August 2015

Genetic Analysis of a Non-Germinating Mutant of Arabidopsis Thaliana

Md Jakir Hossan
The University of Western Ontario

Supervisor
Dr. Yuhai Cui
The University of Western Ontario

Graduate Program in Biology

A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science

© Md Jakir Hossan 2015

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Agricultural Science Commons, Molecular Genetics Commons, and the Plant Biology Commons

Recommended Citation
https://ir.lib.uwo.ca/etd/3062

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact tadam@uwo.ca, wlsadmin@uwo.ca.
GENETIC ANALYSIS OF A NON-GERMINATING MUTANT OF ARABIDOPSIS
THALIANA

(Thesis format: Monograph)

by

MD JAKIR HOSSAN

Graduate Program in Biology

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

© Md Jakir Hossan 2015
ABSTRACT

Seed germination is partially controlled by plant hormone gibberellins (GAs). Chemical mutagenesis yielded an *Arabidopsis thaliana* mutant *gm11*, which has an absolute gibberellin requirement for seed germination. This mutant exhibited phenotypes of GA-rescuable dwarfs, including dark-green leaves, and reduced fertility. However, with repeated GA treatment, *gm11* develops into fertile plants with a nearly wild type phenotype. Bulked-segregant analysis mapped *gm11* to the bottom arm of chromosome 1, and subsequent next-generation mapping revealed that the mutation is a G → A transition in *At1g79460* (*GA2*), creating a premature stop codon. This gene encodes an ent-kaurene synthase (KS) which catalyzes the conversion of copalyl diphosphate to ent-kaurene in the GA biosynthesis pathway. Further genetic analysis suggests that *gm11* is allelic to *ga2* and has been named *ga2-11*. This work demonstrated a genetic finding useful for further understanding the molecular process underlying dormancy and germination.

Keywords:

*Arabidopsis*, Gibberellin, Seed germination, Dormancy, Next-generation mapping, *GA2*
ACKNOWLEDGMENTS

I express my gratitude to my supervisor, Dr. Yuhai Cui, for his invaluable mentorship and for providing me the excellent opportunity to work with him. Thank you for your advice throughout the past two years, and especially for your tremendous encouragement. You are a source of inspiration.

I am thankful to my co-supervisor Dr. Susanne Kohalmi for her valuable comments, and sound advice throughout my study. I would like to extend my appreciation to the members of my committee, Dr. David Smith and Dr. Sangeeta Dhaubhadel for their feedback and guidance.

I would especially like to credit Behnaz Saatian for involving me in this project and for her sharing ideas, patience, and enthusiasm for science. I want to thank Dr. Gang (Gary) Tian for materials and ideas of work. I would also like to thank all my lab-mates: Dr. Chenlong Li, Behnaz Saatian, Chen Chen, Jie Shu, Fu-Yu Hung for valuable discussion and wonderful working atmosphere. I am thankful to Nguyen Vi for helping with proofreading my thesis. I am also thankful to Dr. Ryan Austin for Next generation mapping services.

Last but not least, I would like to extend my gratitude to my beloved parents whose enthusiasm and sacrifice brought me here. Finally, I would like to express my appreciation to my Almighty for His continuous help to finish this work and preparation of this dissertation.
# TABLE OF CONTENTS

ABSTRACT ............................................................................................................................... ii

ACKNOWLEDGMENTS ........................................................................................................... iii

TABLE OF CONTENTS ........................................................................................................ iv

LIST OF TABLES .................................................................................................................... vii

LIST OF FIGURES .................................................................................................................. viii

LIST OF ABBREVIATIONS ..................................................................................................... ix

CHAPTER 1: INTRODUCTION ................................................................................................. 1

1.1 Overview .......................................................................................................................... 1

1.2 Model Plant *Arabidopsis thaliana* ................................................................................. 2

1.3 Seed Development ........................................................................................................... 2

1.4 Seed Dormancy and Germination .................................................................................. 5

1.5 Gibberellin ....................................................................................................................... 9

1.6 Seed Germination Mutants ............................................................................................. 10

1.7 The Role of GA Mutants in Genetics and Plant Breeding .............................................. 11

1.8 Biosynthetic Pathway of GAs ....................................................................................... 12

1.9 Genes Encoding Enzymes for GA Biosynthesis and Signaling ..................................... 15

1.10 Positional Mapping Vs. Next Generation Mapping ....................................................... 17

1.11 Hypothesis and Objectives ........................................................................................... 18

CHAPTER 2: MATERIALS AND METHODS ........................................................................... 19

2.1 Plant Materials ............................................................................................................... 19

2.2 Germination of Seeds and Plant Growth Conditions ..................................................... 19

2.2.1 Sterilization of Seeds and Preparation of MS Media ................................................. 19

2.2.2 Preparation of Gibberellic Acid Supplemented Growth Media .............................. 20

2.2.3 Growth Conditions ...................................................................................................... 20
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3 Mutant Isolation</td>
<td>20</td>
</tr>
<tr>
<td>2.4 Crossing of <em>Arabidopsis</em></td>
<td>21</td>
</tr>
<tr>
<td>2.5 Generation of F2 Mapping Population</td>
<td>24</td>
</tr>
<tr>
<td>2.6 Plant Genomic DNA extraction</td>
<td>24</td>
</tr>
<tr>
<td>2.7 Checking the Quality of Isolated DNA</td>
<td>27</td>
</tr>
<tr>
<td>2.8 Bulked DNA Preparation for Bulked-segregant Analysis (BSA)</td>
<td>27</td>
</tr>
<tr>
<td>2.9 Bulked DNA Preparation for Next-generation Mapping</td>
<td>28</td>
</tr>
<tr>
<td>2.10 Standard PCR reaction</td>
<td>29</td>
</tr>
<tr>
<td>2.11 DNA Gel Electrophoresis</td>
<td>29</td>
</tr>
<tr>
<td>2.12 Genetic and Physical Mapping of Mutants</td>
<td>30</td>
</tr>
<tr>
<td>2.13 Mapping of Mutant Loci with Next-generation Sequencing</td>
<td>30</td>
</tr>
<tr>
<td>2.14 Sequencing of Candidate Genes</td>
<td>32</td>
</tr>
<tr>
<td>2.15 Screening of Homozygous or Heterozygous T-DNA Mutants</td>
<td>32</td>
</tr>
<tr>
<td>2.16 Confirmation of Candidate Gene (Allelism Test)</td>
<td>33</td>
</tr>
<tr>
<td>2.17 Statistical Analysis</td>
<td>33</td>
</tr>
<tr>
<td>CHAPTER 3: RESULTS</td>
<td>35</td>
</tr>
<tr>
<td>3.1 Screening of F2 Mapping Population</td>
<td>35</td>
</tr>
<tr>
<td>3.2 Bulked-Segregant Analysis (BSA) to Locate the Mutation Site in</td>
<td>35</td>
</tr>
<tr>
<td>Chromosome</td>
<td></td>
</tr>
<tr>
<td>3.3 Next-generation Mapping by Genome Sequencing</td>
<td>38</td>
</tr>
<tr>
<td>3.4 Mutation is Localized in <em>GA₂</em> Gene</td>
<td>46</td>
</tr>
<tr>
<td>3.4.1 Identification of Candidate Genes</td>
<td>46</td>
</tr>
<tr>
<td>3.4.2 Genotyping by DNA Sequencing of the Candidate Gene (<em>ga2</em>) in All</td>
<td>47</td>
</tr>
<tr>
<td>F2 Individuals Used for NGM</td>
<td></td>
</tr>
<tr>
<td>3.4.3 <em>sdg8-2</em> Background is Not Involved in <em>gm11</em> Phenotype</td>
<td>51</td>
</tr>
<tr>
<td>3.5 Genetic Evidence Confirming that Mutation in <em>ga2</em> is the Cause</td>
<td>52</td>
</tr>
<tr>
<td>of the Germination Mutant Phenotype of <em>gm11</em></td>
<td></td>
</tr>
</tbody>
</table>
3.5.1 \textit{gm11} is Allelic to \textit{ga2}........................................................................52

3.5.2 Allelism Test...............................................................................................57

3.6 Application of Exogenous GA Rescues Phenotypes of \textit{ga2} Mutant........62

CHAPTER 4: DISCUSSION AND CONCLUSION..................................................65

REFERENCES....................................................................................................72

APPENDIX I. License Agreement for Figure One and Two.............................82

APPENDIX II. License Agreement for Figure Five.........................................84

CURRICULUM VITAE.........................................................................................86
LIST OF TABLES

Table 1. List of identified GA-biosynthesis and signaling genes ........................................ 16

Table 2. List of SSLP primers used for genetic mapping in this study ................................. 31

Table 3. List of other primers used for this study ............................................................... 34

Table 4. Segregation of F2 progeny derived from different crosses ................................. 41

Table 5. Seven identified SNPs at the bottom of chromosome 1 ........................................ 48

Table 6. Germination percentage of T-DNA mutant seeds ............................................... 60

Table 7. Segregation of F1 progeny derived from heterozygous gm11 and SALK_017607C cross .................................................................................................................. 61
LIST OF FIGURES

Figure 1. Arabidopsis seed germination ................................................................. 3

Figure 2. Schematic representation of the process controlling seed dormancy and germination in an Arabidopsis seed ................................................................. 7

Figure 3. GA mutants in genetics and plant breeding ............................................. 13

Figure 4. Screening of GA sensitive mutants ......................................................... 22

Figure 5. Schematic illustration of a strategy of next-generation sequencing-based mapping and identification of EMS mutations ......................................................... 25

Figure 6. The Morphological phenotypes of mutants and wild-type ..................... 36

Figure 7. Bulked-segregant analysis of pooled F2 mutant seedlings ..................... 39

Figure 8. Next generation EMS mutation mapping web application ..................... 42

Figure 9. Genome-wide SNP variation patterns plotted by histograms for each chromosome. .................................................................................................................. 44

Figure 10. DNA sequencing result of AT1G79460 gene for F2 mutants ............... 49

Figure 11. Agarose gel electrophoresis of PCR products demonstrating the genotype of the sdg8-2 locus ......................................................................................... 53

Figure 12. Analysis of T-DNA insertion lines for validation of mutation in ga2 gene .... 55

Figure 13. PCR and sequencing genotyping for identifying heterozygous or homozygous T-DNA and gm11 mutants ......................................................................................... 58

Figure 14. Phenotype of germination mutants in comparison with wild-type .......... 63
LIST OF ABBREVIATIONS

