will be a proof of strategy. By testing a simple trait such as flower color, it can be shown that this technique for analyzing genetic differences in CAA and CAB works, allowing further research to be done, for example on the molecular basis for the ability of CAB to grow in serpentine soils due to its tolerance to heavy metals, low nitrogen, low calcium, and low phosphorous. If the genetic basis behind these features can be determined, it will shed light on subjects such as adaptation to environment and speciation.

CHAPTER II

METHODS

Location of candidate gene

Coding DNA sequence (CDS) from *Arabidopsis thaliana* gene AT1G66370 was entered into a Basic Local Alignment Search Tool (BLAST) on the NCBI website and a blastx search conducted (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The top five hits from the family Brassicaceae with a minimum expect value of 10e-10 were recorded and compared with results from a BLAST done with AT1G66380 as the query. The top *Brassica* hit from the search with AT1G66370 was then used to conduct a BLAST search in CLC Genomics (Bjarne et al.) using the *Caulanthus amplexicaulis* var. *amplexicaulis* and *barbarae* transcriptomes as the databases to be searched. All loci with confidence 1 or greater were accepted and a reciprocal BLAST back to *Arabidopsis* and *Brassica* was done on the NCBI BLAST site (blast.ncbi.nlm.nih.gov/Blast.cgi). The sequences from *Caulanthus* were then aligned using the commercial software packages Geneious (Drummond et al.) and CLC along with the *Brassica* and *Arabidopsis* genes to ensure that a consensus existed. Alignments of both nucleotide and protein sequence were carried out until significant similarity was confirmed.

Verification

The candidate sequences were searched for single nucleotide polymorphisms (SNPs) between CAA and CAB that resulted in different restriction sites and primers were created to isolate these polymorphisms. Two forward primers and four reverse were ordered: FC1-F, FC1-R3, FC1-R4, FC2-F, FC2-R2, and FC2-R3. The primers were then prepared in test tubes. First 10mM Tris was

added to make a concentration of 100pmol/μL, and then each was further diluted in 10mM elution buffer (EB) to a final oligonucleotide concentration of 7.5 pmol/μL.

Next, tubes were prepared for PCR. In an eppi tube, 100 μL 2X GoTaq and 70 μL dH_2O were combined. 17 μL of the mix was pipetted into 8 strip tubes. 1 μL forward primer FC2-F and 1 μL reverse primer FC2-R2 was added to tubes 1-4, and 1 μL primers FC2-F and FC2-R3 were added to tubes 5-8. Lastly, 2.3 μL CAB genomic DNA was transferred to tubes 1 and 5, 1 μL CAA genomic DNA to tubes 2 and 6, 1μL F_1 DNA to tubes 3 and 7, and 1μL water to tubes 4 and 8 as a negative control. After the contents of the tubes had been mixed with a pipette and centrifuged shortly, they were run through a thermal cycler for 2 min at 95°C, 0.45 min at 95°C, 0.45 min at 55°C, and 1 min at 72°C for 35 cycles, then 5 min at 72°C and allowed to cool at 12°C. The reaction mixtures were then loaded onto a 2% agarose gel in 50% TBE along with a 1 Kb+ ladder, subject to electrophoresis, and photographed under ultraviolet light. The same procedure for preparation of the reaction mixtures and running of the gel was repeated using the coding DNA for CAA and CAB.

CHAPTER III

RESULTS

The BLAST search for Brassicaceae homologues yielded MYB2 from *Brassica oleracea* var. *botrytis* cultivar *Stovepipe* as the closest match to *Arabidopsis* gene AT1G66370 with 95% query coverage, 69% identity, an expect value of 7.00E-110, and an overall score of 327. Other close hits, in order of decreasing score, were MYB2 in *Brassica oleracea* var. *botrytis* cultivar *Graffiti*, PAP1 in *Brassica rapa* subsp. *Rapa* cultivar *Tsuda*, MYB 1 in *Brassica oleracea* var. *botrytis* cultivar *Stovepipe*, and MYB 4 in *Brassica oleracea* var. *botrytis* cultivar *Stovepipe* (Table 1). When the most closely related *Brassica* gene (MYB2 from *Stovepipe*) was used as the query in a BLAST of the CAA transcriptome done in CLC, two loci were reported as matches. These were designated as putative locus 2881, and putative locus 35997. The BLAST output is summarized in Table 2. Locus 35997 was isolated as the locus of interest, specifically two transcripts, 2 and 4. A reciprocal BLAST search using transcript 4 back to *Arabidopsis thaliana* yielded gene MYB114 (AT1G66380) as the top hit, making 35997 a likely candidate for involvement in pigment production (Table 3). In CAB, locus 14598 was found to be a secondary candidate.