Note: International System of Units (SI) are not listed

3’            three prime ends of a DNA fragment
5’            five prime ends of a DNA fragment
A             adenine
aa            amino acids
AAFC          Agriculture and Agri-Food Canada
ABA           abscisic acid
ABRC          Arabidopsis Biological Resource Center
BAC           bacterial artificial clone
bp            base pair (s)
BR            brassinosteroids
BSA           Bulked-segregant analysis
Col           Columbia
CPP           copalyl pyrophosphate
CPS           ent-copalyl synthase
CTAB          cetyltrimethyl-ammonium bromide
d             days
DNA           deoxyribonucleic acid
EDTA          ethylenediaminetetraacetic acid
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMS</td>
<td>ethyl methane sulfonate</td>
</tr>
<tr>
<td>F1, F2</td>
<td>first, second filial generation after a cross</td>
</tr>
<tr>
<td>G</td>
<td>guanine</td>
</tr>
<tr>
<td>GA</td>
<td>gibberellin</td>
</tr>
<tr>
<td>GA3</td>
<td>gibberellic acid</td>
</tr>
<tr>
<td>GGDP</td>
<td>geranylgeranyl diphosphate</td>
</tr>
<tr>
<td>Het</td>
<td>heterozygous</td>
</tr>
<tr>
<td>Homo</td>
<td>homozygous</td>
</tr>
<tr>
<td>HS</td>
<td>high sensitivity</td>
</tr>
<tr>
<td>InDel</td>
<td>insertions/deletions</td>
</tr>
<tr>
<td>IRRI</td>
<td>International Rice Research Institute</td>
</tr>
<tr>
<td>kb</td>
<td>kilo base pair(s)</td>
</tr>
<tr>
<td>KAO</td>
<td>entkaurenoic acid oxidase</td>
</tr>
<tr>
<td>KO</td>
<td>ent-kaurene oxidase</td>
</tr>
<tr>
<td>KS</td>
<td>ent-kaurene synthase</td>
</tr>
<tr>
<td>LB</td>
<td>Luria Bertani</td>
</tr>
<tr>
<td>Ler</td>
<td>Ler <em>erecta</em></td>
</tr>
<tr>
<td>M1, M2, M3</td>
<td>first, second, third filial generation after a mutagenesis treatment</td>
</tr>
<tr>
<td>Mb</td>
<td>mega base pair (s)</td>
</tr>
<tr>
<td>MES</td>
<td>2-N-morpholino-ethanesulfonic acid</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>MS</td>
<td>Murashige and Skoog</td>
</tr>
<tr>
<td>Mut</td>
<td>mutant</td>
</tr>
<tr>
<td>NGM</td>
<td>next-generation mapping</td>
</tr>
<tr>
<td>NGS</td>
<td>next-generation sequencing</td>
</tr>
<tr>
<td>NonSyn</td>
<td>nonsynonymous</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>pH</td>
<td>negative decimal logarithm of $H^+$ concentration</td>
</tr>
<tr>
<td>PPR</td>
<td>pentatricopeptide repeat</td>
</tr>
<tr>
<td>RB</td>
<td>T-DNA right border</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>SCPFRC</td>
<td>Southern Crop Protection and Food Research Centre</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>SNP</td>
<td>single-nucleotide polymorphism</td>
</tr>
<tr>
<td>SSLP</td>
<td>simple sequence length polymorphism</td>
</tr>
<tr>
<td>Syn</td>
<td>synonymous</td>
</tr>
<tr>
<td>TAIR</td>
<td>The <em>Arabidopsis</em> Information Resource</td>
</tr>
<tr>
<td>TAE</td>
<td>tris acetic acid EDTA</td>
</tr>
<tr>
<td>Taq</td>
<td><em>Thermus aquaticus</em></td>
</tr>
<tr>
<td>T-DNA</td>
<td>transfer DNA</td>
</tr>
<tr>
<td>Tris</td>
<td>tris (hydroxymethyl)-aminomethane</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>WGS</td>
<td>whole genome sequencing</td>
</tr>
<tr>
<td>WT</td>
<td>wild-type</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet light</td>
</tr>
</tbody>
</table>
CHAPTER 1: INTRODUCTION

1.1 Overview

Seed germination is a critical stage in the life cycle of higher plants, with regards to next-generation plant growth and their overall survival. In nature, many seeds become dormant for a period to optimize germination (Bentsink and Koornneef, 2008). Seed dormancy needs to be broken in order for seed to germinate. A wide variety of environmental factors such as light, temperature, duration of dry storage and genetic factors affect seed germination and dormancy (Donohue et al., 2005; Huang et al., 2010). In the past decades, research has been conducted to improve our understanding of how genetic factors and hormones trigger seed germination. Arabidopsis thaliana is one appropriate genetic model system that can be used to investigate seed germination mutants. Classical and molecular genetic studies in Arabidopsis have uncovered some of the mechanisms of seed dormancy and germination (Graeber et al., 2012). Numerous mutants that exhibit enhanced seed dormancy or reduced seed germination have mutations in the genes involved in the biosynthesis or signaling pathways of plant hormones such as gibberellin (GA), brassinosteroids and ethylene (Koornneef et al., 2002). In particular, GA plays a significant role in seed germination and other aspects of plant growth and development such as, stem enlargement, flower induction, fruit development, and seed maturation (Matsuoka, 2003). Genes that control seed dormancy and germination are mostly associated with the GA metabolic pathway (Koornneef et al., 2002). Despite extensive research efforts devoted to analyzing the GA metabolic pathway, only a few genes from have been identified. Knowledge gaps, however, still
exist in terms of the molecular mechanisms that underlie dormancy and germination in *Arabidopsis*. The purpose of this research was to characterize a recessive GA sensitive mutant of *Arabidopsis*, to improve our understanding of the role of GA signaling pathway in seed germination.

1.2  Model Plant *Arabidopsis thaliana*

*Arabidopsis* is a dicotyledonous small flowering plant that is related to cabbage and mustards. Over the years, it has been widely accepted by the plant biology scientific community as a model organism for hereditary, molecular, biochemical and physiological studies (Meinke et al., 1998; Meyerowitz, 2002). It has the smallest genome among the higher plants (its genome is about 125 Mb in size). One hundred fifteen Mb of the genome, comprised of approximately 26000 genes, was sequenced and annotated late in 2000 (The *Arabidopsis* Initiative, 2000). Also, the efficiency of transformation is particularly high in *Arabidopsis* utilizing *Agrobacterium tumefaciens* (Zhang et al., 2006). Importantly, a considerable number of mutant lines and genomic resources are available publicly, offering an important tool for research on studying gene expression and regulation.

1.3  Seed Development

To understand the biology of seed germination (Figure 1), it is necessary to comprehend the seed-development process. Seed-development includes two crucial phases: embryo
Figure 1. *Arabidopsis* seed germination

Seed development consists of two major phases: embryo development and seed maturation. After imbibition, the seed turns to radicle protrusion and germination occurs. A mature *Arabidopsis* seed can be either dormant or non-dormant. A dormant seed will not germinate even when exposed to the right environmental conditions, such as light and water. However, seed dormancy can be broken by dry storage or a cold imbibition (stratification). Seed germination in *Arabidopsis* is a two-step process: testa rupture followed by endosperm rupture. Adapted from Bentsink and Koornneef (2008).
development (morphogenesis) and seed maturation. Embryogenesis begins when a single-cell zygote is formed then it goes to the heart stage while embryo’s basic structures like an apical meristem, plumule, the radicle, hypocotyl and the cotyledons are established (Mayer et al., 1991) followed by the growth phase while the embryo occupies the seed sac (Goldberg et al., 1994). After the growth phase, cell division in the embryo comes to a halt (Raz et al., 2001). Once the full-sized embryo is formed, seed maturation begins, and storage reserves are synthesized and accumulated, followed by a desiccation process and eventually dormancy achieved (Goldberg et al., 1994).

### 1.4 Seed Dormancy and Germination

Seed dormancy is a mechanism that delays germination until the conditions are optimal to secure the next generation survival and regeneration, which helps the plant to adapt in different habitats (Finkelstein et al., 2008). Thus, seed dormancy is an important mechanism for plant fitness (Huang et al., 2010). In contrast, a lower level of seed dormancy can decrease the quality of seeds that may lead to pre-harvest sprouting (Gubler et al., 2005). Dormancy is a complicated trait, controlled by both genetic and environmental factors. Recent research revealed that the plant hormones abscisic acid (ABA) and gibberellin play a critical role in seed dormancy regulation (Ye and Zhang, 2012). Additional factors that have an impact on dormancy include dormancy-specific genes such as *DELAY OF GERMINATION1*, chromatin factors and non-enzymatic processes (Graeber et al., 2010). Seed germination is an important aspect of higher plants life cycles, as an ecological and agronomic trait that helps to control when plants enter
ecosystems (Graeber et al., 2010). Utilizing *Arabidopsis* as a model system, my research focuses on studying gene(s) that are responsible for seed germination and dormancy, using ethyl methane sulfonate (EMS) induced germination mutants.

Both seed dormancy and germination depend on a balance between growth hormones, environmental stimuli and structural factors surrounding the embryo such as seed coat factors, testa and endosperm (Mayer et al., 1991) (Figure 2). GA plays a significant role in the promotion of seed germination. GA-deficient mutants cannot germinate without exogenous GA (Mitchum et al., 2006). Previous research reported that content and the expression of GA biosynthesis and catabolism genes regulate dormancy release and seed germination (Yamauchi et al., 2007). GA and ABA are the two main hormones that play central and antagonistic roles in seed germination. When GA content increases, there is consistently associated ABA content reduction during seed germination (Holdsworth et al., 2008; Nambara et al., 2010). Genetic analysis has demonstrated that ABA plays a significant role in seed dormancy (Bonetta and McCourt, 1998). Environmental factors regulate the germination of seeds by influencing the signaling pathways of either GA or ABA (Heschel et al., 2007; Seo et al., 2009). Two such environmental factors, light and cold increase GA biosynthesis, which then improves the embryo’s potential to trigger seed germination (Yamauchi et al., 2004). Furthermore, it is evident that phytochrome-mediated pathways are also predominant in breaking the cold-induced dormancy. Mutants lacking phytochrome B (phyB) exhibit a decreased sensitivity to red light, indicating that phyB has a role in seed germination (Donohue et al., 2007). Brassinosteroids (BR), a plant steroid hormone, also plays a role in the management of seed germination by reversing dormancy, induced by ABA. The BR signal decreases the
Figure 2. Schematic representation of the process controlling seed dormancy and germination in an *Arabidopsis* seed

The *Arabidopsis* seed consists of an embryo with two cotyledons and a single cell layer of endosperm. Green arrows indicate germination promoting factors while the red arrows indicate inhibiting factors. Light and cold act as environmental stimuli. Structural factors surrounding the embryos are testa, endosperm, and seed coat factors. Abscisic Acid (ABA), Gibberellin (GA), Brassinosteroids (BR), Nitric Oxide (NO), Auxin Response Factor 10 (ARF10) and microRNA 160 (miR160) are the factors for embryo growth potential. Adapted from Bentsink and Koornneef, (2008).
sensitivity of ABA that stimulates germination, but BR deficient and signaling mutants suggest that BR is not absolutely required for germination (Steber and McCourt, 2001). Mutants in ethylene signaling also impact the seed germination process. Higher plants produce trace quantities of ethylene, which helps control the growth and developmental processes that range from germination to senescence (Kępczyński and Kępczyńska, 1997). Mutants with an altered seed layer or testa also exhibit decreased seed dormancy (Debeaujon and Koornneef, 2000).

1.5 Gibberellin

GA was first found through research on the bakanae disease in rice. Giberella fujikuroi is the causal organism of the bakane disease which causes excessive elongation of infected plants and consequently no seed production (Nirenberg and O’Donnell, 1998). After culturing the Giberella fungus in the laboratory, Japanese scientists in 1930s identified a fungal compound that processes growth promoting activity in plants which they named Gibberellin (Marciniak et al., 2012). Different forms of GAs have been identified from fungi, bacteria and plants and the most bioactive GAs are GA$_1$, GA$_3$, GA$_4$ and GA$_7$ (Yamaguchi, 2008). The gibberellins play diverse roles in plant growth and development (Kucera et al., 2005).
1.6 Seed Germination Mutants

Many researchers have employed the forward genetic screens approach to identify mutants defective in dormancy and germination (Nonogaki, 2014). Since plant hormones are important for the regulation of plant growth and development, isolating plant hormone deficient mutants is useful for the fields of plant genetics and plant physiology. Studies concerning dwarf mutants, including those that analyzed GA content in wild-type and mutant plants, have demonstrated that bioactive GAs exist as endogenous plant hormones, which regulate various developmental processes. In the Arabidopsis seeds, GA’s main role is to promote the seed coat’s breakage, for which de novo GA biosynthesis is necessary. The seeds of numerous GA deficient mutants, including ga1, ga2 and ga3 single mutants, as well as the double homozygous mutants ga3ox1 ga3ox2, do not germinate in darkness and have just ~5% germination success in the light (Mitchum et al., 2006). Nonetheless, when these mutants’ seed coats are harmed mechanically, the mutant embryos are still able to germinate and grow into mature plants (Telfer et al., 1997; Griffiths et al., 2006). Altered seed coat structures in Arabidopsis and tomato plants could rescue the defective seed germination in GA biosynthesis mutants. This phenomenon provides evidence that GA is necessary to overcome the seed coat’s physical inhibitory effects on seed germination and ABA-associated embryo dormancy (Groot and Karssen, 1987; Debeaujon and Koornneef, 2000).
1.7 The Role of GA Mutants in Genetics and Plant Breeding

Historically, gibberellins have been central to the breeding of crops. Altering the GA pathways was critical for the green revolution, as this has led to semi-dwarfism and greater crop yields (Hedden, 2003; Salamini, 2003). Semi-dwarfism is an important trait because it prevents lodging and, therefore, enables farmers to increase the amount of fertilizer they use and maximize the harvest index (Spielmeyer et al., 2002).

In rice, a semi-dwarf variety carrying a mutation in the Semi-Dwarf-1 (SD1) locus which codes for GA 20-oxidase-2 was first developed (Spielmeyer et al., 2002). About the same time, a dwarfing gene (Rht1) that has its origin in a Chinese cultivar (Dee-geo-woo-gen) was introduced to develop both Indica cultivars and Taichung Native 1 (TN-1) variety (Hedden, 2003). One of the greatest successes was the development of semi-dwarf Indica cultivars (Figure 3 A) and Taichung Native 1 (TN-1) variety as part of a 1950’s breeding program in the tropical zone for ‘green revolution’. Subsequently, the International Rice Research Institute (IRRI) in the Philippines produced IR-8. TN-1 and IR-8 which were then employed by breeding programs as parents, leading to the production of semi-dwarf indica cultivars in tropical as well as semi-tropical area (Hedden, 2003).

Another great achievement during the green revolution is the development and launch of semi-dwarf wheat varieties (Figure 3B) carrying the Rht mutations, which resulted in food grain production being doubled (Börner et al., 1996). The first introduced semi-dwarf wheat variety was NORIN-10 which was established in Japan by introducing this Rht gene. Later in the USA, this variety was used to develop a number of high yielding
semi dwarf varieties that were distributed rapidly in Latin America and South Asia and resulted in spectacular yield increases (Hedden, 2003).

### 1.8 Biosynthetic Pathway of GAs

The biosynthetic pathway of GA in higher plants can be categorized into three distinct stages. Biosynthesis starts with conversion of ent-kaurene from geranylgeranyl diphosphate (GGDP) in proplastids. ent-kaurene is then sequentially oxidized to GA$_{12}$ via cytochrome P450 monooxygenases, and finally form C20- and C19-GAs in the cytoplasm (Hedden and Phillips, 2000). The cyclization of GGDP to ent-kaurene is the first committed intermediate in the GA biosynthesis pathway (Sun and Kamiya, 1994). Two distinctive enzymes are involved in these reactions. The ent-kaurene synthetase A enzyme catalyzes the conversion of GGDP to a copalyl pyrophosphate (CPP), whereas the ent-kaurene synthetase B helps the further conversion of CPP to ent-kaurene. Both enzymes have been suggested to be renamed as ent-copalyl diphosphate synthase (CPS) and ent-kaurene synthase (KS), respectively (Sun and Kamiya, 1994; Yamaguchi et al., 1998). The second stage of the GA biosynthesis pathway, starting from a sequential oxidation followed by ring contraction, is catalyzed by ent-kaurene oxidase (KO) and entkaurenoic acid oxidase (KAO) to produce GA$_{12}$, which is the first GA formed in all systems. The subsequent successive oxidation of GA$_{12}$ to GA$_{53}$ by 13-hydroxylation leads to various GA intermediates and bioactive GAs (Sun, 1994).
Figure 3. GA mutants in genetics and plant breeding

A. The tall isogenic line with Semi-dwarf rice cultivars. From left to right: Dee-geo-woo-gen (dwarf *indica* cultivar), woo-gen (tall equivalent). Adapted from Hedden (2003).

B. Mixed varieties of wheat from the mid-1900s were as tall as an ordinary individual. Interestingly, cutting edge wheat research reduced the size of stalks shorter than the stalks of prior cultivars. The semi-dwarf wheat varieties carry mutation in genes encoding proteins that directly regulate gibberellin biosynthesis. Adapted from Salamini (2003).
1.9 Genes Encoding Enzymes for GA Biosynthesis and Signaling

Different techniques have been utilized to study GA biosynthesis and catabolism pathways, including the gas chromatography-mass spectrometry analysis of GA content, purifying GA metabolism enzymes, isolating GA-deficient mutants, and cloning corresponding genes (Hedden and Phillips, 2000; Yamaguchi, 2008). The previous researches in the areas of molecular genetics and biochemistry have assisted in the identification of GA biosynthesis and catabolism genes in *Arabidopsis* as well as in other species (Table 1). A number of GA-deficient dwarf mutants in *Arabidopsis* have been isolated and named *ga1*, *ga2*, *ga3*, *ga4* and *ga5* (Koornneef and Van Der Veen, 1980). After cloning the corresponding genes, it was discovered that *GA1* encodes ent-copalyl synthase (Sun and Kamiya, 1994), *GA2* encodes ent-kaurene synthase (Yamaguchi et al., 1998a), *GA3* encodes ent-kaurene oxidase (Helliwell et al., 2001), *GA4* encodes GA3ox1 (Chiang et al., 1995), and *GA5* encodes GA20ox (Phillips et al., 1995). Null alleles of three early genes in *Arabidopsis* (*ga1*, *ga2* and *ga3*) have been characterized as non-germinating, reduced fertility and severe dwarfs while the *ga4* and *ga5* mutants are semi-dwarf and late flowering (Koornneef and Van Der Veen, 1980).
Table 1. List of identified GA-biosynthesis and signaling genes

<table>
<thead>
<tr>
<th>Locus ID</th>
<th>Name</th>
<th>Protein function</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GA biosynthesis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At4g02780</td>
<td>CPS (GA1)</td>
<td>ent-copalyl synthase</td>
<td>Sun and Kamiya, 1994</td>
</tr>
<tr>
<td>At1g79460</td>
<td>KS (GA2)</td>
<td>ent-kaurene synthase</td>
<td>Yamaguchi et al., 1998</td>
</tr>
<tr>
<td>At5g25900</td>
<td>KO (GA3)</td>
<td>ent-kaurene oxidase</td>
<td>Helliwell et al., 2001</td>
</tr>
<tr>
<td>At1g05160</td>
<td>KAO1</td>
<td>ent-kaurenoic acid oxidase</td>
<td>Helliwell et al., 2001</td>
</tr>
<tr>
<td>At2g32440</td>
<td>KAO2</td>
<td>ent-kaurenoic acid oxidase</td>
<td>Helliwell et al., 2001</td>
</tr>
<tr>
<td>At4g25420</td>
<td>GA20ox1 (GA5)</td>
<td>GA 20-oxidase</td>
<td>Phillips et al., 1995</td>
</tr>
<tr>
<td>At5g51810</td>
<td>GA20ox2</td>
<td>GA 20-oxidase</td>
<td>Phillips et al., 1995</td>
</tr>
<tr>
<td>At5g07200</td>
<td>GA20ox3</td>
<td>GA 20-oxidase</td>
<td>Phillips et al., 1995</td>
</tr>
<tr>
<td>At1G60980</td>
<td>GA20ox4</td>
<td>GA 20-oxidase</td>
<td>Rieu et al., 2008</td>
</tr>
<tr>
<td>At1G44090</td>
<td>GA20ox5</td>
<td>GA 20-oxidase</td>
<td>Chiang et al., 1995</td>
</tr>
<tr>
<td>At1g15550</td>
<td>GA3ox1 (GA4)</td>
<td>GA 3-oxidase</td>
<td>Yamaguchi et al., 1998</td>
</tr>
<tr>
<td>At1g80340</td>
<td>(GA4H)</td>
<td>GA 3-oxidase</td>
<td>Mitchum et al., 2006</td>
</tr>
<tr>
<td>At4g21690</td>
<td>GA3ox3</td>
<td>GA 3-oxidase</td>
<td>Mitchum et al., 2006</td>
</tr>
<tr>
<td>At1g80330</td>
<td>GA3ox4</td>
<td>GA 3-oxidase</td>
<td>Mitchum et al., 2006</td>
</tr>
<tr>
<td><strong>GA signaling</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At3g05120</td>
<td>GID1a</td>
<td>GA receptor</td>
<td>Nakajima et al., 2006</td>
</tr>
<tr>
<td>At3g63010</td>
<td>GID1b</td>
<td>GA receptor</td>
<td>Nakajima et al., 2006</td>
</tr>
<tr>
<td>At5g27320</td>
<td>GID1c</td>
<td>GA receptor</td>
<td>Nakajima et al., 2006</td>
</tr>
<tr>
<td>At2g01570</td>
<td>RGA</td>
<td>GA signaling repressor</td>
<td>Silverstone et al., 1997</td>
</tr>
<tr>
<td>At1g14920</td>
<td>GAI</td>
<td>GA signaling repressor</td>
<td>Peng et al., 1997</td>
</tr>
<tr>
<td>At1g66350</td>
<td>RGL1</td>
<td>GA signaling repressor</td>
<td>Wen and Chang, 2002</td>
</tr>
<tr>
<td>At3g03450</td>
<td>RGL2</td>
<td>GA signaling repressor</td>
<td>Peng et al., 1997</td>
</tr>
<tr>
<td>At5g17490</td>
<td>RGL3</td>
<td>GA signaling repressor</td>
<td>Dill and Sun, 2001</td>
</tr>
<tr>
<td>At3g11540</td>
<td>SPY</td>
<td>GA signaling repressor</td>
<td>Jacobsen and Olszewski, 1993</td>
</tr>
<tr>
<td>At4g24210</td>
<td>SLY1</td>
<td>F-box protein</td>
<td>McGinnis et al., 2003</td>
</tr>
<tr>
<td>At5g48170</td>
<td>SNE</td>
<td>F-box protein</td>
<td>Strader et al., 2004</td>
</tr>
</tbody>
</table>
1.10 Positional Mapping Vs. Next Generation Mapping

Forward genetic screens are a powerful tool to uncover new gene functions in organisms. Although mutant screens can sometimes be performed quickly, the identification of the causative mutation using map-based cloning is highly laborious. Sizable F2 mapping populations with more than 1000 individual mutants are necessary in order to fine-map the part of a chromosome that holds a causative mutation (Jander et al., 2002). This quantity of mutant individuals can be challenging to gather. This is particularly true when dealing with phenotypic traits that (1) are challenging to score, (2) are weakly transmitted, or (3) in organisms that are difficult to propagate. Bulked-segregant analysis is one valuable method for identifying markers that are linked to a mutation (Michelmore et al., 1991). Genetic mapping has traditionally been a laborious and time-consuming task due to the necessity of crossing and phenotyping large populations of F2 plants and their subsequent molecular scoring.

The advent of next-generation sequencing (NGS) technologies have dramatically reduced this genetic mapping effort in a number of model systems, including *Arabidopsis*, by replacing the scoring of molecular markers with whole-genome sequencing (Austin et al., 2014). NGS can be used to substitute map-based cloning by direct sequencing of mutant genomes in order to efficiently identify causative mutations (Austin et al., 2011). Nonetheless, the methods now in use rely on identifying homozygous mutant individuals from a F2 mapping population after outcrossing (Schneeberger et al., 2009) or numerous rounds of backcrossing are necessary (Zuryn et al., 2010). Until now, a number of groups have created important NGS mapping methods for *Arabidopsis*, normally focusing on the
identification of recessive ethyl-methanesulfonate (EMS)-induced mutations (Uchida et al., 2011). To overcome the issue of background noise, most researchers first get an idea of rough mapping of the candidate genes by conducting a bulked-segregant analysis. For instance, one approach in this regard, called SHOREmap, has identified a causative *Arabidopsis* mutation using the Illumina Genome Analyzer (GA) sequencing of a group of 500 pooled F2 lines (Schneeberger et al., 2009).