Comparison of protein sequences revealed locus 35997 to be highly similar to *Arabidopsis* MYB114 with only six differences in amino acid sequence. The major distinction was that 35997 seemed to have a premature stop codon that makes it roughly two thirds the length of MYB114. Locus 14598 in *barbarae* shows the same similarity until the end of the gene where similarity decreases dramatically.

When the primers were used to amplify genomic DNA from CAA, imaging from the gel revealed that little to no amplification occurred (Fig. 1A). It was then theorized that since the primers were designed using CDS, they may have been created to where they span an intron, thus inhibiting proper polymerase binding and preventing amplification. To correct for this, the same primers were used in PCR of coding DNA. The results from the gel loaded with these reaction mixtures showed bands for each mixture except the negative controls (Fig 1B). No F_1 DNA was available to use in this gel.

CHAPTER IV

CONCLUSIONS

The discovery of a stop codon early in the CAA gene was the opposite of what was expected. Under the assumption that the white CAB is a mutant in which pigment production is inhibited, it was predicted that the gene for a flower color transcription factor may have been mutated in CAB. The results of this study do not support this hypothesis, as the candidate transcription factor gene in CAA is much shorter than the gene in CAB. However, the hypothesis cannot be refuted either, as CAB also shows consensus with the equally long MYB113 amino acid sequence up to a certain point, after which the strength of the alignment greatly weakens. This could possibly be evidence of a mutation causing a frame shift, which is also known to result in a nonfunctional protein.

Though it is unknown how genes such as MYB114 and MYB113 function, these results seem to suggest repressor activity. The shorter length of the *amplexicaulis* gene would hint that the resulting protein would be unable to function. White coloration is typically the result of an inhibited pigment production pathway. However, the CAA flower is purple, indicating that perhaps the gene at 35997 acts as a repressor, not an activator. This is consistent with the observation that the longer, and therefore supposedly functional, genes are found in *Arabidopsis* and *barbarae* which both have white flowers. This new hypothesis that the flower color genes act as repressors has yet to be tested. The fact that one CAB and one CAA sequence was found to be similar to all three transcription factors in *Arabidopsis* (AT1G66370, AT1G66380, and AT1G66390) suggests that a duplication of the gene occurred in the lineage leading to

Arabidopsis after divergence from the lineage leading to *Caulanthus*. Further comparison of additional sequences from the orthologous genes in the Brassicaceae family is needed to confirm such an event.

The PCR results were also not according to expectations. Different combinations of primers were designed to be CAA or CAB-specific but gel imaging revealed that regardless of which primers were used, sequence from both variants had been amplified. One possibility is that the transcriptomes used to design the primers were not entirely accurate. An updated and improved assembly of the CAA and CAB transcriptomes is in preparation, and the CAB genome is being sequenced by the Joint Genome Institute of the Department of Energy (website: www.jgi.doe.gov). These new data sets are being searched to refine our gene models of the candidate transcription factors in *Caulanthus*.

REFERENCES

Bjarne Knudsen, Thomas Knudsen, Mikael Flensborg, Henrik Sandmann, Michael Heltzen, Alex Andersen, Mikkel Dickenson, Jakob Bardram, Peter J. Steffensen, Søren Mønsted, Torben Lauritzen, Roald Forsberg, Agnes Thanbichler, Jannick D. Bendtsen, Lasse Görlitz, Jane Rasmussen, David Tordrup, Morten Værum, Mikkel Nygaard Ravn, Christian Hachenberg, Esben Fisker, Patrick Dekker, Jacob de Meza, Anne-Mette K. Hein, Jonas Buur Sinding, Jesper Quorning, Kenneth Hvam, Steffen Mikkelsen, Poul Liboriussen, Jacob Grydholt, Henrik Handberg, Martin Bundgaard, Anika Joecker, Martin Simonsen, Peder R. L. Nielsen, Andreas Joecker, Paul Fleischer, Jesper Jakobsen, Søren Juul, Uwe Appelt, Anthony Fejes, Aske Simon Christensen (2012) CLC Genomics Workbench v5.5.1, Available from www.clcbio.com

Burrell, A. M., Taylor, K. G., Williams, R. J., Cantrell, R. T., Menz, M. A. and Pepper, A. E. (2011), A comparative genomic map for *Caulanthus amplexicaulis* and related species (Brassicaceae). Molecular Ecology, 20: 784–798. doi: 10.1111/j.1365-294X.2010.04981.x

Burrell, A. M., A. K. Hawkins, and Pepper, A. E. (2012), Genetic Analyses of Nickel Tolerance in a North American Serpentine Endemic Plant, *Caulanthus Amplexicaulis* var. *Barbarae* (Brassicaceae). *American Journal of Botany*, 99.11: 1-9.