### 1.11 Hypothesis and Objectives

Although a great deal of information is available concerning seed germination mutants, relatively little research has been conducted to reveal the molecular mechanisms underlying the seed germination mutant phenotype. The goal of my research is to improve our understanding of the seed development biology and the factors that influence seed dormancy and germination. I hypothesize that the germination mutant and its dwarf phenotype are caused by a mutation in gene(s) acting in the GA biosynthesis pathway that promote seed germination in *Arabidopsis*.

The specific objectives of this research are

1. Identification of the gene(s) that cause the dwarf phenotype using next-generation mapping

2. Confirmation and validation of identified gene(s) using genetic approaches
CHAPTER 2: MATERIALS AND METHODS

2.1 Plant Materials

All of the experiments described below were performed at the Southern Crop Protection and Food Research Centre (SCPFRC), Agriculture and Agri-Food Canada, London, Ontario. Wild-type seeds of *Arabidopsis* of ecotype Col, *Ler erecta* and T-DNA insertion lines were obtained from Arabidopsis Biological Resource Centre (ABRC).

2.2 Germination of Seeds and Plant Growth Conditions

2.2.1 Sterilization of Seeds and Preparation of MS Media

For germination, seeds were first surface-sterilized by mixing in 70% ethanol for 30 seconds, followed by 15-min incubation in 20% bleach with 0.1% sodium dodecyl sulfate (SDS) solution, then rinsed three times with sterilized distilled water. The sterilized seeds were then sown on MS medium containing 4.3 g l⁻¹ Murashige and Skoog nutrient mix (Sigma –Aldrich), 3% sucrose, 0.5 g L⁻¹ 2-N-morpholino-ethanesulfonic acid (MES), 0.8% agar and buffered to pH 5.7 with 1 M sodium hydroxide.
2.2.2 Preparation of Gibberellic Acid Supplemented Growth Media

All of the following steps were carried out inside a laminar flow hood. The autoclaved media was allowed to cool in a water bath at 55°C. 1000 μl of 10 mM GA₃ hormone was added per 1L of agar media to yield a 10 μM final concentration. The agar media was gently mixed and poured into Petri Dish. These plates were sealed with parafilm and placed in a cold room at 4°C in the dark for future use.

2.2.3 Growth Conditions

Seeds were vernalized at 4°C for three days before growing directly in pots containing ProMix-BX (Premier Horticulture, Québec) in a growth room. Plants were grown under 16-hour light (22°C) / 8-hour dark (20°C) cycles.

2.3 Mutant Isolation

Seeds of the sdg8-2 T-DNA insertion homozygous mutant lines, from our former lab member Dr. Gang Tian, were identified and subjected to EMS mutagenesis according to standard protocols (Weigel and Glazebrook, 2006). The resulting mutagenized seeds (M1) were directly grown in soil and allowed to set seeds. Seeds from each plant were harvested individually. To isolate non-germinating GA sensitive mutants, individual M2 lines were grown on standard MS medium. After two weeks, the seeds, which had not germinated, were transferred to agar media containing 10 μM GA₃. These dishes were
placed back into the controlled room and allowed to germinate in GA$_3$ media. After
identifying GA sensitive mutant lines, these seedlings were then transferred to soil to
detect any morphological phenotypes caused by the mutation(s) of interest. The resulting
M2 lines, showing the non-germinating dwarf phenotypes were isolated (Figure 4). For
further confirmation, M3 lines were tested for germination behavior and GA sensitivity.
The reason behind using $sdg8$ for EMS mutagenesis is the following. In the past, our lab
conducted an EMS mutant screen for mutants exhibiting ectopic expression of seed
storage protein genes $sdg8$ was allelic to one of the mutant ($essp4$) (Tang et al., 2008;
Tang et al., 2012b; Tang et al., 2012a; Li et al., 2015). The original intention of the
mutant screen in the $sdg8$ background was to identify suppressor of $sdg8$.

### 2.4 Crossing of Arabidopsis

Plants were grown in the growth room until they reached their flowering stage. To
prepare the pistil as a pollen receptor, unopened young flower buds, before the pollen of
the anthers was ripened, were selected. The buds were opened using fine forceps. Sepals,
petals and anthers of receptor flowers were carefully removed. The remaining flowers
and younger buds were also removed from the inflorescence. The stigma of the prepared
carpel was pollinated with mature pollen from recently opened flower buds from the
donor plants. Pollination was repeated after 24 hours to ensure crossing success. The
resulting siliques were later harvested, and seeds were dried in a desiccator at room
temperature. To confirm crossing successes, genotyping was performed using the F1
generation.
Figure 4. Screening of GA sensitive mutants

A) Healthy seedlings on MS media.

B) Non-germinating seed on MS media at 14 days.

C) Germinated seeds on agar media with GA3 (10 µM).

D) Fourteen days Seedlings grown in soil with and without GA. Arrow is indicating mutant and wild-type.

Note: Wild-type (WT); Mutant (Mut)
2.5 Generation of F2 Mapping Population

Since homozygous plants did not form any opened flowers either for pollen donor or receptor, F2 mapping population of GA sensitive mutation(s) was done by crossing heterozygous M2 plants from Col background with *Ler erecta* wild-type plants. F1 seeds, from the resulting crosses, were grown in soil and allowed to self-pollinate and mature (Figure 5). To determine whether GA sensitive mutation is recessive or dominant and if it is controlled by one or more loci, the mutant phenotype ratio was carefully determined.

2.6 Plant Genomic DNA extraction

DNA extraction from *Arabidopsis* plant tissue was done by following the cetyltrimethyl ammonium bromide (CTAB) method (Stewart Jr and Via, 1993). Plant tissue was harvested into 2.0 mL Eppendorf tubes with a copper bead, followed by dipping into liquid nitrogen. The tissues were homogenized, utilizing the Qiagen Tissuelyser, by shaking at 25 rpm for 30 seconds. Then 500 μL of extraction buffer (100 mMTris-HCl, 1.4 M NaCl, 20 mM ethylenediaminetetraacetic acid (EDTA), 2% CTAB) was added to each tube. The tubes were thoroughly mixed by vortexing for 15 sec and incubated at 65°C for at least 20 min. The samples were sanctioned to cool at room temperature. After adding 500 μL of chloroform, samples were mixed thoroughly, followed by centrifugation at 13000 rpm for 10 min. The supernatant was carefully transferred to a clean tube and mixed with an equal volume of isopropanol. DNA samples were precipitated by placing the tubes at -20°C for at least 1h followed by centrifugation for 10
Figure 5. Schematic illustration of a strategy of next-generation sequencing-based mapping and identification of EMS mutations

EMS mutant, for example, Col background, is outcrossed to Ler erecta to generate an F2 mapping population. F2 segregants displaying the mutant phenotype are selected and bulked for genomic DNA extraction and sequencing library preparation. After subjected to whole genome sequencing using next-generation sequencing technology, sequencing reads are aligned and SNPs are identified using a pipeline of computational scripts. Adapted from Zhang et al. (2014).
EM

wild-type
Col-0

mutant
(Col-0)

X

mutant
(Ler)

cross

F2 segregants

DNA extraction → library preparation → NGS and SNP analyses
min at 10,000 rpm. The precipitated DNA pellets were washed with 1 mL 70% ethanol and air dried. Each pellet was dissolved in 20-50 μL of sterile MilliQ water.

2.7 Checking the Quality of Isolated DNA

The quality and concentration of DNA were evaluated using the Thermo Scientific™ NanoDrop Lite Spectrophotometer. Absorbance at 260, 280 and 230 nm was measured, indicating the content of nucleic acids, proteins and other contaminations, respectively. For good quality DNA, the ratios of A260/A280 and A260/A230 should be over 1.8 and in the range of 2.0 to 2.2, respectively. In addition, DNA samples were also run on 1 % agarose gel to confirm the DNA quality and detect the presence of any RNA contamination. Since it is found that spectrophotometric measurements of DNA concentration were not very accurate for DNA extraction using the CTAB method, I switched to using the Qubit 2.0 fluorometer with the high-sensitivity (HS) DNA quantification reagents (Invitrogen, Carlsbad, CA) for accurate measurements of the extracted DNA before library preparation for whole genome sequencing.

2.8 Bulked DNA Preparation for Bulked-segregant Analysis (BSA)

Individual (F2) plants, showing the mutant phenotype, were used for bulked-segregant analysis for rough mapping. A total of 25 F2 individuals from this population were used for bulk DNA preparation. After extraction of Arabidopsis genomic DNA using the CTAB method (Stewart Jr and Via, 1993), DNA samples were diluted with sterile MilliQ
water to about 200 ng/μL. DNA from individual mutants was then pooled together for further analysis.

2.9 Bulked DNA Preparation for Next-generation Mapping

The isolation of genomic DNA protocol, for next generation mapping, was slightly different than that for bulked-segregant analysis. Firstly, about 100 mg of 25 individual mutant plant tissues were first collected in 2.0 ml Eppendorf tubes and then all pooled together for grinding, in liquid nitrogen, using mortar and pestle. Then a required amount of CTAB buffer was added and the pooled tissues were further ground using mortar and pestle. The mixture was then divided into 2.0 ml Eppendorf tubes followed by incubation at 65°C for 45 min. The rest of the CTAB method was followed to complete the DNA extraction, except to avoid RNA contamination, 1 μL of RNase was added to each tube before adding isopropanol, and incubated at 37°C for 30 min. After dissolving the pellet in 20-50 μL of sterile MilliQ water, individual DNA samples were checked against agarose gel to determine the quality. The DNA quantity was measured using the NanoDrop Spectrophotometer. After confirming the quality and quantity, samples were pooled together and the DNA was precipitated with isopropanol followed by washing with 70% ethanol. The final DNA was checked against quality and quantity before sending for next generation mapping.
2.10 Standard PCR reaction

To amplify DNA fragments for genotyping and other analysis, PCR mixtures were created according to the manufacturer’s instructions. For the standard reaction (genotyping and other analysis) regular Taq DNA Polymerase (Invitrogen) was used. Applied Biosystems GeneAmp 9700 or 2720 PCR system was used for carrying out various amplifications. A standard 20 μL PCR protocol was followed according to the program guideline: a denaturing temperature of 94°C for 15-20 sec, followed by an annealing temperature of 55°C-58°C for 20 sec, and primer extension at 72°C for 30 sec based on target DNA size to be amplified. These three steps were repeated for a total of 32 to 37 cycles, followed by a final extension for 7 min. The PCR product was immediately used or stored at -20°C for future analysis.

2.11 DNA Gel Electrophoresis

PCR DNA fragments were mixed with loading buffer and separated on 1-4% agarose gel electrophoresis. The concentration of agarose gels depends on the size of fragments to be resolved. Electrophoresis was performed in 1x Tris-acetate-EDTA (TAE) buffer (40 mM Tris-acetate, 2 mM EDTA,) and 5 μL 100mL⁻¹ Rad Safe Dye (Invitrogen). 1 kb DNA ladder was used to estimate the size of fragments. After electrophoresis, DNA bands were viewed using a Bio Rad UV trans-illuminator.
2.12 Genetic and Physical Mapping of Mutants

Genetic mapping was accomplished from the F2 mapping population possessing the mutant phenotype. Genomic DNA was isolated from individual plants, and bulked DNA was made as described in bulked DNA preparation for bulked-segregant analysis (BSA). Mapping was carried out using ecotype-specific markers (simple sequence length polymorphisms, SSLPs) which are visible through PCR as described by Lukowicz et al. (2000). Previously identified markers were obtained from the Arabidopsis Information Resource (TAIR: https://www.Arabidopsis.org/index.jsp ) and Arabidopsis mapping platform website (http://amp.genomics.org.cn). Useful small insertions/deletions (INDELs for SSLPs) between the publicly available Col sequence and Ler sequence were generated through Monsanto http://www.Arabidopsis.org/browse/Cereon/index.jsp (Jander et al., 2002). Successful marker primers are listed in (Table 2). Additional SSLP markers were designed based on the natural polymorphism existing between Col and Ler to roughly narrow down the genomic interval containing the mutation site(s).

2.13 Mapping of Mutant Loci with Next-generation Sequencing

Mutant phenotypes were screened from the F2 segregating population. Individuals with mutant alleles were selected and pooled to isolate nuclei DNA for Illumina sequencing described in section 2.9. Nuclear DNA was subjected to whole-genome sequencing (WGS) on a MiSeq sequencing machine. Sequencing data were analyzed using a web-based tool designed for the rapid localization of recessive EMS induced mutations within
Table 2. List of SSLP primers used for genetic mapping in this study

<table>
<thead>
<tr>
<th>Chr.</th>
<th>BAC</th>
<th>Forward Primer Sequence (5'→3')</th>
<th>Reverse Primer Sequence (5'→3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F21M12</td>
<td>GGCTTTCTCGAAAATCTGFCG</td>
<td>TTA CTTTTTGCCCTTC TTGFCATG</td>
</tr>
<tr>
<td>1</td>
<td>F1K23</td>
<td>CCCTAGGCTCTGACCCCTCTT</td>
<td>TTCCAAATGCAAAGAAGACGA</td>
</tr>
<tr>
<td>1</td>
<td>CIW1</td>
<td>ACATTTTCTCAATCTTT AACT</td>
<td>GAGAGCTTCTTT ATTTT GAT</td>
</tr>
<tr>
<td>1</td>
<td>NGA280</td>
<td>CTGATCTCAGCGACAAATAGFCC</td>
<td>GGCCTCCATAAAAGAGC ACC</td>
</tr>
<tr>
<td>1</td>
<td>NGA111</td>
<td>TGTTTTTATCGAACAAATGG</td>
<td>CTCCAGTGGAAGCTAAAAGG</td>
</tr>
<tr>
<td>2</td>
<td>F3C11</td>
<td>GCA CAGTCC AAGTC AATCAAC</td>
<td>CGC TACTGCTTTCC TG TAAAG</td>
</tr>
<tr>
<td>2</td>
<td>F12A24</td>
<td>GGCTGATGTGAGCGAGCTCTTGGAAAG</td>
<td>TGCAACACGTGCTCTT CA TAG</td>
</tr>
<tr>
<td>2</td>
<td>F27L4</td>
<td>TTATG ATGTGTTGAGTTGAGTGC</td>
<td>TTGGTAGATAATAGC GCT TAA</td>
</tr>
<tr>
<td>2</td>
<td>NGA168</td>
<td>GAGGACATGTATAAGGACGTCC</td>
<td>TACCTGACTGCACTGCC</td>
</tr>
<tr>
<td>3</td>
<td>NGA162</td>
<td>CTCTGTCACCTCTTTCTCTCTG</td>
<td>CATGCAATTTGCACTGT AAG</td>
</tr>
<tr>
<td>3</td>
<td>CIW11</td>
<td>CCCCCAGTGGAGTATT</td>
<td>GAAGAAATCCCTAAAGCAC T</td>
</tr>
<tr>
<td>3</td>
<td>CIW4</td>
<td>GTTCATTTA AACTTG CGTGTGTTG</td>
<td>TACGGTCAAGATGGATGATTCC</td>
</tr>
<tr>
<td>3</td>
<td>NGA6</td>
<td>ATGGAGAAGCTTACA ATGTGAC</td>
<td>TGGATTTCTCTC TTCTC TCCC</td>
</tr>
<tr>
<td>4</td>
<td>CIW5</td>
<td>GGT TAAA AATTAGGGTACGA</td>
<td>AGATTACGTGGGAAGC A T</td>
</tr>
<tr>
<td>4</td>
<td>F17A8</td>
<td>CAGACG TATCAA AATGACA AAATG</td>
<td>AGATTTACGTGGAAGCAAAT</td>
</tr>
<tr>
<td>4</td>
<td>F13M23</td>
<td>TTA TCTCAAGAGCT CTTGT GTT</td>
<td>GTCCGGTC AATCTG AGTA G</td>
</tr>
<tr>
<td>4</td>
<td>NGA1107</td>
<td>CGACGAAATCG CAGA AAT TAGG</td>
<td>GCCGAAAAAC AAAAAAA AATCCA</td>
</tr>
<tr>
<td>5</td>
<td>CTR1</td>
<td>CCACTGTGTCCTCT CTTG</td>
<td>TATCAACAGAAACGC A CCGA A</td>
</tr>
<tr>
<td>5</td>
<td>MYJ24</td>
<td>CATTA GAGACC GGTATCGGA</td>
<td>GCCCAAAAAAT TTAGAAA AACC</td>
</tr>
<tr>
<td>5</td>
<td>PHYC</td>
<td>CTCAGAGAAATT TCAGAAAT TATC</td>
<td>AA ACTCGAG AGT TTTTTGCTAGA TCA</td>
</tr>
<tr>
<td>5</td>
<td>CIW9</td>
<td>CAGACG TATCAA AAT GAC AAAATG</td>
<td>GACTACTGCTCAAA A ACTATT CGG</td>
</tr>
<tr>
<td>5</td>
<td>K19B1</td>
<td>AATTTCTAGCATGA ATCAG</td>
<td>TGITTCTTACTG CAATG T G</td>
</tr>
</tbody>
</table>
F2 mapping population (http://bar.utoronto.ca/NGM/; (Austin et al., 2011). Potential mutation site(s) were narrowed down by combining the results from rough mapping and WGS. Candidate SNPs were identified by filtering as many linked SNPs as possible, using multiple appropriate criteria including the type of nucleotide change, its location, and potential cause of an amino acid change. Sequence assembly and SNP annotation Illumina Myseq reads from mutant were mapped against the TAIR10 release of the reference genome.

2.14 Sequencing of Candidate Genes

After identification of the final candidate(s), sequences of the same GA sensitive locus were re-sequenced with gene specific primers by the Sanger sequencing method to confirm the results of Illumina sequencing. DNA sequencing was performed by the DNA sequencing facility at the SCPFRC, London, Ontario. Sequences were analyzed by DNASTAR Laser gene 10 core software. Independent F2 mutants and parental line were used for sequencing and verifying lesions.

2.15 Screening of Homozygous or Heterozygous T-DNA Mutants

For the identified candidate gene, The Arabidopsis Information Resource (TAIR: https://www.Arabidopsis.org/index.jsp) and Salk Institute Genomic Analysis Library (SIGnal: http://signal.salk.edu) were searched for available T-DNA insertion lines. T-DNA insertion lines, which most likely have an insertion in an exon area of
selected candidate gene, were acquired from the *Arabidopsis* Biological Resource Centre (ABRC: https://abrc.osu.edu/). Primers for genotyping of the T-DNA insertion lines were designed by the “iSect Toolbox” (http://signal.salk.edu/tdnaprimers.2.html) (Table 3). Homozygous or heterozygous T-DNA insertion plants were identified by PCR using a T-DNA left border-specific primer Lb1.3 and two gene-specific T-DNA primers (LP and RP). PCR conditions were: 94°C for 5 min; 37 cycles of 94°C for 30 s, 58°C for 30 sec and 72°C for 1 min, followed by a final extension at 72°C for 7 min.

2.16 Confirmation of Candidate Gene (Allelism Test)

Screening of homozygous or heterozygous T-DNA insertion lines confirmed the gene responsible for the mutant phenotype. Mutant alleles of the candidate gene were used for pair-wise crosses with EMS induced parental line and the progeny was examined for segregation ratio of the mutant phenotype.

2.17 Statistical Analysis

Simple statistics (means, standard deviations) were calculated using Excel 2007. Statistical analysis was performed using SPSS version 14.0 (SPSS, Inc., Chicago, Illinois). The Chi-square test was used to determine whether the observed data are in agreement with those of the predicted data based on the genetic explanation of wild-type to mutant phenotype ratio. A p-value of 0.05 or less indicated statistically significant differences.
Table 3. List of other primers used for this study

<table>
<thead>
<tr>
<th>Primers</th>
<th>Forward Primer sequence (5'→3')</th>
<th>Reverse Primer sequence (5'→3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>sdg8-2</td>
<td>TAAAGAGGCTGCATCATGG</td>
<td>CACTGTCAGATAAGCGCTGG</td>
</tr>
<tr>
<td>SALK_017607C</td>
<td>TCAGGGCATTCATTCTCATTTC</td>
<td>TGTTGGATTTAGGCGTCTTTG</td>
</tr>
<tr>
<td>ga2-1</td>
<td>TTCCATCACCAGCTCCCA</td>
<td>TGAACCCAAATGGAATCGTCA</td>
</tr>
<tr>
<td>LB1.3</td>
<td>ATTTTGCGATTTCCGAAC</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 3: RESULTS

3.1 Screening of F2 Mapping Population

As described in section 2.3, a germination mutant (temporarily named gml1) was identified from an EMS mutagenized population in the sdg8-2 background. To identify germination related mutants, F2 mapping population was grown on MS medium and the ratio of germinated vs. non-germinated seeds were determined (Table 4). A total of 141 F2 mutant seedlings were isolated. The morphology of these mutants was almost similar to gml1 and germination can be restored almost completely by GA3. Without further GA3 application at the initial stage of the plant life cycle, the seedlings looked like wild type, but at the later stages, upon transferring to soil, these seedlings developed into dark green dwarf plants (Figure 6). Apart from these phenotypic characteristics, petals and stamens were found to be very poorly developed while pistils and sepals were almost normal. Even though no seeds were being produced, seeds could be obtained by further application of GA once per week.

3.2 Bulked-Segregant Analysis (BSA) to Locate the Mutation Site in Chromosome

Since homozygous mutants could not form any seeds, the heterozygous gml1 line was crossed with wild-type Ler accession to generate a segregating mapping line. Plants with wild type and mutant phenotypes segregated in a 16:1 ratio (Table 4) indicating that the
Figure 6. The Morphological phenotypes of mutants and wild-type

A. Healthy seeds are showing no major defects prior to sowing on MS media. Scale bar indicates 1000 µm.

B. Non-germinating mutant and a germinating seedling in MS media. Scale bar indicates 1000 µm.

C. Non-germinating seeds were transferred to MS media supplemented with 10 µM GA$_3$. Germinated seedlings grown in the soil after 14 days, showing dwarf mutants and wild type.

D. Mutants were grown for 45 days in separate pots compared to wild-type.

E. Close up view of a dark green dwarf mutant.

F. Typical dwarf unopened flower phenotype in a late stage of the plants life cycle.
gm11 germination mutant is caused by a single locus recessive mutation. To determine the chromosomal location of the mutation, a bulked-segregant analysis was conducted for rough mapping with 22 SSLP markers (Figure 7 A). If any marker is linked to the mutant locus, it should show a clear bias towards the Col ecotype specific mutant pool, indicating low recombination with the mutant locus (Lukowitz et al., 2000). Three DNA samples, from wild-type Col-0, F1 plants and pooled F2 seedlings, were prepared and used for a bulked-segregants analysis. The first two samples were homozygous and heterozygous, respectively, for all genetic loci and served as controls. The third sample, pooled F2 seedlings, was homozygous for markers linked to the mutation but essentially heterozygous for unlinked markers. Marker NGA111 showed a clear bias towards the Col specific band in the mutant pool compared to the heterozygous control (F1). The bulked-segregant analysis data clearly indicated that the mutation is localized at the bottom arm of chromosome 1, near marker NGA111 (27.35 Mb) (Figure 7 B).

### 3.3 Next-generation Mapping by Genome Sequencing

Genome sequencing was carried out at AAFC, London, Ontario. Pooled genomic DNA from 25 F2 mutants (rescued by GA3 application) was used for NGS. The samples were prepared and loaded into the flowcell of the Illumina MiSeq according to manufacturer’s instructions. For detecting SNPs and InDel between mutant and Col-0, reads were aligned against the TAIR reference version Col-0 sequence. After uploading the NGM data into
Figure 7. Bulked-segregant analysis of pooled F2 mutant seedlings

A. Schematic representation of 22 SSLP marker positions used in the rough genetic mapping experiment. Centromeres are represented by open circles. Rectangle around NGA11 represents mutation site.