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Heled J, Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Swidan F, Thierer T, Wilson A (2012) Geneious v5.6, Available from http://www.geneious.com

http://blast.ncbi.nlm.nih.gov/Blast.cgi

Sequencher® version 4.8 sequence analysis software, Gene Codes Corporation, Ann Arbor, MI USA http://www.genecodes.com

The Arabidopsis Information Resource (TAIR), [http://www.arabidopsis.org/servlets/TairObject?id=137234&type=locus], on www.arabidopsis.org, [September 14, 2012].

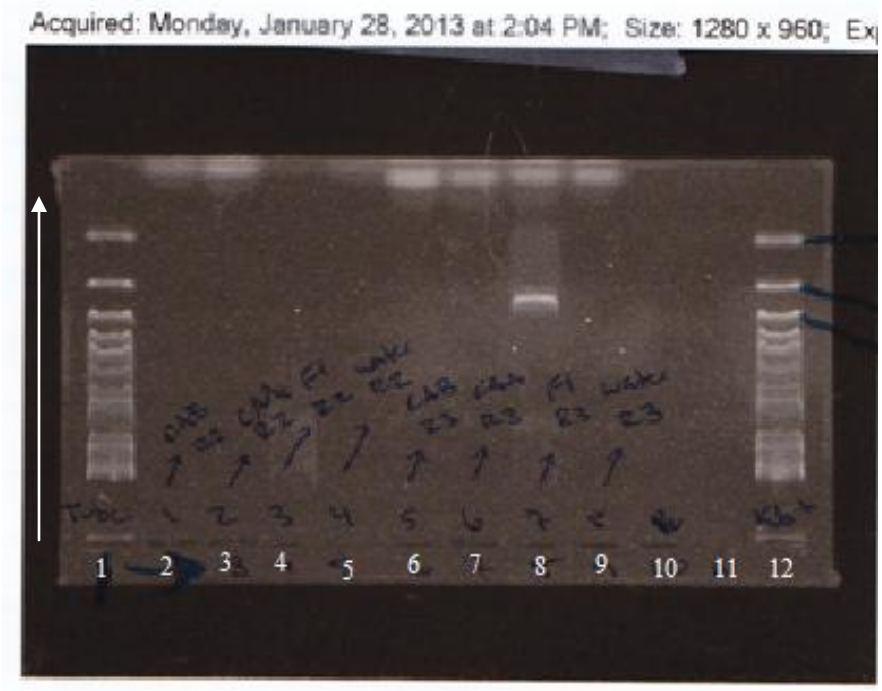
The Arabidopsis Information Resource (TAIR), [http://www.arabidopsis.org/servlets/TairObject?id=137236&type=locus], on www.arabidopsis.org, [September 14, 2012].

The Arabidopsis Information Resource (TAIR), [http://www.arabidopsis.org/servlets/TairObject?id=137239&type=locus], on www.arabidopsis.org, [September 14, 2012].

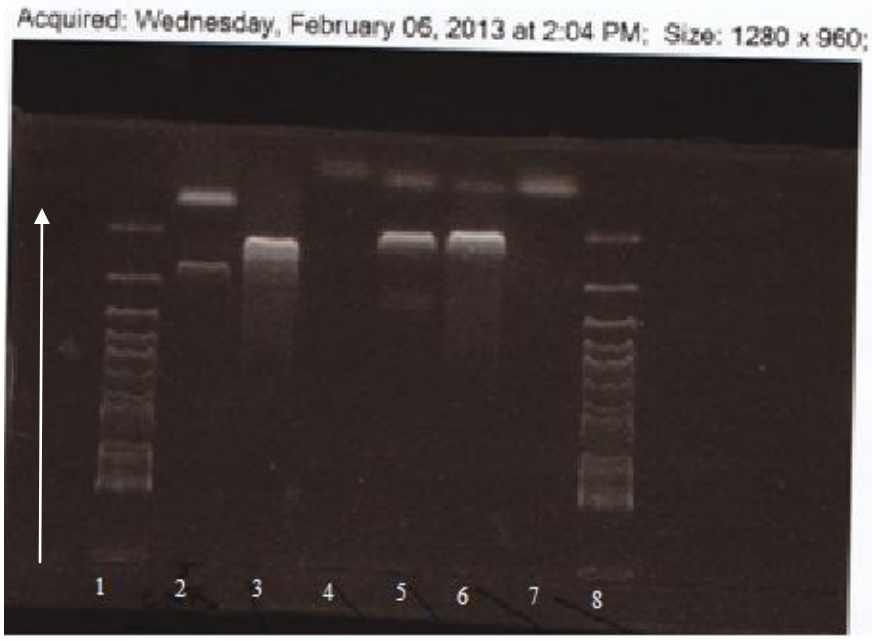
www.jgi.doe.gov

Figures:

Figure 1. Gel Imaging of Agarose Gels



A



B

Figure 1. Gels loaded with genomic DNA (A) and coding DNA (B) after amplification by PCR using loading protocol in Table 4. Sizes of ladder fragments start at 100 and increase in increments of 100 from top to bottom.

Tables:

Table 1. Top 5 NCBI blastx results for query AT1G66370.

Hit #	Organism	Gene	Score	Query coverage	E value	% identity
1	*Brassica oleracea* var. *botrytis* cultivar *Stovepipe*	MYB2	327	95%	7.00E-110	69%
2	*Brassica oleracea* var. *botrytis* cultivar *Graffiti*	MYB2	327	95%	1.00E-109	69%
3	*Brassica rapa* subsp. *Rapa* cultivar *Tsuda*	PAP1	322	99%	1.00E-107	67%
4	*Brassica oleracea* var. *botrytis* cultivar *Stovepipe*	MYB1	319	99%	1.00E-106	69%
5	*Brassica oleracea* var. *botrytis* cultivar *Stovepipe*	MYB4	295	97%	4.00E-97	65%

Table 2. Loci obtained for BLAST of *Brassica oleracea* var. *botrytis* cultivar *Stovepipe* gene MYB 2 against CAA database.

Locus	Transcript	confidence
2881	4/7	0
2881	6/7	0
2881	7/7	0
35997	2/6	1
35997	4/6	2

Table 3. Reciprocal BLAST of locus 35997 transcript 4/6 back to *Arabidopsis*.

Hit #	Organism	Gene	Score	Query coverage	E value	% identity
1	*Arabidopsis thaliana*	MYB114	139	52%	5.00E-39	93%
2	*Arabidopsis thaliana*	unknown	139	52%	6.00E-39	93%
3	*Arabidopsis lyrata*	MYB transcription factor	142	52%	9.00E-39	92%
4	*Arabidopsis lyrata*	Hypothetical protein	141	52%	2.00E-38	91%
5	*Arabidopsis thaliana*	Putative transcription factor	140	52%	2.00E-38	91%
6	*Arabidopsis thaliana*	MYB75	140	52%	3.00E-38	91%
7	*Arabidopsis thaliana*	MYB transcription factor	140	52%	3.00E-38	91%
8	*Arabidopsis thaliana*	MYB75	139	52%	7.00E-38	90%
9	*Arabidopsis thaliana*	MYB75	139	52%	9.00E-38	90%
10	*Arabidopsis thaliana*	MYB90	139	52%	9.00E-38	92%
11	*Brassica oleracea* var. *botrytis* cultivar *Graffiti*	MYB2	139	52%	1.00E-37	90%
12	*Brassica rapa*	PAP1	138	52%	2.00E-37	88%
13	*Brassica oleracea* var. *botrytis*	MYB3	133	52%	5.00E-37	88%
14	*Brassica oleracea* var. *botrytis* cultivar *Stovepipe*	MYB2	137	52%	7.00E-37	88%
15	*Brassica oleracea* var. *botrytis*	MYB1	137	52%	8.00E-37	91%
16	*Arabidopsis thaliana*	MYB75	136	52%	9.00E-37	90%
17	*Brassica oleracea* var. *botrytis*	MYB4	134	52%	1.00E-35	88%

Table 4. Loading protocol for Fig. 1

Lane in Fig. 1A	Lane in Fig. 1B	PCR Contents
1	1	1Kb+ ladder
2	2	CAB DNA FC2-F FC2-R2
3	3	CAA DNA FC2-F FC2-R2
4		F_1 DNA FC2-F FC2-R2
5	4	Water FC2-F FC2-R2
6	5	CAB DNA FC2-F FC2-R3
7	6	CAA DNA FC2-F FC2-R3
8		F_1 DNA FC2-F FC2-R3
9	7	Water FC2-F FC2-R3
10		
11		
12	8	1 Kb+ ladder