B. Gel electrophoresis of PCR products for SSLP markers in chromosome 1. PCR products for markers in four other chromosomes are not shown due to space restraints. Col-0 and F1 were used as controls for locating the marker linkage with pooled F2 mutant sample. Col specific bands are marked with an asterisk (*). The mutation in Col background linked to the markers, NGA111, is judged by the clear bias toward amplification of the Col band from the genomic DNA of F2 mutants. The ratio of Col to Ler erecta amplification of other markers is about the same between the heterozygous control and F2 mutants, suggesting that the mutation is unlinked to other loci on chromosome 1.
Table 4. Segregation of F2 progeny derived from different crosses

<table>
<thead>
<tr>
<th>Cross</th>
<th>Total</th>
<th>G⁺ᵃ</th>
<th>G⁻ᵃ</th>
<th>Expected ratio</th>
<th>χ²ᵇ</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>gm11 x Ler F2-12</td>
<td>600</td>
<td>558</td>
<td>42</td>
<td>16:1</td>
<td>0.15</td>
<td>0.45</td>
</tr>
<tr>
<td>gm11 x Ler F2-13</td>
<td>700</td>
<td>651</td>
<td>49</td>
<td>16:1</td>
<td>0.50</td>
<td>0.48</td>
</tr>
<tr>
<td>gm11 x Ler F2-14</td>
<td>700</td>
<td>650</td>
<td>50</td>
<td>16:1</td>
<td>0.32</td>
<td>0.57</td>
</tr>
</tbody>
</table>

A chi-square test was used to ensure that observed data are in agreement with the predicted ratio of wild-type to mutant phenotype. ᵃ G⁺ Number of F2 germinated seeds; G⁻, number of non-germinated seeds after 14 days on MS media; ᵇ, Test for 16:1 segregation of the germination phenotype; P, the significance level
Figure 8. Next generation EMS mutation mapping web application

A. Screenshot of the initial phase of region determination and SNP annotation. Col-0 TAIR-10 reference genome was used for SNP data analysis.

B. Screenshot of the last phase of region determination and SNP annotation. SNPs were filtered using specific criteria including EMS-affected mutations and whether they produced a non-synonymous amino acid change within the coding sequence. The NGM tool is accessible at http://bar.utoronto.ca/NGM.
NGM - Next-generation EMS mutation mapping

Preprocess/Upload Data | Map to chromosome | Charity belt partitioning | Fine Map and Annotate

**Preprocess:**
NGM needs to preprocess your MAQ or Samtools generated SNP calls and add discordant charity information. The applet provided below will create an "emap" file for upload in the next section.

**Important Notes:**
1. You need JRE 1.5 or higher and Java enabled in your browser to run the applet.
2. As the applet reads and writes to your local Reponomy, you have to click "Allow" in the Security dialog box that pops up.
3. Samtools has deprecated the "plink" function and replaced it with "plink2". Therefore, Samtools version 0.1.16 or earlier must be used with NGM. It is available here.
4. Samtools v1.15 data in the form of a VCF file can now be processed using this beta script. Note that VCF should be quality-filtered by the user and the NGM filters box below unchecked.

**Note:** If there are problems loading the applet, you need to clear Java applications cache by following instructions here. If there are problems with filtering, please clear your web browser cache.

Start Applet

The applet will store results in a local file created with a name of your choice.

Alternatively, Perl scripts can be downloaded and run locally against SamTools or MAQ generated data to produce the necessary file for upload.

To try NGM mapping using some example data, click here or here or here.

---

NGM - Next-generation EMS mutation mapping

Preprocess/Upload Data | Map to chromosome | Charity belt partitioning | Fine Map and Annotate

**Upload Data:**

<table>
<thead>
<tr>
<th>Desktop\NextGen\NGM</th>
<th>Browse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Select gene reference used for mapping</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Select SNP data by quality criteria (MAQ/SAM only)</td>
<td></td>
</tr>
<tr>
<td>G: Minimum depth</td>
<td></td>
</tr>
<tr>
<td>250: Maximum depth</td>
<td></td>
</tr>
<tr>
<td>40: Minimum neighbour quality</td>
<td></td>
</tr>
<tr>
<td>20: Minimum best read quality</td>
<td></td>
</tr>
<tr>
<td>20: Minimum consensus quality</td>
<td></td>
</tr>
</tbody>
</table>

Upload and analyze | Eject

---

NGM - Next-generation EMS mutation mapping

Preprocess/Upload Data | Map to chromosome | Charity belt partitioning | Fine Map and Annotate

**Current file:** nucmerGen

Select interval for SNP annotation: produce line [30000, 100] [500000, 100] [10000, 100] [150000, 100] [50000, 100]

If you want several SNP tests on the last page, you can display them using the slide next to the chart.

Heterozygous/Homozygous signal ratio (0.5 / 2 kernel = 1.80)

Position (bp)

Drag the red bars to change the interval for which annotations are revealed.

Select the signal level:
- Moderate deviation
- High deviation
- Exon deviation
- Conserved mutation
- Low conservation
- High conservation

**SNP annotations:**

<table>
<thead>
<tr>
<th>Chrom Position</th>
<th>Ref base</th>
<th>SNP base</th>
<th>Depth</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>A</td>
<td>100</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>B</td>
<td>50</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>C</td>
<td>10</td>
<td>0.1</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
<td>D</td>
<td>5</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Figure 9. Genome-wide SNP variation patterns plotted by histograms for each chromosome.

A. Chromosome 1, as a control, exhibited the default pattern of genome-wide natural variation. Histograms of the highly reproducible frequency of genome wide SNPs between Col-0 and Ler *erecta* accessions (250 kb bins). Adapted from Austin et al. (2014).

B. SNPs were plotted based on their abundance on each chromosome, indicating bin size of 250 kb. Non recombinant region, for mutant, around the bottom arm on chromosome 1 is indicated by a star. All other chromosomes exhibit the default pattern recombination of natural variation.
The University of Toronto Next Generation EMS mutation website (http://bar.utoronto.ca/NGM/) (Figure 8), SNPs frequencies of this mutant were plotted using bin size of 250 kb based on their abundance over each chromosome. Non-recombinant regions lacking SNPs within the bottom arm of chromosome 1 were indicated by a star (Figure 9). Together, the bulked-segregant analysis results and the NGM data strongly suggest that the mutation is indeed located in the bottom arm of chromosome 1.

3.4 Mutation is Localized in $GA_2$ Gene

3.4.1 Identification of Candidate Genes

After filtering SNPs on chromosome I, mutations in the Col genetic background were analyzed to identify the mutant loci and candidate gene(s). Seven new SNPs (Table 5) were detected in coding sequences, 4 of which were synonymous and three non-synonymous on chromosome I. After filtering the SNPs data three non-synonymous mutations at positions 27586767, 28022871 and 29891246 were predicted as candidate loci for causal mutations. Three non-synonymous candidates SNPs, identified by NGM, were found in genes $AT1G73370$, $AT1G74580$ and $AT1G79460$, respectively. To confirm whether one of the candidate SNPs were indeed responsible for the $gm11$ dwarf phenotype, I searched the TAIR website for the function(s) of these three genes. $AT1G73370$ encodes Arabidopsis sucrose synthase 6, the functions included sucrose biosynthetic and metabolic process, callose deposition in the phloem. A mutation in $AT1G73370$ shows minor phenotypic variation during vegetative growth, flowering stalks
are reduced in height and all inflorescence parts are smaller than those wild-type plants (Paul et al., 2011). \textit{AT1G74580} encodes a Pentatricopeptide Repeat (PPR), is a superfamily protein, which has different functions including biological process, cellular component and plant structure. Currently, there are no other publicly available articles or phenotypic information available for this gene. The third candidate SNP is found in \textit{AT1G79460}. This accession is also referred to as \textit{ga2}. It encodes an ent-kaurene synthase which catalyzes the conversion of copalyl diphosphate to ent-kaurene in the GA biosynthesis pathway (Yamaguchi et al., 1998a). The publicly available database suggests that \textit{ga2} mutants have almost the same phenotypic characteristics compared to \textit{gm11} mutant. In \textit{Arabidopsis}, six GA-sensitive germination mutants (\textit{ga-1} through \textit{ga-6}) have been isolated and characterized (Koornneef and Van Der Veen, 1980). Each mutant is blocked in a particular step in GA biosynthesis. \textit{ga2} mutant seeds are non-germinating, and the seedlings are extremely dwarfs. My NGM data identified a G $\rightarrow$ A transition at the 855\textsuperscript{th} \textit{ga2}’s coding sequence, creating a premature stop codon. Based on all of the gathered information, I predict that the \textit{gm11} phenotype is caused by a mutation in the \textit{ga2} gene (\textit{AT1G79460}).

3.4.2 Genotyping by DNA Sequencing of the Candidate Gene (\textit{ga2}) in All 25 F\textsubscript{2} Individuals Used for NGM

As described in section 3.4.1, NGM identified an EMS mutagenesis-induced G $\rightarrow$ A nucleotide base change at the 855\textsuperscript{th} position of \textit{ga2}/\textit{AT1G79460}’s coding sequence, which is predicted to be the mutation causing the \textit{gm11} mutant phenotype. As a first step
Table 5. Seven identified SNPs at the bottom of chromosome 1

<table>
<thead>
<tr>
<th>Chro.</th>
<th>Position</th>
<th>Ref. base</th>
<th>Alternate base</th>
<th>Accession</th>
<th>Type</th>
<th>Ref. codon</th>
<th>Alternate codon</th>
<th>Ref. aa</th>
<th>New aa</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>26621549</td>
<td>G</td>
<td>A</td>
<td>AT1G70600.1</td>
<td>Syn</td>
<td>GGG</td>
<td>GGA</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>1</td>
<td>27121329</td>
<td>C</td>
<td>T</td>
<td>AT1G72080.1</td>
<td>Syn</td>
<td>ACC</td>
<td>ACT</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>1</td>
<td>27586767</td>
<td>C</td>
<td>T</td>
<td>AT1G73370.1</td>
<td>Nonsyn</td>
<td>CAG</td>
<td>TAG</td>
<td>Q</td>
<td>Stop</td>
</tr>
<tr>
<td>1</td>
<td>27687981</td>
<td>C</td>
<td>T</td>
<td>AT1G73640.1</td>
<td>Syn</td>
<td>TTC</td>
<td>TTT</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>1</td>
<td>28022871</td>
<td>G</td>
<td>A</td>
<td>AT1G74580.1</td>
<td>Nonsyn</td>
<td>GTC</td>
<td>ATC</td>
<td>V</td>
<td>I</td>
</tr>
<tr>
<td>1</td>
<td>28101295</td>
<td>G</td>
<td>A</td>
<td>AT1G74790.1</td>
<td>Syn</td>
<td>GCG</td>
<td>GCA</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>1</td>
<td>29891246</td>
<td>G</td>
<td>A</td>
<td>AT1G79460.1</td>
<td>Nonsyn</td>
<td>TGG</td>
<td>TGA</td>
<td>W</td>
<td>Stop</td>
</tr>
</tbody>
</table>

Three SNPs were non-synonymous, and 4 were synonymous. Three non-synonymous SNPs at 27586767 and 29891246 introduced a stop codon and at 28022871 introduced amino acid substitution.
Figure 10. DNA sequencing result of *AT1G79460* gene for F2 mutants

Sequencing diagram represents all 25 F2 mutants. Only gene specific forward primer was used for this sequencing. Arrow sign indicates point mutation (G→A) at 855 position of the candidate gene.
Forward Strand Sequencing of gene *AT1G79460*
to confirming that ga2 is indeed allelic to gm11, I amplified and sequenced the mutation sites in all the 25 F2 mutants used for NGM (Figure 10). Col-0 was analyzed as a control and the results showed that Col-0 did not contain the point mutation at position 855 but all the 25 mutants demonstrated the point mutation. For further validation, apart from these 25 mutants, other different mutant lines were also re-sequenced to determine whether the point mutation was at the same location compared to previously identified candidate SNPs (data not shown).

3.4.3 sdg8-2 Background is Not Involved in gm11 Phenotype

As mentioned in section 2.3, germination mutants (gm11) were isolated from an EMS mutagenized population in the sdg8-2 background. To eliminate the potential involvement of sdg8-2 in conferring the gm11 phenotype, twenty-five individual F2 mutants were genotyped by PCR to characterize their genotypes at the SDG8-2 gene. The PCR data showed that among the twenty-five individual mutants, two were heterozygous for the sdg8-2 T-DNA insertion (Figure 11), thus eliminating the contribution of the sdg8-2 background to the gm11 phenotype.
3.5 Genetic Evidence Confirming that Mutation in \textit{ga2} is the Cause of the Germination Mutant Phenotype of \textit{gm11}

3.5.1 \textit{gm11} is Allelic to \textit{ga2}

To further confirm whether \textit{gm11} is allelic to \textit{ga2}, two T-DNA insertion lines of \textit{AT1G79460} were obtained from ABRC: SALK_095437 and SALK_17607C. The insertion sites were in the 7\textsuperscript{th} and 13\textsuperscript{th} exons of \textit{AT1G79460}, respectively (Figure 12 A). T-DNA insertion lines were grown in parallel to wild type Col-0 on MS medium to observe their germination behavior and phenotypic characteristics. The T-DNA lines and other seeds were germinated and grown on MS medium in the absence of GA\textsubscript{3} (Figure 12 B). Non-germinated seeds were later transferred to MS medium supplemented with 10 µM GA\textsubscript{3} to observe the germination behavior (Figure 12 C). The ratio of plants with wild type and mutant phenotypes for SALK_017607C were 3:1 (Table 6), indicating that \textit{gm11} mutation is a single recessive locus and allelic to \textit{ga2}. However, the ratio of plants with wild type and mutant phenotypes for SALK_095437 was not consistent with the 3:1 ratio. PCR-based genotyping showed that only SALK_017607C had the T-DNA insertion (Figure 12 D). To determine whether the phenotype is similar to \textit{gm11} mutants, T-DNA line mutants were grown in soil in parallel with other mutants. SALK_17607C T-DNA insertion mutants showed almost identical phenotype compared to \textit{gm11} mutants (Figure 12 E), which strongly suggests that the mutation in \textit{ga2} is indeed responsible for this dark green dwarf phenotype.
Figure 11. Agarose gel electrophoresis of PCR products demonstrating the genotype of the sdg8-2 locus

Confirmation of homozygous or heterozygous genotypes for the sdg8-2 T-DNA insertion for all 25 F2 mutants. #5 and #15 harboring heterozygous mutations in their SDG8-2 locus while this locus is homozygous for all other mutants. The absence of bands indicate wild-type (WT). M, DNA marker; Col-0, WT control; WT, wild type; H, homozygote; Het, heterozygote.
Figure 12. Analysis of T-DNA insertion lines for validation of mutation in \( ga2 \) gene

A. Gene Structure of \( ATIG79460 \) and locations of T-DNA insertion sites. Boxes represent exons and lines represent introns and untranscribed regions.

B. Germination behavior of T-DNA lines, \( gm11 \) and wild-type seeds without GA3 supplement.

C. Germinating seeds, after 14 days, for each T-DNA mutant in MS media supplemented with 10 \( \mu \)M GA3.

D. Agarose gel electrophoresis of PCR products demonstrating the genotype of T-DNA insertion for SALK_17607C.

E. Morphological variation of the mutant showing WT Col-0 as a control at 45 days
3.5.2 Allelism Test

An allelism test was performed by crossing parental (gm11) and SALK_017607C T-DNA insertion mutants, and analyzed the segregation ratio of F1 seeds. Since a homozygous mutant for this allele did not form any seeds without further exogenous GA application, heterozygous plants for both T-DNA and gm11 were used for the crosses. Plants, either heterozygous or homozygous for the T-DNA insertion, were identified by PCR-based genotyping following standard protocol (http://signal.salk.edu/tdnaprimers.html) (Figures 13 A and 13 C). Sequencing was done by ga2 specific primers for the identification of heterozygous or homozygous plants for gm11 mutation (13B and 13D). Plants heterozygous for SALK_017607C and gm11 lines were used for crosses each other. SALK_017607C was used as the mother plant and gm11 line as pollen donor. In the F1 generation, plants with wild-type and mutant phenotypes segregated in a 3:1 ratio (Table 7), indicating that the mutant is allelic to ga2. To identify true F1 plants, 20 mutant plants were genotyped with SALK_017607C T-DNA specific primers to determine whether an individual mutant was heterozygous for the T-DNA. PCR results showed that every F1 mutant was heterozygous for the T-DNA. For further validation, 20 F1 mutants were also sequenced with ga2 gene specific primers. The sequencing results confirmed heterozygous for the causal mutation site (data not shown). The combined sequencing results strongly suggested that the mutation was indeed in the ga2 gene, i.e., gm11 is allelic to ga2.
Figure 13. PCR and sequencing genotyping for identifying heterozygous or homozygous T-DNA and gm11 mutants

A. Identification of heterozygous SALK_O176076C T-DNA insertion plants. The number indicates either heterozygous or wild type for T-DNA insertion. The absence of bands indicates no T-DNA insertion. M, DNA marker; Col-0, WT control; WT, wild-type; Het, heterozygote.

B. Arrow head indicates reverse sequences with GA2 gene specific primers for the identification of heterozygous plants.

C. Identification of homozygous SALK_O176076C T-DNA insertion for all mutants. The number indicates either homozygous or heterozygous for T-DNA insertion. The absence of bands indicates no T-DNA in these plants. M, DNA marker; Col-0, WT control; WT, wild-type; Het, heterozygote.

D. Arrow head indicates forward sequences with ga2 gene specific primers for the identification of homozygous gm11.
Table 6. Germination percentage of T-DNA mutant seeds

<table>
<thead>
<tr>
<th>T-DNA</th>
<th>Total</th>
<th>G&lt;sup&gt;a&lt;/sup&gt;</th>
<th>G&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Expected ratio</th>
<th>χ&lt;sup&gt;2b&lt;/sup&gt;</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salk_017607c</td>
<td>91</td>
<td>69</td>
<td>22</td>
<td>3:1</td>
<td>0.05</td>
<td>0.86</td>
</tr>
<tr>
<td>Salk_095437</td>
<td>113</td>
<td>95</td>
<td>18</td>
<td>3:1</td>
<td>4.75</td>
<td>0.025</td>
</tr>
</tbody>
</table>

A chi-square test was used to determine whether the observed data are in agreement with the predicted ratio of wild-type to mutant phenotype. <sup>a</sup> G<sup>+</sup>, Number of germinated seeds for T-DNA insertion lines; G<sup>a</sup>, number of non-germinated seeds after 14 days on MS media; <sup>b</sup>, Test for 3:1 segregation of the germination phenotype; P, the significance level.
Table 7. Segregation of F1 progeny derived from heterozygous *gm11* and SALK_017607C cross

<table>
<thead>
<tr>
<th>Cross</th>
<th>Total</th>
<th>G⁺ᵃ</th>
<th>G⁻ᵇ</th>
<th>Expected ratio</th>
<th>χ²ᵇ</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-F-4 x SALK_017607C</td>
<td>95</td>
<td>70</td>
<td>25</td>
<td>3:1</td>
<td>0.23</td>
<td>0.76</td>
</tr>
</tbody>
</table>

A chi-square test was used to determine whether observed data are in agreement with the predicted ratio of wild-type to mutant phenotype. ᵃ G⁺, Number of F₁ seeds germinated; ᵇ G⁻, number of non-germinated after 14 days on MS media; ᵇ Test for 3:1 segregation of the germination phenotype; P, significance level
3.6 Application of Exogenous GA Rescues Phenotypes of ga2 Mutant

For further confirmation of whether ga2 mutant phenotypes were due to defects in the GA biosynthesis pathway, I examined whether exogenously applied GA$_3$ could rescue the ga2 mutant phenotype to a mature stage. Plants were sprayed with GA$_3$ once per week (Soh, 2006). Indeed, the dwarf and sterile flowering phenotypes were restored to normal phenotypes by the GA$_3$ treatment (Figure 14). In addition, these mutants successfully formed seeds and had increased number of inflorescences and siliques compared to wild type. However, the quantity of seeds per silique was less contrasted with wild type (data not shown). These results further confirm that ga2 mutant phenotypes caused by interfering with GA-biosynthesis pathway.
Figure 14. Phenotype of germination mutants in comparison with wild-type

A. Mutant seedlings under standard growth conditions on MS media with supplemented 10 µM GA₃ followed by growing in pots at 45 days.

B. Morphological variation of the mutants compared to wild-type plants at the same age after foliar application of 100 µM GA₃.
CHAPTER 4: DISCUSSION AND CONCLUSION

The study of gibberellin (GA) deficient mutants has been intimately linked with seed germination. GA-deficient mutants are incapable of germinating without the supplement of exogenous GA (Mitchum et al., 2006). Previous studies revealed that seed germination mutants are controlled by genetic as well as other environmental factors (Huang et al., 2010). In Arabidopsis and tomato GA deficient mutants were isolated by selecting non-germinating seeds that failed to germinate without an exogenous application of GA. To date, several GA deficient mutants have been identified and characterized. Most germination mutants were caused by a mutation in GA biosynthesis genes, and only a few were responsive to exogenous GA (Steber et al., 1998). Mutations in genes GA1, GA2, and GA3, acting in early GA biosynthesis, displays a number of GA-rescuable phenotypes, including failure of germination, dwarfism, dark green leaf, underdeveloped petals and reduced fertility (Koornneef and Van Der Veen, 1980; Wilson et al., 1992). However, previous studies reported that semi-dwarfism is a valuable trait for crop breeding due to resistant to damage by wind and rain, allowing for crops with a better harvest index (Spilmeyer et al., 2002). Doubling the productions of wheat and rice yields during the ‘Green Revolution’ was made possible by the introduction of dwarfing traits into plants. Identified genes responsible for these traits showed that they interfere with the action or production of GA. From these two important discussion points, germination mutants and the green revolution, the present study was directed in an attempt to identify a novel gene for germination mutants using next generation-mapping. Although several GA response mutants had been isolated in different species, uncovering
new dwarf traits in *Arabidopsis* is still important for a better understanding of the plant developmental processes. The results obtained from this study are discussed below.

This project was conducted to characterize a seed germination mutant (*gm11*). To generate the *gm11* mutant (Figure 4), an EMS mutagenized population in the *sdg8-2* background was screened. In this study, F2 mapping population was used for next-generation mapping. The results from genotyping with *sdg8-2* T-DNA insertion primers suggested that there was no involvement of *sdg8-2* in conferring the *gm11* genotype (Figure 11). My research data suggest that phenotypic characteristics of the *gm11* mutant were consistent with those of a previously identified *ga2* mutant (Figure 4). Former studies have reported that *ga2* mutants exhibit GA-rescuable phenotypes, similar to those of the *gm11* mutant, including failure of germination, dark green leaf, underdeveloped petals and stamen accompanied by reduced fertility, an expanded number of buds per inflorescence, deferred blossoming, diminished apical dominance, and postponed senescence (Koornneef and Van Der Veen, 1980; Wilson et al., 1992). Based on the phenotypic and genotypic results, I presumed that *gm11* is not conferred by *sdg8-2*. Thus, I can predict that *gm11* is a GA sensitive mutant resembling a GA biosynthesis mutant, *ga2*.

I proposed that *gm11* might act by interfering with the signal transduction pathway of the GA. Besides the phenotypic characteristics, it is also necessary to characterize the mutation site by genetic mapping. Essentially, to determine whether the *gm11* mutant is recessive or dominant and if it is controlled by one or more loci, the ratio of wild-type to mutant phenotype must be observed in the F2 mapping population. It was found that plants with wild-type to mutant phenotypes segregated in a 16:1 ratio (Table 4). It is
predicted that if the homozygous mutant parent is used for crossing with *Ler erecta* wild-type, the phenotype of wild-type to mutant will segregate in the ratio of 3:1. Since the homozygous mutant showed reduced fertility and did not form any seeds, a heterozygous *gml1* line was crossed with *Ler erecta* wild-type to generate the segregating mapping population. Taking this into consideration, my results also suggest that *gml1* is conferred by a single recessive mutant. Based on this hypothesis, to locate the mutation site, I performed bulked-segregant analysis with 22 SSLP primers (Table 2) followed by next generation mapping. Bulked-segregant analysis is the most commonly used method to identify markers that are genetically linked to a mutation (Michelmore et al., 1991). Resources for sequence polymorphisms and molecular markers are now available, and can be used for fine mapping with a polymerase chain reaction-based method. In *Arabidopsis* map-based cloning process, simple sequence length polymorphism (SSLP) markers have been developed to map genes of interest (Lukowitz et al., 2000). The bulked-segregant analysis data from DNA samples of pooled F$_2$ mutant seedlings, indicates that the NGA111 marker is linked to the mutant locus near the bottom arm of chromosome 1, showing a clear bias to the Col specific band in the mutant pool compared to the heterozygous control (F$_1$) (Figure 7). This is in agreement with the concept of bulked-segregant analysis discussed by Lukowitz et al., (2000) which states that if any marker is linked to the mutant locus, it should show a clear bias towards the Col ecotype specific mutant pool compared with heterozygous control (F$_1$), representing low recombination with the mutant locus.
NGS technology is a powerful tool to map genetic mutations underlying the phenotype of interest in *Arabidopsis* using whole genome sequencing against pooled members of F2 mutants. The sequence data generated from the pooled mutant population is aligned against the reference genome. Polymorphism between the mutant and mapping lines are used to identify the homozygous non-recombinant region of the mutation site (Austin et al., 2014). I used 25 homozygous pooled F2 mutants for whole genome sequencing. SNP frequencies of this mutant, using bin size of 250 kb, showed that the non-recombinant regions lacking SNPs are within the bottom arm of chromosome 1 (Figure 9). Combining the data from bulked-segregant analysis and NGM, the results strongly suggest that the mutation is a single recessive locus located in bottom arm of chromosome 1.

Filtering the candidate SNPs around the mutation site was the most challenging task. NGM data suggested around seven candidate SNPs in the bottom arm of chromosome 1 (Table 5), four of which were synonymous and three non-synonymous. After filtering SNPs, mutations at positions 27586767, 28022871 and 29891246 were identified as candidate loci for causal mutations. Despite these complications, I was able to identify a short list of candidate genes from the NGM data. NGM identified three of the candidate SNPs, in genes *AT1G73370*, *AT1G74580* and *AT1G79460*. After searching the TAIR website for the function(s) of these three genes, I found that *AT1G73370* is potentially involved in sucrose biosynthetic and metabolic process and mutation in this gene has minor phenotypic variations like reduced flowering stalks and floral parts compared with wild-type (Paul Barratt et al., 2011). *AT1G74580* is a member of a superfamily protein that has structural plant function rather than potential involvement in the GA biosynthesis
pathway. Finally, one of the candidate SNPs for gm11, identified in the AT1G79460 gene (GA2, ent-kaurene synthase), contains a G → A base pair transition, resulting in a premature stop codon at position 855. The ga2 mutant has been studied and a similar phenotype to my gm11 mutant was reported (Koornneef and Van Der Veen, 1980). The biochemical characterization of this mutant demonstrated that ga2 mutant is impaired the catalyzes process conversion of copalyl diphosphate to ent-kaurene in the GA biosynthesis pathway (Yamaguchi et al., 1998b). In addition, I found that sequencing of gm11 mutant identified a G → A substitution in the coding region of GA2. This is similar to the ga2-1 allele that has a single base substitution which introduces a premature stop codon (Yamaguchi et al., 1998b) (Figure 10). Based on all of the gathered information, it suggests that GA2 is the best candidate for a germination mutant. SALK T-DNA line that I ordered for the GA2 gene, showed the particular ratio of non-germinating dwarf phenotype, similar to my gm11 mutant. The ratio of plants with wild-type to mutant phenotypes, for this T-DNA, was also 3:1 (Table 6).

Furthermore, allelism tests were conducted to observe recessive mutants exhibiting the same gm11 phenotype after complementation. Complementation of the gm11 mutant and SALK T-DNA line in the F1 generation exhibited a ratio of 3:1 wild-type to mutant (Table 7), indicating that gm11 is allelic to ga2. Genotyping results also showed that all mutants were heterozygous for the T-DNA insertion in the F1 generation. Sequencing with ga2 gene specific primer also revealed heterozygosity for all mutants in the F1 generation. Thus, the combined results strongly suggest that the mutation is indeed in the ga2 gene, i.e., gm11 is allelic to ga2.
It has been reported (Soh, 2006) that exogenous application of GA3 could rescue mutant phenotypes to a mature stage, suggesting that germination mutants were caused by defects in GA signaling. Based on this finding, I sprayed 100 μM GA3 solutions once a week. Interestingly, the dwarf and reduced fertile plants were able to reach maturation following the GA3 treatment (Figure 14). However, the quantity of seeds per silique was less compared to that of wild-type. These results suggest that the gml1 mutant phenotypes are caused by interference in the GA biosynthesis pathway.

A rational approach to improving crop varieties depends on a better understanding of plant biology. Gibberellin plays an important role in diverse aspects of seed germination and other developmental process (Kucera et al., 2005). GA signaling pathways are prime targets for manipulating genetic traits in crop yield. The introduction of dwarfing genes into cereal crops was crucial for ‘green revolution’ (Hedden, 2003; Salamini, 2003). There are several strategies for identification of genes affecting different phenotypes in plants. One approach is to analyze plant mutants that exhibit a mutant phenotype and then identify the responsible gene(s). Next-generation mapping platforms are most commonly employed to map genetic mutations using whole genome sequencing. The identified gene (ga2-11), responsible for dwarf traits, interferes with the production of the gibberellin (GA) plant hormones. Since now the complete GA synthesis pathway is known, future challenges on GA are 1) to understand how this pathway is dynamically regulated in the context of specific plant growth and developmental phases; 2) to dissect the mechanisms underlying its diverse roles in plant growth, development and responses to pathogens and stresses; and 3) its crosstalk with other plant hormones. New knowledge gained from such future studies will allow us ways to precisely control plant growth and development.
at specific stages, including seed germination, to maximize crop productivity and benefits to agriculture and our environment.
REFERENCES


APPENDIX I. License Agreement for Figure One and Two

American Society of Plant Biologists LICENSE
TERMS AND CONDITIONS

Jul 20, 2015

This is a License Agreement between University of Western Ontario -- Md Jakir Hossan ("You") and American Society of Plant Biologists ("American Society of Plant Biologists") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by American Society of Plant Biologists, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number 3673231237650
License date Jul 20, 2015
Licensed Content Publisher American Society of Plant Biologists
Licensed Content Publication Publication1
Licensed Content Title The arabidopsis book
Licensed Content Author None
Licensed Content Date Jan 1, 2002
Type of Use Thesis/Dissertation
Requestor type Academic institution
Format Print, Electronic
Portion image/photo
Number of images/photos requested 2

Title or numeric reference of the portion(s) Figure 1. Schematic presentation of processes controlling seed dormancy and germination in an Arabidopsis seed, Figure 2. Arabidopsis seed germination

Title of the article or chapter the portion is from SEED DORMANCY AND GERMINATION

Editor of portion(s) N/A
Author of portion(s) Le Änne Bentsinka, Maarten Koornneef
Volume of serial or monograph. N/A
Page range of the portion 2-3
Publication date of portion 2008
Rights for Main product
Duration of use Current edition and up to 5 years
Creation of copies for the disabled no
With minor editing privileges no
For distribution to Canada
TERMS AND CONDITIONS

The following terms are individual to this publisher:

OTHER TERMS AND CONDITIONS:

STANDARD TERMS AND CONDITIONS

1. Description of Service; Defined Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the “Work(s)”). Copyright Clearance Center, Inc. (“CCC”) grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the “Rightsholder”). “Republication”, as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Confirmation. “User”, as used herein, means the person or entity making such republication.

2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a “freelancer” or other third party independent of User and CCC, such party shall be deemed jointly a “User” for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.
APPENDIX II. License Agreement for Figure Five

From: Wiley Global Permissions <permissions@wiley.com>
Sent: 15 July 2015 05:23
To: Md Jakir Hossan
Subject: RE: Permission to Use Copyrighted Material in a Master’s Thesis

Dear Md Jakir Hossan,

Thank you for your request.

Permission is granted for you to use the material requested for your thesis/dissertation subject to the usual acknowledgements and on the understanding that you will reapply for permission if you wish to distribute or publish your thesis/dissertation commercially. You must also duplicate the copyright notice that appears in the Wiley publication in your use of the Material.

Permission is granted solely for use in conjunction with the thesis, and the material may not be posted online separately.

Any third party material is expressly excluded from this permission. If any of the material you wish to use appears within our work with credit to another source, authorisation from that source must be obtained.

Yours Sincerely,

Rebecca Cook
Permissions Assistant
John Wiley & Sons Ltd
The Atrium
Southern Gate, Chichester
West Sussex, PO19 8SQ
UK

WILEY
Dear,

I am a University of Western Ontario graduate student completing Master's thesis entitled "Identification of Non-germinating gibberellin sensitive mutant in Arabidopsis thaliana using Next generation mapping". My thesis will be available in full-text on the internet for reference, study and/or copy. Except in situations where a thesis is under embargo or restriction, the electronic version will be accessible through the Western Libraries web pages, the Library's web catalogue, and also through web search engines. I will also be granting Library and Archives Canada and ProQuest/UMI a non-exclusive license to reproduce, loan, distribute, or sell single copies of my thesis by any means and in any form or format. These rights will in no way restrict republication of the material in any other form by you or by others authorized by you.

I would like permission to allow inclusion of the following material in my thesis in Figure 5. I need to include one image from this figure: "Figure 7.18.1 Schematic Illustration of a strategy of next-gen sequencing-based mapping and identification of EMS mutations" from this article UNIT 7.18 Next-Gen Sequencing-Based Mapping and Identification of Ethyl Methanesulfonate-Induced Mutations in Arabidopsis thaliana. Published Online: 1 OCT 2014 DOI: 10.1002/0471142727.mh0718s108

Current protocol of Molecular Biology

The material will be attributed through a citation.

Please confirm in writing or by email that these arrangements meet with your approval. I have also attached the pdf version of this article.

Sincerely

Md Jakir Hossain
MSc Candidate
Molecular Biology Unit
Department of Biology
University of Western Ontario
Agriculture and Agri-Food Canada
CURRICULUM VITAE

Name: Md Jakir Hossan

Post-secondary Education and Degrees:
The University of Western Ontario, London, Ontario, Canada

2013- present, M.Sc. Molecular Biology

Sher-e-Bangla Agricultural University, Dhaka, Bangladesh
2004-2008

B.Sc. (Honours), Agriculture

Honours and Awards:
Western Graduate Research Scholarship
2013-2015

Undergraduate Stipend, Sher-e-Bangla Agricultural University
2004-2009

Australian Government Work and Holiday Visa Award
2012-2013

Related Work Experience:
Teaching and Research Assistant
The University of Western Ontario
2013-2015

Research Assistant
Southern Crop Protection and Food Research Centre
Agriculture and Agri-food Canada
2013- Present

Research Assistant
Department of Horticulture, Sher-e-Bangla Agricultural University
Dhaka, Bangladesh
2010-2012
Publications:

