ANALYSIS OF THE GID1 FAMILY OF GIBBERELLIN RECEPTORS IN LAND PLANTS AND ELUCIDATION OF GIBBERELLIN-DRIVEN TRANSCRIPTIONAL PROGRAMS DURING SOYBEAN (*GLYCINE MAX*) GERMINATION

RAJESH KUMAR GAZARA

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE DARCY RIBEIRO – UENF

CAMPOS DOS GOYTACAZES – RJ

FEBRUARY 2019

ANALYSIS OF THE GID1 FAMILY OF GIBBERELLIN RECEPTORS IN LAND PLANTS AND ELUCIDATION OF GIBBERELLIN-DRIVEN TRANSCRIPTIONAL PROGRAMS DURING SOYBEAN (*GLYCINE MAX*) GERMINATION

RAJESH KUMAR GAZARA

Thesis submitted to the Centro de Biociências e Biotecnologia of Universidade Estadual do Norte Fluminense Darcy Ribeiro, as partial fulfillment of the requirements for obtaining the degree of Doctor in Biosciences and Biotechnology.

Advisor: Prof. Dr. Thiago Motta Venancio

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE DARCY RIBEIRO – UENF

CAMPOS DOS GOYTACAZES – RJ

FEBRUARY 2019

FICHA CATALOGRÁFICA
UENF - Bibliotecas
Elaborada com os dados fornecidos pelo autor.

G289 Gazara, Rajesh Kumar.

Analysis of the GID1 family of gibberellin receptors in land plants and elucidation of gibberellin-driven transcriptional programs during soybean (Glycine max) germination / Rajesh Kumar Gazara. - Campos dos Goytacazes, RJ, 2019.

127 f. : il.
Bibliografia: 13-18, 62-65, 98-102.

Tese (Doutorado em Biociências e Biotecnologia) - Universidade Estadual do Norte Fluminense Darcy Ribeiro, Centro de Biociências e Biotecnologia, 2019.
Orientador: Thiago Motta Venancio.

1. General Introduction. 2. Expansion and Diversification of GID1s. 3. Transcriptional landscape of soybean (Glycine max) embryonic axes during germination in the presence of PBZ. 4. General Discussion. I. Universidade Estadual do Norte Fluminense Darcy Ribeiro. II. Título.

CDD - 570

ANALYSIS OF THE GID1 FAMILY OF GIBBERELLIN RECEPTORS IN LAND PLANTS AND ELUCIDATION OF GIBBERELLIN-DRIVEN TRANSCRIPTIONAL PROGRAMS DURING SOYBEAN (*GLYCINE MAX*) GERMINATION

RAJESH KUMAR GAZARA

Thesis submitted to Centro de Biociências e Biotecnologia of Universidade Estadual do Norte Fluminense Darcy Ribeiro, as partial fulfillment of the requirements for obtaining the degree of Doctor in Biosciences and Biotechnology.

Advisor: Prof. Dr. Thiago Motta Venancio

Date Defended/Approved: 18/02/2019.

Examining committee:

Prof. Dr. Thiago Motta Venancio - UENF

Prof. Dr. Katia Valevski Sales Fernandes - UENF

Prof. Dr. Gonçalo Apolinário de Souza Filho - UENF

Prof. Dr. Robson Francisco de Souza - USP

Prof. Dr. Rodrigo Nunes da Fonseca - UFRJ

CAMPOS DOS GOYTACAZES – RJ
February 2019

ACKNOWLEDGEMENTS

This thesis is the pinnacle of my journey in obtaining Ph.D. degree. Many people have earned my gratitude for their contribution in my Ph.D. journey. This doctoral dissertation would not have come to successful completion without their support. I would like to take an opportunity to be grateful to all of those who made this thesis possible and a memorable experience for me.

First and foremost, I would like to express my sincere gratitude and special thanks to my Ph.D. advisor, Dr. Thiago Motta Venancio, for his support, patience and consistent encouragement. I appreciate all his contributions of time, ideas, and efforts to make my Ph.D. experience dynamic and inspiring that I have received throughout my Ph.D. He always kept me on the right track to solve the research problems. He boosted me up and encouraged me to bring the best out of me. My warm appreciation goes to him for all his help and guidance.

I would like to thank Dr. Elenir and Dr. Eduardo Oliveira for the support in the experimental stage, without which this work would not be possible. I am also grateful to Dr. Arnoldo Rocha Facanha, course coordinator of the Center of Bioscience and Biotechnology, Dr. Olga Machado (Ex-course coordinator), and the secretaries Beatriz Almeida and Marlene dos Santos for their endless support for all postgraduate students in Biosciences and Biotechnology.

I acknowledge FAPERJ, CAPES and CNPq for financial support. I also thank the Life Sciences Core Facility (LaCTAD) of State University of Campinas (UNICAMP) for library preparation and RNA sequencing.

I appreciate my project defence committee members, Dr. Kátia Fernandes, Dr. Gustavo Rezende, and Dr. Jorge Hernandez Fernandez. During my project defence, I received extremely valuable comments and insights from them that greatly enriched my work. I would like to thank my qualification exam committee, Dr. Clícia Grativol and Dr. Gonçalo Souza Filho for their precious suggestions.

I extend my thank to all the present lab mates: Filipe, Kanhu, Franscisnei, Fabricio (big), Fabricio (small), Hemanuel, Daniela, Dayana, Késia, Isabella, Luiz and past lab mates: Daniel and Lupis. They always helped me during my hard

time and provided a pleasant atmosphere to me. I thank their constructive suggestions provided during our rational discussions. I wish special thank to Kanhu for his discussions on science and non-science topics. I also needed to thank him because he inspired me to do things with full of passion which mainly include science.

I would like thank to my previous lab mates, Dr. Shadab and Dr. Sandhya. I cannot forget to mention them because I started my research career with them and they provided full support and guidance at any time whenever I required the most during my earlier research career. It was a great time working with them and learning new essential skills which helped me to complete this Ph.D. journey.

There are not enough words to express my deepest gratitude towards my parents. I am deeply indebted towards my father, Mr. Ghanshyam Das Gazara and my mother, Mrs. Jasoda Gazara for providing me strength and love. They provided me so many opportunities throughout my life to grow, fly and live my dreams. They are my greatest strength and I find myself extremely fortunate to have them as my parents.

I convey my admiration to my brother Gulshan Gazara and my bhabhi Rekha Gazara for their constant support, love and determined belief. The talks on Skype and whatsapp with my niece Bhavika and nephew Karan always re-energized me. I thank my siblings for always believing in me and for the warm memories I have of my childhood years. Their love inspired me in my whole-life. Though thank you doesn’t seem sufficient, but it is said with appreciation and respect. I own my everlasting gratitude to my Mama Ji, Laxman Mawani, to share his knowledge which facilitated me to face problems with smile and positive attitude throughout my life.

At last I would like to thank everyone whose direct or indirect involvement was important for the successful completion of the thesis.

Above all, I thank God for giving me the strength, insight and, health to carry out this research task and enabling me to its completion.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATION

ABA: Abscisic acid
ABF: Abscisic acid responsive element-binding factor
ACS: 1-Amino-cyclopropane-1-carboxylate synthases
AGP: Arabinogalactan protein
AHP: His-containing phosphotransfer protein
AMY: Alpha amylase-like
ARFA: alpha-L-arabinofuranosidase
A-RR: Type-A response regulator
ATP: Adenosine triphosphate
Aux/IAA: Auxin/Indole-3-Acetic Acid
BAK1: BRI1-associated receptor kinase 1
BES1: BRI1-ethyl methanesulfonate-suppressor 1
BGLU: Beta glucosidase
bHLH: Basic helix-loop-helix
BIN2: Brassinosteroid-insensitive 2
BKI1: BRI1 kinase inhibitor
BLAST: Basic local alignment search tool
BRI1: Brassinosteroid-insensitive 1
B-RR: Type-B response regulator
BRs: Brassinosteroid
BSK: Brassinosteroid signaling kinases
BSU: BRI1-suppressor
BXL: Beta-xylosidase
bZIP: Basic Leucine Zipper
BZR1: Brassinazole-resistant 1
C2H2: C2H2 zinc finger
CK: Cytokinin
COI1: CORONATINE INSENSITIVE 1
CPS: Ent-copalyl diphosphate synthase

CRE1: CYTOKININ RESPONSE 1

DEGs: Differentially expressed genes

DNA: Deoxyribonucleic acid

ent-CDP: ent-copalyl diphosphate

ERF: Ethylene response factor

EXP: Expansin

FDR: False discovery rate

FLA: FASCICLIN-like arabinogalactan-protein

FPKM: Fragment per kilobase of transcript per million mapped reads

GA: Gibberellin

GEA: Gene expression atlas

GGDP: Geranyl geranyl diphosphate

GH: Glycosyl hydrolase family protein

GH3: Auxin-responsive Gretchen Hagen3

GID1: GIBBERELLIN INSENSITIVE DWARF1

GO: Gene ontology

GRP: Glycine-rich protein

GSDS: Gene structure display server

GST: Glutathione-S-transferase

HAI: Hours of imbibition

HD-ZIP: Homeodomain-leucine zipper

HRGP: Hydroxyproline-rich glycoprotein family protein

HSLs: Hormone sensitive lipases

HSPs: Heat shock proteins

IA Ox N-oxide: Indol-3-acetaldoxime N-oxide

IAA: Indol acidic acid

IAOx: Indol-3-acetaldoxime

IG: Indole glucosinolates

IPA: Indole-3-pyruvate

JA: Jasmonic acid

JAR1: JASMONATE RESISTANT1

JAZ: JASMONATE ZIM-domain

JIN1: JASMONATE INSENSITIVE 1

KAO: Ent-kaurenoic acid oxidase

KEGG: Kyoto Encyclopedia of Genes and Genomes

KO: Ent-kaurene oxidase

KOBAS: KEGG Orthology-Based Annotation System

KS: Ent-kaurene synthase

K_s: Synonymous substitution rate

LAC: Laccase

LSD: LESION SIMULATING DISEASE

MCMC: Markov chain monte carlo

MEME: Multiple expectation maximization for motif elicitation

ML: Maximum likelihood

MSA: Multiple sequence alignment

MYB: Myelobastosis

NAC: No apical meristem (NAM), ATAF, and CUC (cup-shaped cotyledon) family

NCED: 9-cis-epoxycarotenoid dioxygenase

PAML: Phylogenetic analysis by maximum likelihood

PBZ: Paclobutrazol

PCA: Principal component analysis

PDB: Protein data bank

PER: Peroxidase superfamily protein

PL: Pectin lyase-like superfamily protein

PMEI: Pectin methylesterase inhibitor superfamily protein

PP2C: Phosphatase 2C

PROMALS3D: PROfile Multiple Alignment with predicted Local Structures and 3D constraints

PRP: Proline-rich protein

PYL: PYR-like

PyMOL: Python-enhanced molecular graphics tool

PYR: Pyrabactin resistance

RAxML: Randomized axelerated maximum likelihood

REViGO: Reduce visualization gene ontology

RNA: Ribonucleic acid

RPKM: Reads per kilobase of transcript per million mapped reads

RuBisCO: Ribulose-1,5-bisphosphate carboxylase/oxygenase

SAUR: Small auxin upregulated RNA

SIM: Small ubiquitin-like modifier -interaction motif

SnRK2: Sucrose non-fermenting 1-related protein kinases subfamily 2

SUMO: Small ubiquitin-like modifier

TAR: Tryptophan aminotransferase related

TFs: Transcription factors

WAK: Wall associated kinase

WGD: Whole-genome duplication

WGT: Whole genome triplication

XTH: Xyloglucan endotransglucosylase/hydrolase

ZF-HD: Zinc finger homeodomain

RESUMO

A giberelina (GA) é um fitormônio essencial que regula positivamente a germinação de sementes. Esse hormônio controla uma ampla variedade de genes através da interação com os receptores GIBBERELLIN INSENSITIVE DWARF1 (GID1), que evoluíram de uma grande família de lipases sensíveis a hormônios. A interação GA-GID1 promove a degradação dos repressores transcricionais DELLA pela via do proteassomo 26S e, consequentemente, a ativação da sinalização por GA. Os receptores GID1 de eudicotiledôneas podem ser divididos nos subgrupos GID1ac e GID1c. Entretanto, diversos aspectos acerca da evolução e diversificação funcional dessas subfamílias permanecem desconhecidos. Além disso, como a giberelina estimula a germinação de sementes, também é essencial compreender esse processo a nível molecular, investigando os genes que são regulados por GA, por exemplo. A presente tese de doutorado inclui dois estudos relacionados que envolvem abordagens de genômica comparativa e RNA-seq. Os objetivos maiores desses estudos são ajudar a compreender a história evolutiva dos receptores GID1 em plantas terrestres e descobrir os mecanismos de transcrição regulados por GA em sementes de soja (*Glycine max*) durante a germinação. No primeiro estudo, que se trata da evolução da família GID1 em plantas terrestres, nós descobrimos que a duplicação completa do genoma contribuiu para a expansão e diversificação de ambas as subfamílias (isto é, GID1ac e GID1b) em eudicotiledôneas. Este estudo revelou, ainda, características estruturais compartilhadas e divergentes entre os subgrupos GID1ac e GID1b em eudicotiledôneas que fornecem insights sobre suas funções. Notadamente, nós encontramos importantes resíduos divergentes no sítio de ligação de GA a GID1b que poderiam conferir maior afinidade a GA. Os níveis de expressão gênica em diferentes espécies endossaram que GID1b especializou-se em condições de baixas concentrações de GA, como raízes. O segundo estudo buscou identificar genes induzidos por GA em eixos embrionários de sementes de soja durante a germinação. O transcriptoma de soja foi analisado em um experimento de RNA-*seq* ao longo do tempo (12, 24 e 36 horas após a embebição) na presença de paclobutrazol

(PBZ), um inibidor da biossíntese de GA. Genes relacionados à modificação da parede celular, biossíntese e sinalização hormonal foram diferencialmente expressos e analisados a fundo através da integração de dados da literatura. Este estudo também mostrou que as famílias de fatores de transcrição MYB, bHLH e bZIP são alvos de GA que regulam os mecanismos de transcrição durante a germinação.

Palavras chave: Giberelina; GID1; Paclobutrazol; Soja; Transcriptoma, Expressão gênica

ABSTRACT

Gibberellin (GA) is an essential phytohormone that positively regulates seed germination. It controls a wide variety of genes by interacting with GIBBERELLIN INSENSITIVE DWARF1 (GID1) receptors, which evolved from a large family of Hormone Sensitive Lipases. GA-GID1 interaction promotes the degradation of DELLA transcriptional repressors by the 26S proteasome pathway and, hence, the activation of GA signaling. Eudicot GID1s can be separated in the GID1ac and GID1b subgroups. However, several aspects of the evolution and functional diversification of these subfamilies remain unknown. Further, because GA enhances seed germination, it is also essential to understand this process at the molecular level, for example by investigating the genes that are regulated by GA. The present doctoral thesis comprises two related studies involving comprehensive comparative genomics approaches and high-throughput RNA sequencing. The ultimate goals of these studies are to help understand the evolutionary history of the GID1 family in land plants and to uncover the GA-regulated transcriptional program in germinating soybean (*Glycine max*) seeds. In the first study, regarding the evolution of the GID1 family in land plants, we found that whole-genome duplication contributed to the expansion and diversification of both subfamilies (i.e. GID1ac and GID1b) in eudicots. This study further revealed shared and divergent structural features between the GID1ac and GID1b subgroups in eudicots that provide mechanistic insights on their functions. Remarkably, we found important divergent residues in the GID1b GA-binding pocket that could provide increased GA affinity. Gene expression in several species supported that GID1b has specialized in conditions of low GA concentrations (e.g. roots). The second study aimed to identify GA responsive genes in the embryonic axes of germinating soybean seeds. The transcriptome was assessed by a time-course RNA-Seq experiment (12, 24 and 36 hours after germination, HAI) in the presence of paclobutrazol (PBZ), a GA biosynthesis inhibitor. Genes related to cell wall modification, hormone biosynthesis and signaling were differentially expressed and analyzed in depth by integrating primary literature data. This study also showed the MYB, bHLH and bZIP

transcription factors are probable downstream GA targets that drive the GA transcriptional programs during germination.

Keywords: Gibberellin; GID1; Paclobutrazol; Soybean; Transcriptome, Gene expression

Chapter 1 General introduction

1.1 INTRODUCTION

1.1.1 History of soybean domestication

The beginning of soybean domestication has been a topic of intense debate for decades. It is believed that cultivated soybean was domesticated from wild soybean (G. soja Sieb. & Zucc.) in China ~5,000 years ago and later introduced to Korea, and then to Japan ~2,000 years ago, to North America in 1765, and to Central and South America during the first half of the last century (Wilson 2008). Based on morphological, cytogenetic, and biochemical evidence, different regions of China were suggested as the single center of soybean domestication (Broich and Palmer 1981, Hymowitz 2004, Hymowitz and Kaizuma 1981). Based on molecular studies of hundreds of markers and accessions, the Yellow River basin (Li et al. 2010) and the Yangtze region (Southern China) (Guo et al. 2010) were proposed as the origin of soybean domestication. On the other hand, chloroplast sequence variation and archaeological evidence indicated the southern areas of Japan and China as secondary centers of domestication (Lee et al. 2011, Xu et al. 2002). However, a single soybean domestication event is supported by whole genome re-sequencing data (Chung et al. 2014, Lam et al. 2010, Zhou et al. 2015). Using high-density SNP data, Wang and his group suggested the domestication center as northern and central China (Wang et al. 2016), whereas another recent study that used specific-locus amplified fragment sequencing data has proposed central China surrounding the Yellow River as domestication center (Han et al. 2016).

Finally, long-lasting debate regarding soybean domestication came to a conclusion with the complex hypothesis (Sedivy et al. 2017), which combined the results of two different studies: i) whole genome comparison of one wild soybean ecotype to one soybean cultivar (Kim et al. 2010) and ii) pan-genome comparison of 7 wild soybean ecotypes (Li et al. 2014). The complex hypothesis

states that before the domestication of soybean, the ancestor of domesticated soybean first diverged from *G. soja* 0.27 (Kim et al. 2010) or 0.8 million years ago (Li et al. 2014), by creating an intermediate species, *G. gracilisa*, which represents a *G. soja/G. max* complex. Therefore, it can be assumed that the early-domesticated *G. soja* or *G. soja/G. max* complex introduced from China to Korean and Japan, and later experience different domestication event. Nevertheless, It is believed that *G. max* was emerged from *G. soja* or *G. soja–G. max* complex through a long and slow domestication process (Sedivy et al. 2017).

1.1.2 Soybean genome and transcriptomes

Soybean is an economically important crop mainly due to its protein (~38%) and oil (~20%) contents (Hou et al. 2009). Soybean is the largest source of animal protein feed and the second largest source of vegetable oil, after palm oil (http://www.neoda.org.uk). According to United States Department of Agriculture, to meet the growing global needs for food, animal feed and biofuels, soybean production has been significantly increased over the past decade, from 212 million tons in 2008 to over 300 million tons in 2017-2018. USA (119.5 million tons), Brazil (115 million tons) and Argentina (40 million tons) are top most soybean producers in the world followed by China (14.2 million tons) and India (9 million tons) (https://www.fas.usda.gov/) (Figure 1.1). One of the key factors for the Brazilian competitiveness in soybean production is the optimized use of nitrogen-fixing Bradyrhizobium strains, which form a well-characterized symbiotic association with soybean roots. Consequently, chemical nitrogen fertilization in soybean farms is extremely reduced in Brazil (Chang et al. 2015).

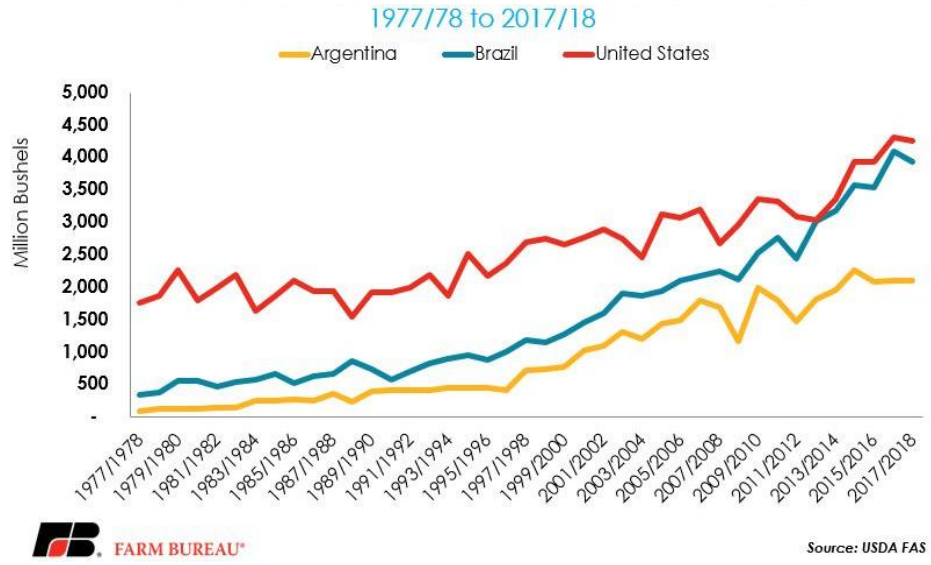


Figure 1.1 USA, Brazil and Argentina soybean production (figure source: USDA FAS).

DNA sequencing revolutionized nearly all fields of biology (França et al. 2002). Around 2007, the release of a new generation of sequencing technologies (e.g. Illumina/Solexa, ABI/SOLiD, 454/Roche, and Helicos) dramatically changed DNA sequencing and genomics. The development of these second (or next) generation sequencing methods has been fueled over the past 12 years, mainly because of the sequencing of many genomes, including the human genome (Grada and Weinbrecht 2013). One of the major applications of next-generation sequencing is transcriptomics (Morozova and Marra 2008). A transcriptome is the set of all RNAs, including mRNAs, rRNAs, tRNAs and other non-coding RNAs expressed in a cell (Peano et al. 2013). Transcriptome studies are essential to understand expressed gene complement of any organism under a particular condition or developmental stage. With the current RNA-sequencing technologies, it is now possible to identify differentially expressed genes (DEGs) in various conditions, with greater precision and reproducibility if compared with microarrays (Marioni et al. 2008).

Due to their large size, polyploidy and abundant repetitive regions, assembling plant genomes is typically more challenging than animal and

microorganism genomes. A whole-genome shotgun approach was used to sequence the ~1.1 gigabase (Gb) soybean genome (Williams 82, Glyma1.01) (Schmutz et al. 2010). Most of the genome was captured in 20 chromosomes, comprising 397 scaffolds with well-organized physical maps covering 937.3 Mb. Additionally, 1,148 unanchored sequence scaffolds comprise 17.7 Mb, mainly filled with repetitive sequences. The initial Williams 82 genome contains 46,430 protein-coding genes, 4,991 single nucleotide polymorphisms (SNPs) and 874 simple sequence repeats. A second version of soybean genome (Wm82.a2.v1) was later released with several improvements, including the prediction of 56,044 protein-coding loci and 88,647 transcripts (Song et al. 2016). The soybean genome has been strongly affected by two polyploidization events, one at the base of the legume (Papilionoideae) lineage and other at the base of the Glycine genus (Schmutz et al. 2010, Severin et al. 2011).

Taking the advantage of a good reference genome and modern RNA-Seq technologies, multiple transcriptome studies have been published over the past few years (Bellieny-Rabelo et al. 2016, Libault et al. 2010, Libault et al. 2010, Prince et al. 2015, Severin et al. 2010, Song et al. 2016, Wang et al. 2014). For example, a key study reported transcriptome profiles in 14 different tissues, including leaf, flower, pod, pod-shell, root, nodules and seven seed developmental stages (Severin et al. 2010). This work provided an important initial soybean transcriptome atlas. Similarly, transcriptome profiles of 14 different tissues, mainly underground tissues, were reported by another research group around the same time (Libault et al. 2010). This work supported the transcription of 55,616 annotated genes, out of which 13,529 are putative pseudogenes (Libault et al. 2010). Dozens of other soybean transcriptome studies have been published afterwards, covering virtually all lifecycle stages and many stress conditions. In order to understand the molecular mechanisms of canopy-wilting and response to drought, transcriptome sequencing was performed in drought-susceptible Pana (DS) and drought-tolerant PI 567690 (DT) cultivars (Prince et al. 2015). Other studies identified genes related to drought and flood stresses in

roots and leaves (Chen et al. 2016, Song et al. 2016). Recently, our group has explored soybean transcriptome during germination, uncovering many aspects of metabolic reactivation, cell wall remodeling and hormonal regulation (Bellieny-Rabelo et al. 2016).

1.1.3 Seed germination

Seed germination is a critical process in plant life-cycle. It determines the successful crop production. Seed germination starts with water uptake (imbibition) by dry seeds and ends with the emergence of embryonic axis (Bewley 1997, Bewley et al. 2013). In general, every seed is divided into three major compartments: 1) seed coat, which is an outer most layer that protects embryo and endosperm, and also play important role in controlling factors which initiate seed germination, 2) an embryo, which will become new plant after germination process and 3) endosperm, a tissue which provides energy and nutrient for embryo to grow (Bewley 1997, Bewley et al. 2013). Germination of most eudicot seeds comprises three phases: quick water uptake (phase I), also known as seed rehydration stage; lag phase (phase II) and; a second rapid water uptake phase (phase III) (Bewley 1997, Bewley et al. 2013). Morphologically seed germination is divided into testa rupture, endosperm rupture and radicle protrusion (Bewley 1997, Müller et al. 2006). Previous studies have shown that seed germination is regulated by multiple factors such as temperature, water, soil type, oxygen, light and plant hormones (Bewley 1997, Bewley et al. 2013).

1.1.3.1 Transcription during germination

Intensive metabolic changes take place during phase I and II, resulting in radicle protrusion. Seed dehydration and rehydration during maturation and imbibitions, respectively, are linked with oxidative stress, resulting in DNA damage. Therefore, during germination, DNA repair is an essential step, mainly conducted by DNA ligase via joining of single- and double-strand breaks. *De novo* nuclear and mitochondrial DNA synthesis also take place in the radicle shortly upon imbibition (De Castro et al. 1995). All components required for transcription and

translation (except polysomes) are already available in dry seeds. Polysome formation takes place early during germination, in the transition from quiescence to a fully imbibed and metabolically active state (Dommes and Walle 1990).

Transcription initiates during the first few hours after the imbibition, as well as the synthesis of enzymes involved in glycolysis, pentose phosphate pathway and respiration (Botha et al. 1992). During imbibition, there is an increasing intake of oxygen (Logan et al. 2001), resulting in accumulation of ROS, which is important for germination, endosperm weakening and programmed cell death (El-Maarouf-Bouteau and Bailly 2008).

1.1.4 Hormones in seed germination control

Phytohormones concentrations and interactions play important regulatory roles during seed germination (Kucera et al. 2005). GA is perhaps the most well-studied promoter of seed germination and the *de novo* GA biosynthesis in imbibed seeds is essential for germination (Ikuma and Thimann 1960, Yomo and Iinuma 1966). Severe GA-deficient mutants such as *ga1-3* and *ga2-1* fail to germinate (Koornneef and van der Veen 1980). GA is important during early and late germination. Although present in dry and after-ripened seeds, bioactive GA concentrations increase during late germination (Ogawa et al. 2003). In contrast, endogenous ABA content decreases during imbibition and early phase II, what is necessary for the completion of seed germination (Müller et al. 2006). ABA and GA are also antagonize each other in their influences on developmental processes (e.g. flowering) (Razem et al. 2006). Because of the rapid ABA degradation, GA/ABA ratio increased during seed germination (Ogawa et al. 2003). However, it was observed that exogenous GA application did not affect the ABA content in GA-deficient (ga1-3) *Arabidopsis* mutant during early seed germination (Ogawa et al. 2003). Further, expression of *GA3ox1* transcripts were decreased in the *cyp707a2* ABA-overproducing *Arabidopsis* mutant (Yano et al. 2009).

Ethylene also promotes germination, mainly through inhibition of ABA signaling. Ethylene concentration increases during seed germination of several plants, such as wheat, corn, soybean and rice (Pennazio and Roggero 1991, Zapata et al. 2004). The enzyme 1-aminocyclopropane-1 carboxylic acid oxidase, which is essential for ethylene production, was also shown to enhance radicle protrusion (Petruzzelli et al. 2000, Petruzzelli et al. 2003).

Brassinosteroids (BRs) are also important ABA antagonists, promoting embryo growth and seed germination (Finch-Savage and Leubner-Metzger 2006, Leubner-Metzger 2001). After interaction with BR, a leucine-rich-repeat receptor like kinase (BRI1) binds with BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1 (BAK1) and phosphorylates BRI1 kinase inhibitor 1 (BKI1) (Li et al. 2002, Nam and Li 2002, Wang et al. 2001). This event activates trans-phosphorylation between BRI1 and BAK1, releasing phosphorylated Brassinosteroid-Signaling Kinases (BSKs) (Wang and Chory 2006). These phosphorylated BSKs induce BR signaling (Li and Jin 2007). Further, ABA can rapidly inhibit BR signaling and change the expression of BR-responsive genes (Zhang et al. 2009).

The role of auxin in seed germination is largely unclear, with inhibitory effects reported in wheat (Ramaih et al. 2003) and soybean (Shuai et al. 2017). Nevertheless, expression of genes related to polar auxin transport and genes encoding CYP79B2 and CYP79B3, necessary for formation of indoleacetic acid, were up-regulated by exogenous GA during germination of *Arabidopsis* seeds (Ogawa et al. 2003). Cytokinins were also shown to release seed dormancy and enhance seed germination under various stress conditions (Atici et al. 2005, Khan and Ungar 1997, Nikolić et al. 2006, Peleg and Blumwald 2011). Since the focus of this thesis is on GA signaling and GA-responsive gene regulation, the next section addresses GA biosynthesis, regulation and signaling in more detail.

1.1.4.1 Gibberellins

GAs are a large family of diterpenoid compounds that can be divided in two groups with regard to their number of carbons: ent-gibberellane (C20) and 20-nor-ent-gibberellane (C19) carbon skeletons. In C_{19} GAs, carbon C-20 released in form of CO_2 and lactone ring is formed between carbon C-19 and carbon C-10 (MacMillan 2001, Sponsel and Hedden 2010). In 1935, GA was first isolated from *Gibberella fujikuroi* (*G. fujikuroi*, reclassified as *Fusarium fujikuroi*), a fungal rice pathogen that causes the disease known as 'bakanae' or 'foolish seedling' (Yabuta 1935). Since the 1950s, different studies demonstrated the activity of GAs in regulating plant growth, resulting in the GA classification as plant hormones (Brian et al. 1954, Brian and Hemming 1955, Phinney 1956, Radley 1956). Currently, ~136 GAs are known in plants, fungi and bacteria (http://www.plant-hormones.info/gibberellin_nomenclature.htm), although most of them are precursor or inactive forms (MacMillan 2001, Sponsel and Hedden 2010). The most active GAs in higher plants are GA_1, GA_3 and GA_4. GA_1 and GA_4 are typically abundant in higher plants, whereas GA_3 a major GA product of *F. funikuroi,* which is produced commercially for agronomic, horticultural and other scientific uses (Hedden and Thomas 2012). GAs are essential regulators of multiple plant growth and development processes, including seed germination, root and stem elongation, leaf expansion, flower and fruit development (Olszewski et al. 2002, Tanimoto and Hirano 2013).

1.1.4.2 Gibberellins Biosynthesis

In higher plants, the enzymes involved in the biosynthesis of bioactive GAs are categorized into three classes; 1) terpene synthases (TPSs), 2) cytochrome P450 monooxygenases (P450s), and 3) 2-oxoglutarate–dependent dioxygenases (2ODDs) (Graebe 1987) (Figure 1.2). GA biosynthesis starts with geranylgeranyl diphosphate (GGPP) production from isopentenyl diphosphate (IPP), via the terpenoid biosynthesis pathway. GGPP is converted in tetracyclic hydrocarbon ent-kaurene in plastids, in a two-step reaction, in which GGPP is catalyzed by ent-copalyl diphosphate synthase (CPS) and ent-kaurene synthase (KS), with an

intermediate, ent-copalyl diphosphate (Hedden and Thomas 2012, Olszewski et al. 2002). Next, ent-kaurene is oxidized by cytochrome P450 mono-oxygenases to form GA_{12}. The synthesis of GA_{12} requires several oxidation steps, catalyzed by two mono-oxygenases; ent-kaurene oxidase (KO) and ent-kaurenoic acid oxidase (KAO), localized in the endoplasmic reticulum. The ent-kaurene is converted into ent-kaurenoic acid through ent-kaurenol and ent-kaurenal by KO. Oxidation of ent-kaurenoic acid to form GA_{12} is catalysed by KAO (Hedden and Thomas 2012, Olszewski et al. 2002). GA_{12} is converted into bioactive form by 2ODDS, via oxidation of C-20 and C-3 by GA20 oxidases and GA3 oxidases, respectively (Figure 1.2). As a consequence of these multiple steps, various GA intermediates are found in cytoplasm before bioactive forms are harnessed.

To regulate the effective concentration of bioactive GA, plants can also inactivate GA by means of 2β-hydroxylation reactions catalyzed by GA2-oxidase (GA2ox) (Hedden and Thomas 2012, Olszewski et al. 2002). Hence, the concentration of bioactive GA in a given situation depends on a balance between synthesis and deactivation. Another deactivation mechanism including epoxidation of non-13-hydroxylated GA in rice (Zhu 2006) and methylation of GA in Arabidopsis (Varbanova et al. 2007) also have been identified.

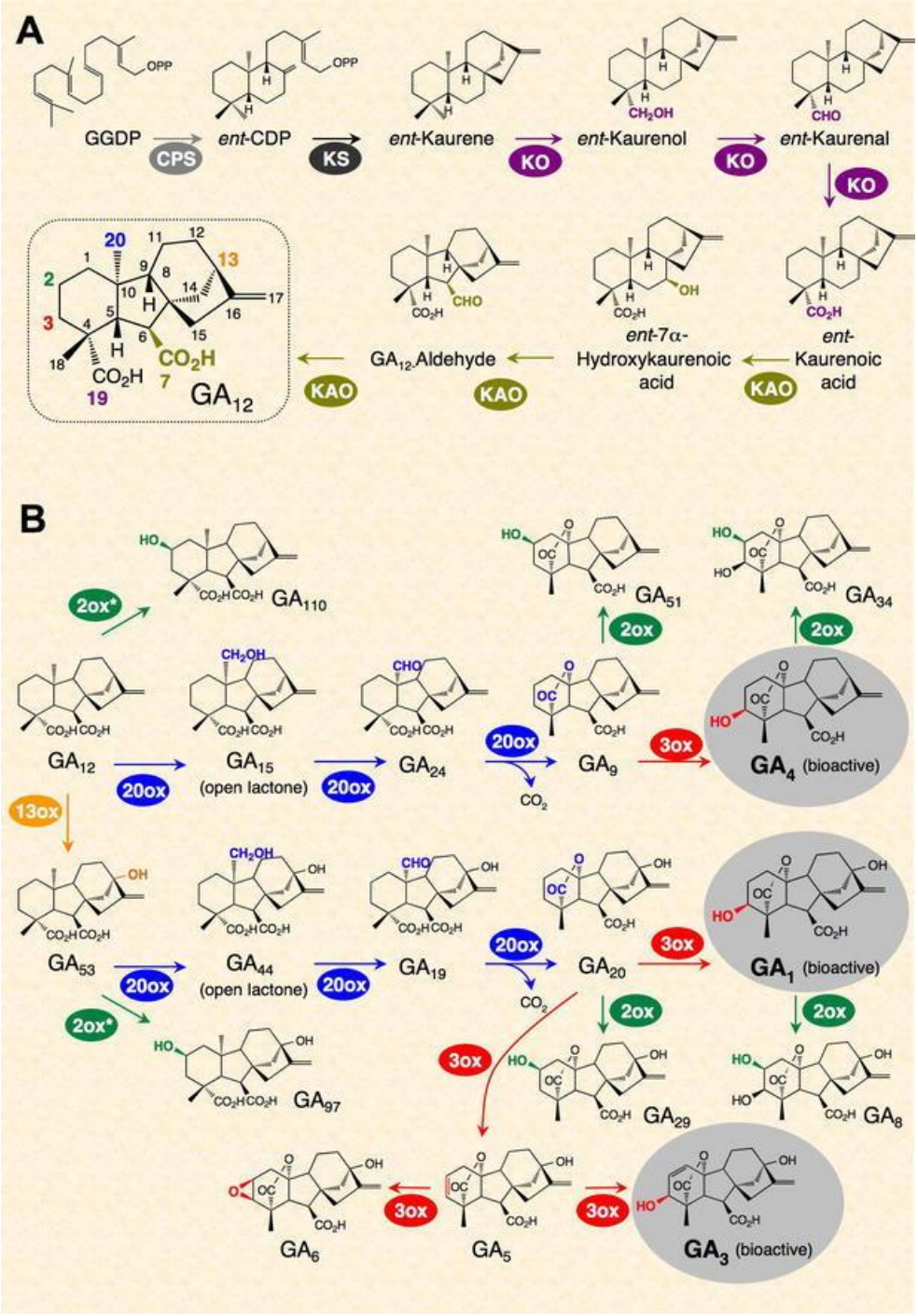


Figure 1.2 GA Biosynthetic and Catabolic Pathways in Plants.
(A) Synthesis of GA12 from GGDP. (B) GA biosynthesis and deactivation (by GA2ox) pathways from GA12. The three active GAs are highlighted in grey circles. GA7 (13-nonhydroxy GA3), another active GA, is synthesized from GA9 (not shown) (Sun 2008).

1.1.4.3 GA regulation and signaling

The GA levels in higher plants are maintained by a feedback mechanism. It has been shown that *GA20ox* and *GA3ox* are down-regulated by GA (Olszewski et al. 2002), as opposed to early GA biosynthesis genes (e.g. *CPS*, *KS* and *KO*) (Helliwell et al. 1998). In *Arabidopsis*, *GA2ox* is up-regulated by GA treatment (Thomas et al. 1999). Other factors that regulate GA metabolism are light (Oh et al. 2006), temperature (Penfield et al. 2005), stress (Yamaguchi 2008), tissue type, transport, developmental stage, levels of GA conjugates (Schneider and Schliemann 1994, Yamaguchi 2008), other plant hormones (e.g. auxin and ABA) (Ross et al. 2001, Yamaguchi 2008).

The main GA signaling pathway involves the recognition of bioactive GA (i.e. GA_3) by GIBBERELLIN INSENSITIVE DWARF1 (GID1) receptors, originally identified in rice (OsGID1) using genetic approaches (Ueguchi-Tanaka et al. 2005). The interaction with GA promotes a conformational change in GID1, increasing its affinity for DELLA proteins, which are transcriptional co-repressors of GA signaling (Richards et al. 2001). The GA-GID1-DELLA complex is recognized by the SCF^{SLY1} ubiquitin ligase complex, which ubiquitinates and induces the proteasomal degradation of DELLA (Dill et al. 2004, Gomi et al. 2004, McGinnis et al. 2003). Therefore, the down-regulation of DELLA is the process that ultimately triggers canonical GA effects (Fleet and Sun 2005) (Figure 1.3). Other known positive GA regulators are *DWARF1* (*D1*) (Ueguchi-Tanaka et al. 2000), *PHOTOPERIOD RESPONSIVE 1* (*PHOR1*) (Dale 1998, Willert and Nusse 1998), MYB Transcription Factors (Woodger et al. 2003), *SLEEPY* and *PICKLE* (*PKL*) (Ogas et al. 1997, Ogas et al. 1999). On the other hand, DELLA (Peng et al. 1997, Peng et al. 1999, Silverstone et al. 1998, Wen and Chang 2002), *SPINDLY* (*SPY*) (Filardo and Swain 2003, Jacobsen and Olszewski 1993) and *SHORT INTERNODES* (*SHI*) (Fridborg et al. 2001) are some of the negative regulators of GA biosynthesis.

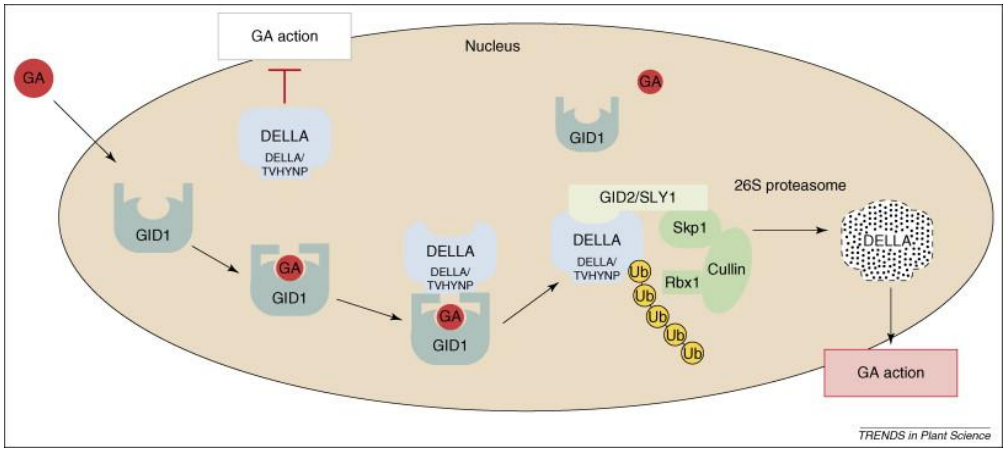


Figure 1.3 Model for GA action through DELLA degradation.
In the absence of GA, GA action is repressed by DELLA. When GA is present, the GID1 binds GA. The GID1–GA complex interacts with DELLA via DELLA´s TVHYNP motifs, resulting in the recognition of DELLA by the SCF^{SLY1} E3-ligase complex (consisting of Skp1, Cullin, F-box protein, and Rbx1). Upon polyubiquitination, DELLA is degraded through the 26S proteasome pathway and the GA response is released. Abbreviations: Ub, ubiquitin (Hirano et al. 2008).

1.1.5 REFERENCES

Atici Ö, Ağar G, Battal P. 2005. Changes in phytohormone contents in chickpea seeds germinating under lead or zinc stress. Biologia Plantarum.49:215-222.

Bellieny-Rabelo D, Alves Gamosa de Oliveira E, da Silva Ribeiro E, Pessoa Costa E, Elenir Amâncio Oliveira A, Motta Venancio T. 2016. Transcriptome analysis uncovers key regulatory and metabolic aspects of soybean embryonic axes during germination. Scientific Reports.6:36009.

Bewley JD. 1997. Seed Germination and Dormancy. THE PLANT CELL ONLINE.9:1055-1066.

Bewley JD, Bradford KJ, Hilhorst HWM, Nonogaki H. 2013. Seeds: Physiology of Development, Germination and Dormancy.

Botha FC, Potgieter GP, Botha A-M. 1992. Respiratory metabolism and gene expression during seed germination. Plant Growth Regulation.11:211-224.

Brian PW, Elson GW, Hemming HG, Radley M. 1954. The plant growth promoting properties of gibberellic acid, a metabolic product of the fungus gibberella fujikuroi. Journal of the Science of Food and Agriculture.5:602-612.

Brian PW, Hemming HG. 1955. The Effect of Gibberellic Acid on Shoot Growth of Pea Seedlings. Physiologia Plantarum.8:669-681.

Broich SL, Palmer RG. 1981. Evolutionary studies of the soybean: the frequency and distribution of alleles among collections of Glycine Max and G. Soja of various origin. Euphytica.30:55-64.

Chang W-s, Lee H-i, Hungria M. 2015. Soybean Production in the Americas. In: Principles of Plant-Microbe Interactions. Springer, Cham. p. 393-400.

Chen W, Yao Q, Patil GB, Agarwal G, Deshmukh RK, Lin L, Wang B, Wang Y, Prince SJ, Song L, et al. 2016. Identification and Comparative Analysis of Differential Gene Expression in Soybean Leaf Tissue under Drought and Flooding Stress Revealed by RNA-Seq. Frontiers in Plant Science.7.

Chung W-H, Jeong N, Kim J, Lee WK, Lee Y-G, Lee S-H, Yoon W, Kim J-H, Choi I-Y, Choi H-K, et al. 2014. Population Structure and Domestication Revealed by High-Depth Resequencing of Korean Cultivated and Wild Soybean Genomes. DNA Research.21:153-167.

Dale TC. 1998. Signal transduction by the Wnt family of ligands. The Biochemical journal.329 (Pt 2:209-223.

De Castro RD, Zheng X, Bergervoet J, De Vos C, Bino RJ. 1995. [beta]-Tubulin Accumulation and DNA Replication in Imbibing Tomato Seeds. Plant physiology.109:499-504.

Dill A, Thomas SG, Hu J, Steber CM, Sun T-P. 2004. The Arabidopsis F-box protein SLEEPY1 targets gibberellin signaling repressors for gibberellin-induced degradation. The Plant cell.16:1392-1405.

Dommes J, Walle CVD. 1990. Polysome formation and incorporation of new ribosomes into poiysomes dming germination of the embryonic axis of maize. Physiologia Plantarum.79:289-296.

El-Maarouf-Bouteau H, Bailly C. 2008. Oxidative signaling in seed germination and dormancy. Plant signaling & behavior.3:175-182.

Filardo FF, Swain SM. 2003. SPYing on GA Signaling and Plant Development. Journal of Plant Growth Regulation.22:163-175.

Finch-Savage WE, Leubner-Metzger G. 2006. Seed dormancy and the control of germination. New Phytologist.171:501-523.

Fleet CM, Sun TP. 2005. A DELLAcate balance: The role of gibberellin in plant morphogenesis. Current Opinion in Plant Biology.8:77-85.

França LTC, Carrilho E, Kist TBL. 2002. A review of DNA sequencing techniques. Quarterly reviews of biophysics.35:169-200.

Fridborg I, Kuusk S, Robertson M, Sundberg E. 2001. The Arabidopsis protein SHI represses gibberellin responses in Arabidopsis and barley. Plant physiology.127:937-948.

Gomi K, Sasaki A, Itoh H, Ueguchi-Tanaka M, Ashikari M, Kitano H, Matsuoka M. 2004. GID2, an F-box subunit of the SCF E3 complex, specifically interacts with phosphorylated SLR1 protein and regulates the gibberellin-dependent degradation of SLR1 in rice. Plant Journal.37:626-634.

Grada A, Weinbrecht K. 2013. Next-generation sequencing: methodology and application. The Journal of investigative dermatology.133:e11.

Guo J, Wang Y, Song C, Zhou J, Qiu L, Huang H, Wang Y. 2010. A single origin and moderate bottleneck during domestication of soybean (Glycine max): implications from microsatellites and nucleotide sequences. Annals of Botany.106:505-514.

Han Y, Zhao X, Liu D, Li Y, Lightfoot DA, Yang Z, Zhao L, Zhou G, Wang Z, Huang L, et al. 2016. Domestication footprints anchor genomic regions of agronomic importance in soybeans. New Phytologist.209:871-884.

Hedden P, Thomas SG. 2012. Gibberellin biosynthesis and its regulation. Biochemical Journal.444:11-25.

Helliwell C, Sheldon C, Olive M, Walker A, Zeevaart J, Peacock W, Dennis E. 1998. Cloning of the Arabidopsis ent-kaurene oxidase gene GA3. Proceedings of the National Academy of Sciences of the United States of America.95:9019-9024.

Hirano K, Ueguchi-Tanaka M, Matsuoka M. 2008. GID1-mediated gibberellin signaling in plants. Trends Plant Sci. Apr;13:192-199. Epub 2008/03/14.

Hou A, Chen P, Alloatti J, Li D, Mozzoni L, Zhang B, Shi A. 2009. Genetic Variability of Seed Sugar Content in Worldwide Soybean Germplasm Collections. Crop Science.49:903.

Hymowitz T. 2004. Speciation and cytogenetics. In: Soybeans: Improvement, Production, and Uses. p. 97-136.

Hymowitz T, Kaizuma N. 1981. Soybean seed protein electrophoresis profiles from 15 Asian countries or regions: Hypotheses on paths of dissemination of soybeans from China. Economic Botany.35:10-23.

Ikuma H, Thimann KV. 1960. Action of Gibberellic Acid on Lettuce Seed Germination. Plant physiology.35:557-566.

Jacobsen SE, Olszewski NE. 1993. Mutations at the SPINDLY locus of Arabidopsis alter gibberellin signal transduction. The Plant cell.5:887-896.

Khan MA, Ungar IA. 1997. Alleviation of Seed Dormancy in the Desert ForbZygophyllum simplexL. from Pakistan. Annals of Botany.80:395-400.

Kim MY, Lee S, Van K, Kim TH, Jeong SC, Choi IY, Kim DS, Lee YS, Park D, Ma J, et al. 2010. Whole-genome sequencing and intensive analysis of the undomesticated soybean (Glycine soja Sieb. and Zucc.) genome. Proc Natl Acad Sci U S A. Dec 21;107:22032-22037. Epub 2010/12/07.

Koornneef M, van der Veen JH. 1980. Induction and analysis of gibberellin sensitive mutants in Arabidopsis thaliana (L.) heynh. Theoretical and Applied Genetics.58:257-263.

Kucera B, Cohn MA, Leubner-Metzger G. 2005. Plant hormone interactions during seed dormancy release and germination. Seed Science Research.15:281-307.

Lam H-M, Xu X, Liu X, Chen W, Yang G, Wong F-L, Li M-W, He W, Qin N, Wang B, et al. 2010. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nature genetics.42:1053-1059.

Lee G-A, Crawford GW, Liu L, Sasaki Y, Chen X. 2011. Archaeological Soybean (Glycine max) in East Asia: Does Size Matter? PLoS ONE.6:e26720.

Leubner-Metzger G. 2001. Brassinosteroids and gibberellins promote tobacco seed germination by distinct pathways. Planta.213:758-763.

Li J, Jin H. 2007. Regulation of brassinosteroid signaling. Trends in Plant Science.12:37-41.

Li J, Wen J, Lease KA, Doke JT, Tax FE, Walker JC. 2002. BAK1, an Arabidopsis LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signaling. Cell. Jul 26;110:213-222. Epub 2002/08/02.

Li Y-H, Li W, Zhang C, Yang L, Chang R-Z, Gaut BS, Qiu L-J. 2010. Genetic diversity in domesticated soybean (Glycine max) and its wild progenitor (Glycine soja) for simple sequence repeat and single-nucleotide polymorphism loci. New Phytologist.188:242-253.

Li YH, Zhou G, Ma J, Jiang W, Jin LG, Zhang Z, Guo Y, Zhang J, Sui Y, Zheng L, et al. 2014. De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. Nat Biotechnol. Oct;32:1045-1052. Epub 2014/09/15.

Libault M, Farmer A, Brechenmacher L, Drnevich J, Langley RJ, Bilgin DD, Radwan O, Neece DJ, Clough SJ, May GD, et al. 2010. Complete Transcriptome of the Soybean Root Hair Cell, a Single-Cell Model, and Its Alteration in Response to Bradyrhizobium japonicum Infection. PLANT PHYSIOLOGY.152:541-552.

Libault M, Farmer A, Joshi T, Takahashi K, Langley RJ, Franklin LD, He J, Xu D, May G, Stacey G. 2010. An integrated transcriptome atlas of the crop model Glycine max, and its use in comparative analyses in plants. The Plant Journal.63:no-no.

Logan DC, Millar AH, Sweetlove LJ, Hill SA, Leaver CJ. 2001. Mitochondrial biogenesis during germination in maize embryos. Plant physiology.125:662-672.

MacMillan J. 2001. Occurrence of Gibberellins in Vascular Plants, Fungi, and Bacteria. Journal of Plant Growth Regulation.20:387-442.

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. 2008. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. Genome Research.18:1509-1517.

McGinnis KM, Thomas SG, Soule JD, Strader LC, Zale JM, Sun T-p, Steber CM. 2003. The Arabidopsis SLEEPY1 gene encodes a putative F-box subunit of an SCF E3 ubiquitin ligase. The Plant cell.15:1120-1130.

Morozova O, Marra MA. 2008. Applications of next-generation sequencing technologies in functional genomics. Genomics.92:255-264.

Müller K, Tintelnot S, Leubner-Metzger G. 2006. Endosperm-limited Brassicaceae Seed Germination: Abscisic Acid Inhibits Embryo-induced Endosperm Weakening of Lepidium sativum (cress) and Endosperm Rupture of Cress and Arabidopsis thaliana. Plant and Cell Physiology.47:864-877.

Nam KH, Li J. 2002. BRI1/BAK1, a receptor kinase pair mediating brassinosteroid signaling. Cell. Jul 26;110:203-212. Epub 2002/08/02.

Nikolić R, Mitić N, Miletić R, Nešković M. 2006. Effects of Cytokinins on In Vitro Seed Germination and Early Seedling Morphogenesis in Lotus corniculatus L. Journal of Plant Growth Regulation.25:187-194.

Ogas J, Cheng JC, Sung ZR, Somerville C. 1997. Cellular differentiation regulated by gibberellin in the Arabidopsis thaliana pickle mutant. Science (New York, NY).277:91-94.

Ogas J, Kaufmann S, Henderson J, Somerville C. 1999. PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in Arabidopsis. Proceedings of the National Academy of Sciences.96:13839-13844.

Ogawa M, Hanada A, Yamauchi Y, Kuwahara A, Kamiya Y, Yamaguchi S. 2003. Gibberellin biosynthesis and response during Arabidopsis seed germination. The Plant cell.15:1591-1604.

Oh E, Yamaguchi S, Kamiya Y, Bae G, Chung W-I, Choi G. 2006. Light activates the degradation of PIL5 protein to promote seed germination through gibberellin in Arabidopsis. The Plant Journal.47:124-139.

Olszewski N, Sun T-P, Gubler F. 2002. Gibberellin signaling: biosynthesis, catabolism, and response pathways. The Plant cell.14 Suppl:S61-S80.

Peano C, Pietrelli A, Consolandi C, Rossi E, Petiti L, Tagliabue L, De Bellis G, Landini P. 2013. An efficient rRNA removal method for RNA sequencing in GC-rich bacteria. Microbial Informatics and Experimentation.3:1.

Peleg Z, Blumwald E. 2011. Hormone balance and abiotic stress tolerance in crop plants. Current Opinion in Plant Biology.14:290-295.

Penfield S, Josse E-M, Kannangara R, Gilday AD, Halliday KJ, Graham IA. 2005. Cold and Light Control Seed Germination through the bHLH Transcription Factor SPATULA. Current Biology.15:1998-2006.

Peng J, Carol P, Richards DE, King KE, Cowling RJ, Murphy GP, Harberd NP. 1997. The Arabidopsis GAI gene defines a signaling pathway that negatively regulates gibberellin responses. Genes and Development.11:3194-3205.

Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ, Worland AJ, Pelica F, et al. 1999. 'Green revolution' genes encode mutant gibberellin response modulators. Nature.400:256-261.
Pennazio S, Roggero P. 1991. Effects of exogenous salicylate on basal and stress-induced ethylene formation in soybean. Biologia Plantarum.33:58-65.
Petruzzelli L, Coraggio I, Leubner-Metzger G. 2000. Ethylene promotes ethylene biosynthesis during pea seed germination by positive feedback regulation of 1-aminocyclo-propane-1-carboxylic acid oxidase. Planta.211:144-149.
Petruzzelli L, Sturaro M, Mainieri D, Leubner-Metzger G. 2003. Calcium requirement for ethylene-dependent responses involving 1-aminocyclopropane-1-carboxylic acid oxidase in radicle tissues of germinated pea seeds*. Plant, Cell and Environment.26:661-671.
Phinney BO. 1956. Growth Response of Single-Gene Dwarf Mutants in Maize To Gibberellic Acid. Proceedings of the National Academy of Sciences of the United States of America.42:185-189.
Prince SJ, Joshi T, Mutava RN, Syed N, Joao Vitor MdS, Patil G, Song L, Wang J, Lin L, Chen W, et al. 2015. Comparative analysis of the drought-responsive transcriptome in soybean lines contrasting for canopy wilting. Plant Science.240:65-78.
Radley M. 1956. Occurrence of Substances Similar to Gibberellic Acid in Higher Plants. Nature.178:1070-1071.
Ramaih S, Guedira M, Paulsen GM. 2003. Relationship of indoleacetic acid and tryptophan to dormancy and preharvest sprouting of wheat. Functional Plant Biology.30:939.
Razem FA, Baron K, Hill RD. 2006. Turning on gibberellin and abscisic acid signaling. Current opinion in plant biology.9:454-459.
Richards DE, King KE, Ait-ali T, Harberd NP. 2001. HOW GIBBERELLIN REGULATES PLANT GROWTH AND DEVELOPMENT : A Molecular Genetic Analysis of Gibberellin Signaling. Annual Review of Plant Physiology and Plant Molecular Biology.52:67-88.
Ross JJ, O'Neill DP, Wolbang CM, Symons GM, Reid JB. 2001. Auxin-Gibberellin Interactions and Their Role in Plant Growth. Journal of Plant Growth Regulation.20:346-353.
Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, et al. 2010. Genome sequence of the palaeopolyploid soybean. Nature.463:178-183.
Schneider G, Schliemann W. 1994. Gibberellin conjugates: an overview. Plant Growth Regulation.15:247-260.
Sedivy EJ, Wu F, Hanzawa Y. 2017. Soybean domestication : the origin , genetic architecture and molecular bases. New Phytologist.539-553.
Severin AJ, Cannon SB, Graham MM, Grant D, Shoemaker RC. 2011. Changes in twelve homoeologous genomic regions in soybean following three rounds of polyploidy. Plant Cell. Sep;23:3129-3136. Epub 2011/09/16.
Severin AJ, Woody JL, Bolon Y-T, Joseph B, Diers BW, Farmer AD, Muehlbauer GJ, Nelson RT, Grant D, Specht JE, et al. 2010. RNA-Seq Atlas of Glycine max: A guide to the soybean transcriptome. BMC Plant Biology.10:160.
Shuai H, Meng Y, Luo X, Chen F, Zhou W, Dai Y, Qi Y, Du J, Yang F, Liu J, et al. 2017. Exogenous auxin represses soybean seed germination through decreasing the gibberellin/abscisic acid (GA/ABA) ratio. Scientific Reports.7:12620.
Silverstone AL, Ciampaglio CN, Sun T. 1998. The Arabidopsis RGA gene encodes a transcriptional regulator repressing the gibberellin signal transduction pathway. The Plant cell.10:155-169.
Song L, Prince S, Valliyodan B, Joshi T, Maldonado dos Santos JV, Wang J, Lin L, Wan J, Wang Y, Xu D, et al. 2016. Genome-wide transcriptome analysis of soybean primary root under varying water-deficit conditions. BMC Genomics.17:57.
Song Q, Jenkins J, Jia G, Hyten DL, Pantalone V, Jackson SA, Schmutz J, Cregan PB. 2016. Construction of high resolution genetic linkage maps to improve the soybean genome sequence assembly Glyma1.01. BMC Genomics.17:33.
Sponsel VM, Hedden P. 2010. Gibberellin Biosynthesis and Inactivation. In: Plant Hormones Dordrecht: Springer Netherlands. p. 63-94.

Tanimoto E, Hirano K. 2013. Plant Roots: The Hidden Half, Fourth EditionFourth ed. p. 13-11-13-11.

Thomas SG, Phillips aL, Hedden P. 1999. Molecular cloning and functional expression of gibberellin 2- oxidases, multifunctional enzymes involved in gibberellin deactivation. Proceedings of the National Academy of Sciences of the United States of America.96:4698-4703.

Ueguchi-Tanaka M, Ashikari M, Nakajima M, Itoh H, Katoh E, Kobayashi M, Chow T-y, Hsing Y-iC, Kitano H, Yamaguchi I, et al. 2005. GIBBERELLIN INSENSITIVE DWARF1 encodes a soluble receptor for gibberellin. Nature.437:693-698.

Ueguchi-Tanaka M, Fujisawa Y, Kobayashi M, Ashikari M, Iwasaki Y, Kitano H, Matsuoka M. 2000. Rice dwarf mutant d1, which is defective in the alpha subunit of the heterotrimeric G protein, affects gibberellin signal transduction. Proceedings of the National Academy of Sciences.97:11638-11643.

Varbanova M, Yamaguchi S, Yang Y, McKelvey K, Hanada A, Borochov R, Yu F, Jikumaru Y, Ross J, Cortes D, et al. 2007. Methylation of Gibberellins by Arabidopsis GAMT1 and GAMT2. THE PLANT CELL ONLINE.19:32-45.

Wang J, Chu S, Zhang H, Zhu Y, Cheng H, Yu D. 2016. Development and application of a novel genome-wide SNP array reveals domestication history in soybean. Scientific Reports.6:20728.

Wang L, Cao C, Ma Q, Zeng Q, Wang H, Cheng Z, Zhu G, Qi J, Ma H, Nian H, et al. 2014. RNA-seq analyses of multiple meristems of soybean: novel and alternative transcripts, evolutionary and functional implications. BMC Plant Biology.14:169.

Wang X, Chory J. 2006. Brassinosteroids regulate dissociation of BKI1, a negative regulator of BRI1 signaling, from the plasma membrane. Science. Aug 25;313:1118-1122. Epub 2006/07/22.

Wang ZY, Seto H, Fujioka S, Yoshida S, Chory J. 2001. BRI1 is a critical component of a plasma-membrane receptor for plant steroids. Nature. Mar 15;410:380-383. Epub 2001/03/27.

Wen C-K, Chang C. 2002. Arabidopsis RGL1 encodes a negative regulator of gibberellin responses. The Plant cell.14:87-100.

Willert K, Nusse R. 1998. Beta-catenin: a key mediator of Wnt signaling. Current opinion in genetics & development.8:95-102.

Wilson RF. 2008. Soybean: Market Driven Research Needs. In: Genetics and Genomics of Soybean New York, NY: Springer New York. p. 3-15.

Woodger FJ, Gubler F, Pogson BJ, Jacobsen JV. 2003. A Mak-like kinase is a repressor of GAMYB in barley aleurone. The Plant Journal.33:707-717.

Xu D, Abe J, Gai J, Shimamoto Y. 2002. Diversity of chloroplast DNA SSRs in wild and cultivated soybeans: evidence for multiple origins of cultivated soybean. TAG Theoretical and Applied Genetics.105:645-653.

Yabuta T. 1935. Biochemistry of the 'bakanae' fungus of rice. Agric Hort (Tokyo).10:17-22.

Yamaguchi S. 2008. Gibberellin Metabolism and its Regulation. Annual Review of Plant Biology.59:225-251.

Yano R, Kanno Y, Jikumaru Y, Nakabayashi K, Kamiya Y, Nambara E. 2009. CHOTTO1, a putative double APETALA2 repeat transcription factor, is involved in abscisic acid-mediated repression of gibberellin biosynthesis during seed germination in Arabidopsis. Plant physiology.151:641-654.

Yomo H, Iinuma H. 1966. Production of gibberellin-like substance in the embryo of barley during germination. Planta.71:113-118.

Zapata PJ, Serrano Ma, Pretel MT, Amorós A, Botella MÁ, María, Pretel MT, Amorós A, Botella MÁ. 2004. Polyamines and ethylene changes during germination of different plant species under salinity. Plant Science.167:781-788.

Zhang S, Cai Z, Wang X. 2009. The primary signaling outputs of brassinosteroids are regulated by abscisic acid signaling. Proceedings of the National Academy of Sciences.106:4543-4548.

Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, Yu Y, Shu L, Zhao Y, Ma Y, et al. 2015. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nature Biotechnology.33:408-414.

Zhu Y. 2006. ELONGATED UPPERMOST INTERNODE Encodes a Cytochrome P450 Monooxygenase That Epoxidizes Gibberellins in a Novel Deactivation Reaction in Rice. THE PLANT CELL ONLINE.18:442-456.

Chapter 2 Expansion and diversification of the gibberellin receptor GIBBERELLIN INSENSITIVE DWARF1 (GID1) family in land plants

2.1 INTRODUCTION

Gibberellins (GAs) are hormones that regulate various processes in plant development, particularly during seed germination, flowering, pollen development and stem elongation (Olszewski et al. 2002). The classic GA signaling pathway is characterized by the recognition of bioactive GA (e.g. GA_3 and GA_4) by the GIBBERELLIN INSENSITIVE DWARF1 (GID1) receptor. GID1 is a nucleocytoplasmic protein (Livne and Weiss 2014) that was initially identified in rice (OsGID1, *Oryza sativa*) (Ueguchi-Tanaka et al. 2005). Upon interaction with GA, GID1 undergoes a conformational change that increases its affinity for DELLA, proteins that typically inhibit GA signaling by: interacting and blocking the activity of transcription factors that drive GA transcriptional programs (Murase et al. 2008); co-activating negative regulators of GA signaling or; recruiting chromatin remodeling proteins to specific promoter regions (Nelson and Steber 2016). In the canonical GA signaling pathway, the GA-GID1-DELLA complex is recognized by the SCF^{SLY1} ubiquitin ligase, which ubiquitinates DELLA proteins, promoting their proteasomal degradation (Dill et al. 2004; Fu et al. 2004; Gomi et al. 2004; McGinnis et al. 2003; Peng et al. 1997). Therefore, the down-regulation of DELLA ultimately triggers the classic GA effects (Fleet and Sun 2005). Alternative GA signaling pathways have also been proposed, such as a GA-independent (GID1-mediated) (Yamamoto et al. 2010) and DELLA-independent pathways (Fuentes et al. 2012). Interestingly, canonical and alternative pathways rely on GID1, which appears to have a central role in GA signaling.

GID1 receptors evolved from a larger family of Hormone Sensitive Lipases (HSLs). Comparison of HSLs with the rice GID1 revealed important differences: the His from the HSL catalytic triad (Ser-Asp-His) is replaced by Val in GID1; the last Gly of the HGGG motif is substituted by Ser in GID1 and; the extensive

divergence between the N-terminal lid of GID1 and HSLs (Hirano et al. 2012). Detailed structural analyses of the GA-GID1a-DELLA complex support that these changes are critical for GA binding. Other GID1a amino acid residues were also found to be involved in GA interaction: Gly^{114}, Gly^{115}, Ser^{116}, Ile^{126}, Tyr^{127}, Ser^{191}, Phe^{238}, Val^{239}, Asp^{243}, Arg^{244}, Tyr^{247}, Gly^{320}, Tyr^{322}, Leu^{323} (core domain residues) and; Ile^{24}, Phe^{27}, Lys^{28}, Tyr^{31}, Arg^{35} (N-terminal extension residues) (Murase et al. 2008).

Three GID1 receptor genes have been characterized in *Arabidopsis thaliana* (GID1a, GID1b and GID1c). Although some level of functional redundancy was found between these genes, each of them apparently play specific roles in different developmental stages (Griffiths et al. 2006; Iuchi et al. 2007; Suzuki et al. 2009; Willige et al. 2007). GID1 receptors were also characterized in several other plants, such as ferns (Hirano et al. 2007), cotton (Aleman et al. 2008), barley (Chandler et al. 2008) and wheat (Li et al. 2013). A previous phylogenetic reconstruction of GID1 receptors uncovered the presence of three major groups: eudicot GID1ac, eudicot GID1b and monocot GID1, supporting that a diversification of this family occurred after the divergence of monocots and eudicots (Voegele et al. 2011). In addition to the phylogenetic separation of GID1ac and GID1b subfamilies, a number of important features related to the functional specialization of GID1 subfamilies have been described: 1) a remarkable difference in their transcriptional profiles across several tissues, such as in roots (Griffiths et al. 2006) and during germination (Bellieny-Rabelo et al. 2016); 2) GA-mediated transcriptional down-regulation of GID1ac, but not GID1b (Voegele et al. 2011); 3) The different affinity of GID1 subfamilies for GA, with GID1b displaying greater affinity for GA_3 and GA_4 than GID1a and GID1c (Nakajima et al. 2006) and; 4) The preference of specific GID1 proteins for particular DELLA groups (Hirano et al. 2007), potentially increasing the complexity involved in GA signaling.

Although important aspects of the GID1 family have been elucidated since its discovery and structural determination, important questions remain to be answered regarding the expansion and diversification of the family, the distribution of GID1ac and GID1b subfamilies in major eudicot lineages and the major evolutionary forces shaping the eudicot GID1 subfamilies at the sequence and transcriptional levels. Here we performed a comprehensive survey of GID1 proteins in 54 plant genomes and integrate this data with protein structure and gene expression data. Our results provide important insights on the evolutionary history of the GID1 family in land plants, including findings such as: 1) a detailed phylogenetic reconstruction of GID1s and the identification of the main expansion and diversification events, including a contribution of whole-genome duplication (WGD) events to the structure of the GID1 family in eudicots; 2) the conservation and divergence of key amino acid residues involved in GA and DELLA binding by GID1b and GID1ac and; 3) the important contribution of gene expression divergence in the establishment of the GID1ac and GID1b subfamilies in eudicots. Finally, we discuss theoretical aspects regarding the evolution of GA perception mechanisms, which can fuel future computational and experimental studies.

2.2 MATERIALS AND METHODS

2.2.1 Identification of GID1 proteins in land plants

To identify the GID1 proteins in land plants, predicted proteins of 47 angiosperms, two gymnosperms, one lycophyte and three bryophytes were downloaded from various sources (Table 2.1). GID1 homologs were identified in four steps: 1) BLASTP (Altschul et al. 1997) searches using experimentally characterized GID1s from *Arabidopsis. thaliana*, *Lepidium sativum* and rice to search the predicted proteomes of each species (a total 2,041,985 proteins), with e-value and similarity thresholds of ≤ $1e^{-5}$ and ≥ 38%, respectively. This step resulted in a total of 259 proteins; 2) Only the 245 sequences with the conserved motifs HGG and GXSXG, also shared with HSLs and other plant

carboxylesterases (Ueguchi-Tanaka et al. 2005; Voegele et al. 2011), were retained; 3) Bona-fide GID1s were separated from plant carboxylesterases using a phylogenomic approach, as follows: carboxylesterases of *Ar. thaliana* (AT5G23530) and rice (ABA92266) (Hirano et al. 2007) were aligned with the 245 GID1 candidates using PROMALS3D (Pei et al. 2008). The phylogenetic reconstruction was performed with FastTree (Price et al. 2010). A total of 141 GID1s clearly clustered in a monophyletic clade (Figure 2.1) and were separated from carboxylesterases; 4) redundancy was removed with the aid of BLASTCLUST (95% coverage and 95% identity thresholds) (Altschul et al. 1997). These steps allowed us to identify 132 GID1s. Our collection was supplemented with *Triticum aestivum* and *Le. sativum* GID1s (three from each) (Li et al. 2013; Voegele et al. 2011). One GID1 from *Cajanus cajan* was excluded because of the absence of a start codon. The coding sequences of the identified GID1s were also searched in their respective genomes using BLASTN with an e-value threshold of ≤ $1e^{-6}$ (Altschul et al. 1997), which allowed us to identify an additional *Glycine soja* GID1.

By using the pipeline described above, we have not found GID1 genes in the downloaded proteome/genome of *Picea glauca* and found that one of two GID1 genes of *Selaginella moellendorffi* was fragmented. We believe that these problems were due to assembly incompleteness or gene prediction problems. We obtained GID1 sequences from these two species from individual Genbank entries [*Pi. glauca* (Genbank: BN001188.1) and *Se. moellendorffii* (Refseq: XP_002993392.1, XP_002993392.1)]. Overall, a total of 141 GID1s were used in the analyses (Table 2.2). Species names were abbreviated by the first letter of genus followed by the four first letters of the species name (e.g. Athal corresponds to *Ar. thaliana*) (Table 2.2). Eudicot GID1s were classified in GID1a, GID1b and GID1c using *Ar. thaliana* GID1s as reference. Non-eudicot GID1s were simply numbered, as there is no subfamily division in these species.

Table 2.1 List of plant species used in this study.

Species	Species code name	Taxonomic group	Number of genes	Source
Cucumis sativus	Csati	Eudicot	26548	Cucurbit Genomics Database
Jatropha curcas	Jcurc	Eudicot	57437	ftp://ftp.kazusa.or.jp/pub/jatropha/
Lotus japonicas	Ljapo	Eudicot	39734	ftp://ftp.kazusa.or.jp/pub/lotus/lotus_r3.0
Vigna angularis	Vangu	Eudicot	37769	NCBI
Vigna radiate	Vradi	Eudicot	35143	NCBI
Citrullus lanatus	Clana	Eudicot	23440	ftp://www.icugi.org/pub/genome/watermelon/97103/v1/
Cajanus cajan	Ccaja	Eudicot	48680	http://gigadb.org/dataset/100028
Pyrus x bretschneideri	Pbret	Eudicot	42369	http://gigadb.org/dataset/100083
Actinidia chinensis	Achin	Eudicot	39040	Kiwifruit Genome Database
Arachis duranensis	Adura	Eudicot	42562	NCBI
Cicer arietinum	Carie	Eudicot	33107	NCBI
Glycine soja	Gsoja	Eudicot	50399	NCBI
Gossypium hirsutum	Ghiru	Eudicot	90927	NCBI
Lepidium sativum	Lsati	Eudicot	NA	NCBI
Amborella trichopoda	Atric	Basal angiosperm	26846	Phytozome11.0
Aquilegia coerulea	Acoer	Basal eudicot	24823	Phytozome11.0
Arabidopsis lyrata	Alyra	Eudicot	32670	Phytozome11.0
Arabidopsis thaliana	Athal	Eudicot	27416	Phytozome11.0
Brassica rapa	Brapa	Eudicot	40492	Phytozome11.0
Boechera stricta	Bstri	Eudicot	27416	Phytozome11.0

Capsella grandiflora	Cgran	Eudicot	24805	Phytozome11.0
Capsella rubella	Crube	Eudicot	26521	Phytozome11.0
Carica papaya	Cpapa	Eudicot	27751	Phytozome11.0
Manihot esculenta	Mescu	Eudicot	33033	Phytozome11.0
Ricinus communis	Rcomm	Eudicot	31221	Phytozome11.0
Glycine max	Gmax	Eudicot	56044	Phytozome11.0
Medicago truncatula	Mtrun	Eudicot	50894	Phytozome11.0
Phaseolus vulgaris	Pvulg	Eudicot	27197	Phytozome11.0
Gossypium raimondii	Graim	Eudicot	37505	Phytozome11.0
Theobroma cacao	Tcaca	Eudicot	29452	Phytozome11.0
Fragaria vesca	Fvesc	Eudicot	32831	Phytozome11.0
Malus domestica	Mdome	Eudicot	63514	Phytozome11.0
Prunus persica	Ppers	Eudicot	26873	Phytozome11.0
Populus trichocarpa	Ptric	Eudicot	41335	Phytozome11.0
Solanum lycopersicum	Slyco	Eudicot	34727	Phytozome11.0
Solanum tuberosum	Stube	Eudicot	35119	Phytozome11.0
Vitis vinifera	Vvini	Eudicot	26346	Phytozome11.0
Musa acuminate	Macum	Monocot	36542	Phytozome11.0
Brachypodium distachyon	Bdist	Monocot	34310	Phytozome11.0
Brachypodium stacei	Bstac	Monocot	29898	Phytozome11.0
Oryza sativa	Osati	Monocot	42189	Phytozome11.0
Panicum hallii	Phall	Monocot	37232	Phytozome11.0
Panicum virgatum	Pvirg	Monocot	98007	Phytozome11.0
Phoenix dactylifera	Pdact	Monocot	38570	NCBI

Sorghum bicolor	Sbico	Monocot	34211	Phytozome11.0
Setaria italic	Sital	Monocot	34584	Phytozome11.0
Zea mays	Zmays	Monocot	63480	Phytozome11.0
Triticum aestivum	Taest	Monocot	99386	Phytozome11.0
Pinus taeda	Ptaed	Gymnosperm	33708	http://dendrome.ucdavis.edu/ftp/Proteome_Data/protein/Pita/protein_Pita.fasta
Picea glauca	Pglau	Gymnosperm	6445	NCBI
Selaginella moellendorffii	Smoel	Lycophyte	22285	Phytozome11.0
Sphagnum fallax	Sfall	Bryophyte	26939	Phytozome12.0
Marchantia polymoprha	Mpoly	Bryophyte	19287	Phytozome12.0
Physcomitrella patens	Ppate	Bryophyte	32926	Phytozome11.0

Table 2.2 GID1s identified in 54 plant species.

Gene_ID	Assigned_name	Length (aa)	Number of Introns	Best hit in *Arabidopsis*	Similarity with *Arabidopsis* best hit
Achn259581	Achin.GID1b	389	2	Athal.GID1b	79.3
Achn083581	Achin.GID1c	328	0	Athal.GID1c	80.73
Aquca_026_00095.1	Acoer.GID1	343	1	Athal.GID1c	78.13
XP_015951044.1	Adura.GID1b1	344	1	Athal.GID1b	77.19
XP_015969395.1	Adura.GID1b2	354	1	Athal.GID1b	70.09
XP_015968550.1	Adura.GID1c	345	1	Athal.GID1c	75.8
477795	Alyra.GID1a	344	1	Athal.GID1a	97.95
486805	Alyra.GID1b	358	1	Athal.GID1b	96.09
351756	Alyra.GID1c	344	1	Athal.GID1c	96.51

AT3G05120.1	Athal.GID1a	345	1	Athal.GID1a	100
AT3G63010.1	Athal.GID1b	358	1	Athal.GID1b	100
AT5G27320.1	Athal.GID1c	344	1	Athal.GID1c	100
evm_27.model.AmTr_v1.0_scaffold00197.22	Atric.GID1	364	2	Athal.GID1c	72.67
Bradi2g25600.1.p	Bdist.GID1	355	1	Athal.GID1c	62.22
Brara.E03404.1.p	Brapa.GID1a	346	1	Athal.GID1a	84.93
Brara.D00038.1.p	Brapa.GID1b1	360	1	Athal.GID1b	91.62
Brara.G01992.1.p	Brapa.GID1b2	358	1	Athal.GID1b	88.3
Brara.F02873.1.p	Brapa.GID1c	345	1	Athal.GID1c	92.17
Brast08G119400.1.p	Bstac.GID1	357	1	Athal.GID1c	62.15
Bostr.2570s0176.1.p	Bstri.GID1a	349	1	Athal.GID1a	97.1
Bostr.13158s0303.1.p	Bstri.GID1b	358	1	Athal.GID1b	95.81
Bostr.29827s0050.1.p	Bstri.GID1c	343	1	Athal.GID1c	96.8
XP_004496054.1	Carie.GID1b1	343	1	Athal.GID1b	76.38
XP_004493628.1	Carie.GID1b2	348	1	Athal.GID1b	72.62
XP_004495762.1	Carie.GID1c	345	1	Athal.GID1c	76.45
C.cajan_10477	Ccaja.GID1b	344	1	Athal.GID1b	76.38
C.cajan_06890	Ccaja.GID1c	344	1	Athal.GID1c	79.36
Cagra.2179s0025.1.p	Cgran.GID1a	345	1	Athal.GID1a	97.1
Cagra.0664s0039.1.p	Cgran.GID1b	358	1	Athal.GID1b	91.34
Cla014721	Clana.GID1b	342	1	Athal.GID1b	75.15
Cla011311	Clana.GID1c	332	0	Athal.GID1c	82.23
evm.model.supercontig_731.1	Cpapa.GID1b	344	1	Athal.GID1b	78.07
evm.model.supercontig_84.118	Cpapa.GID1c	233	2	Athal.GID1c	83.42

Carubv10014098m	Crube.GID1a	345	1	Athal.GID1a	97.1
Carubv10017514m	Crube.GID1b	358	1	Athal.GID1b	90.78
cucumber.evm.model.Chr7.1626	Csati.GID1b	342	1	Athal.GID1b	76.32
cucumber.evm.model.Chr1.2651	Csati.GID1c	345	0	Athal.GID1c	80.64
mrna27756.1-v1.0-hybrid	Fvesc.GID1b1	400	2	Athal.GID1b	77.71
mrna20092.1-v1.0-hybrid	Fvesc.GID1b2	361	0	Athal.GID1b	75.38
mrna22353.1-v1.0-hybrid	Fvesc.GID1c	361	1	Athal.GID1c	77.25
XP_016667676.1	Ghiru.GID1c1	344	1	Athal.GID1c	82.56
XP_016675315.1	Ghiru.GID1c2	373	1	Athal.GID1c	82.27
XP_016680245.1	Ghiru.GID1c3	344	1	Athal.GID1c	81.98
XP_016667852.1	Ghiru.GID1c4	344	1	Athal.GID1c	81.82
XP_016673692.1	Ghiru.GID1c5	344	1	Athal.GID1c	81.69
XP_016667850.1	Ghiru.GID1c6	360	1	Athal.GID1c	81.63
XP_016696047.1	Ghiru.GID1b1	344	1	Athal.GID1b	79.88
XP_016730328.1	Ghiru.GID1b2	330	1	Athal.GID1b	79.69
XP_016733251.1	Ghiru.GID1b3	345	1	Athal.GID1b	76.97
Glyma.02G151100.1.p	Gmax.GID1b1	342	1	Athal.GID1b	75.44
Glyma.10G022900.1.p	Gmax.GID1b2	343	1	Athal.GID1b	74.85
Glyma.03G148300.1.p	Gmax.GID1b3	346	1	Athal.GID1b	72.25
Glyma.10G158000.1.p	Gmax.GID1c1	344	1	Athal.GID1c	79.07
Glyma.20G230600.1.p	Gmax.GID1c2	344	1	Athal.GID1c	78.78
Gorai.008G007200.1	Graim.GID1b1	344	1	Athal.GID1b	79.82
Gorai.007G132800.1	Graim.GID1b2	344	1	Athal.GID1b	79.3
Gorai.004G214300.1	Graim.GID1b3	345	1	Athal.GID1b	77.84

Gorai.013G203500.1	Graim.GID1c1	344	1	Athal.GID1c	82.56
Gorai.004G244100.1	Graim.GID1c2	344	1	Athal.GID1c	82.56
Gorai.N020400.1	Graim.GID1c3	307	1	Athal.GID1c	81.58
KHN46608.1	Gsoja.GID1b1	342	1	Athal.GID1b	75.44
KHN03762.1	Gsoja.GID1b2	343	1	Athal.GID1b	74.85
KHN07465.1	Gsoja.GID1b3	347	1	Athal.GID1b	72.05
KHN45395.1	Gsoja.GID1b4	341	1	Athal.GID1b	71.39
KN660624.1:1345549-1343940	Gsoja.GID1c1	344	1	Athal.GID1c	79.07
KHN20191.1	Gsoja.GID1c2	344	1	Athal.GID1c	78.78
Jcr4S06840.20	Jcurc.GID1b	344	1	Athal.GID1b	80.99
Jcr4S00235.70	Jcurc.GID1c	347	1	Athal.GID1c	86.65
Lj1g3v1686830.1	Ljapo.GID1b	349	1	Athal.GID1b	76.68
Lj0g3v0154849.1	Ljapo.GID1c	344	1	Athal.GID1c	76.09
HQ003455	Lsati.GID1a	349	0	Athal.GID1a	96.52
HQ003456	Lsati.GID1b	358	0	Athal.GID1b	93
HQ003457	Lsati.GID1c	343	0	Athal.GID1c	94.77
GSMUA_Achr3P24930	Macum.GID1_1	306	2	Athal.GID1c	65.31
GSMUA_Achr5P27430	Macum.GID1_3	308	2	Athal.GID1c	64.16
GSMUA_Achr8P05910	Macum.GID1_4	305	2	Athal.GID1c	63.45
GSMUA_Achr10P12580	Macum.GID1_5	310	2	Athal.GID1c	62.79
GSMUA_Achr6P16090	Macum.GID1_6	307	2	Athal.GID1c	62.21
GSMUA_Achr5P27450	Macum.GID1_2	338	3	Athal.GID1c	64.74
MDP0000929994	Mdome.GID1b1	346	1	Athal.GID1b	74.71
MDP0000319522	Mdome.GID1b2	415	2	Athal.GID1b	74.08

MDP0000319301	Mdome.GID1c1	364	1	Athal.GID1c	78.43
MDP0000445131	Mdome.GID1c2	387	1	Athal.GID1c	77.91
Manes.05G070500.1.p	Mescu.GID1b1	344	1	Athal.GID1b	80.12
Manes.01G212300.1.p	Mescu.GID1b2	344	1	Athal.GID1b	79.24
Manes.09G161600.1.p	Mescu.GID1c1	344	1	Athal.GID1c	82.46
Manes.08G124700.1.p	Mescu.GID1c2	344	1	Athal.GID1c	82.27
Medtr1g089310.1	Mtrun.GID1b1	360	2	Athal.GID1b	78.07
Medtr7g093950.1	Mtrun.GID1b2	350	1	Athal.GID1b	72.91
Medtr1g082210.1	Mtrun.GID1c	345	1	Athal.GID1c	77.03
LOC_Os05g33730.1	Osati.GID1	354	1	Athal.GID1c	63.14
Pbr041942.1	Pbret.GID1b	434	2	Athal.GID1b	73.8
Pbr006571.1	Pbret.GID1c1	331	0	Athal.GID1c	79.75
Pbr017104.1	Pbret.GID1c2	364	1	Athal.GID1c	79.71
XP_008795099.1	Pdact.GID1_1	348	1	Athal.GID1c	72.41
XP_017701655.1	Pdact.GID1_2	345	1	Athal.GID1c	69.65
BN001188.1	Pglau.GID1	352	0	Athal.GID1c	65.51
Pahal.C03051.1	Phall.GID1	350	1	Athal.GID1a	62.64
Prupe.8G249800.1.p	Ppers.GID1b	344	1	Athal.GID1b	78.07
Prupe.6G332800.1.p	Ppers.GID1c	344	1	Athal.GID1c	79.94
protein_Pita_169159248	Ptaed.GID1	357	0	Athal.GID1c	66.37
Potri.002G213100.1	Ptric.GID1b1	344	1	Athal.GID1b	80.17
Potri.014G135900.1	Ptric.GID1b2	346	1	Athal.GID1b	78.26
Potri.013G028700.1	Ptric.GID1c1	344	1	Athal.GID1c	84.59
Potri.005G040600.1	Ptric.GID1c2	344	1	Athal.GID1c	84.3
Pavir.J26042.1.p	Pvirg.GID1	321	0	Athal.GID1a	60.38

Phvul.007G167600.1	Pvulg.GID1b1	344	1	Athal.GID1b	75.51
Phvul.001G146300.1	Pvulg.GID1b2	344	1	Athal.GID1b	73.76
Phvul.007G183400.1	Pvulg.GID1c	344	1	Athal.GID1c	79.07
29703.m001506	Rcomm.GID1b	345	1	Athal.GID1b	78.36
30128.m008695	Rcomm.GID1c	344	1	Athal.GID1c	82.56
Sobic.009G134600.1.p	Sbico.GID1	355	1	Athal.GID1c	62.32
Seita.3G246300.1.p	Sital.GID1	350	1	Athal.GID1a	62.93
Solyc06g008870.2.1	Slyco.GID1b1	345	1	Athal.GID1b	79.24
Solyc09g074270.2.1	Slyco.GID1b2	345	1	Athal.GID1b	76.32
Solyc01g098390.2.1	Slyco.GID1c	349	1	Athal.GID1c	79.7
PGSC0003DMP400038057	Stube.GID1b1	345	1	Athal.GID1b	78.95
PGSC0003DMP400006843	Stube.GID1b2	345	1	Athal.GID1b	77.78
PGSC0003DMP400049733	Stube.GID1c	348	1	Athal.GID1c	80.18
FR668557_TaGID1-A1	Taest.GID1_1	355	0	Athal.GID1c	62.5
FR668558_TaGID1-B1	Taest.GID1_2	355	0	Athal.GID1c	61.93
FR668556_TaGID1-D1	Taest.GID1_3	354	0	Athal.GID1c	61.54
Thecc1EG004494t1	Tcaca.GID1b	344	1	Athal.GID1b	79.3
Thecc1EG026504t1	Tcaca.GID1c	344	1	Athal.GID1c	81.98
XP_017412437.1	Vangu.GID1b1	341	1	Athal.GID1b	74.78
XP_017418977.1	Vangu.GID1b2	413	1	Athal.GID1b	73.14
XP_017437543.1	Vangu.GID1c	344	1	Athal.GID1c	78.49
XP_014514500.1	Vradi.GID1b1	344	1	Athal.GID1b	75.8
XP_014496142.1	Vradi.GID1b2	344	1	Athal.GID1b	73.47
XP_014514775.1	Vradi.GID1c	344	1	Athal.GID1c	78.49
GSVIVT01011037001	Vvini.GID1b	280	3	Athal.GID1b	63.85

GSVIVT01022014001	Vvini.GID1c	388	3	Athal.GID1c	82.35
GRMZM2G173630_P01	Zmays.GID1_2	351	1	Athal.GID1c	62.46
GRMZM2G016605_P02	Zmays.GID1_1	350	1	Athal.GID1c	62.93
Pp3c3_16760V3.1	Ppate.GID1_1	336	2	Athal.GID1c	42
Pp3c13_21400V3.1	Ppate.GID1_2	343	2	Athal.GID1c	39
XP_002993392.1	Smoel.GID1_1	378	1	Athal.GID1c	56.38
XP_002975655.1	Smoel.GID1_2	371	2	Athal.GID1c	55.79
Sphfalx0013s0062.1.p	Sfall.GID1	350	1	Athal.GID1c	40.41
Mapoly0133s0031.1.p	Mpoly.GID1_1	403	1	Athal.GID1c	40.61
Mapoly0003s0300.1.p	Mpoly_GID1_2	335	0	Athal.GID1c	38.94

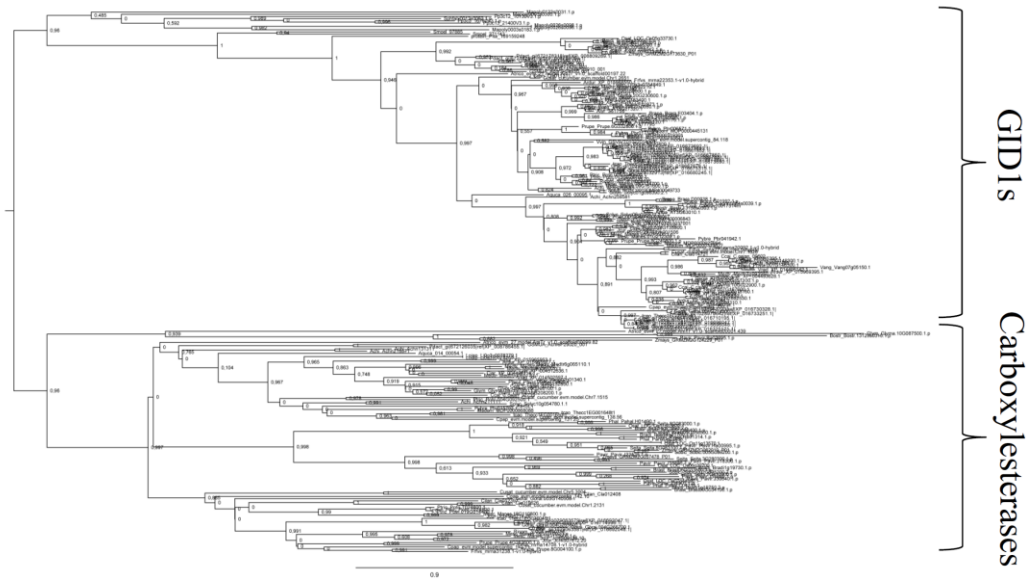


Figure 2.1 Phylogenetic reconstruction of GID1s and carboxylesterases.
Multiple sequence alignment was carried out using PROMALS3D and a phylogenetic tree was constructed using FastTree and visualized with FigTree. A total of 141 GID1s clustered apart from carboxylesterases and were retrieved for downstream analyses.

2.2.2 Sequence analysis, phylogenetic reconstruction and microsynteny analyses

Multiple sequence alignment of GID1 proteins was carried out using PROMALS3D (Pei et al. 2008) and visually inspected with Jalview (Waterhouse et al. 2009). Large N- and C-terminal gaps were removed. Conserved motifs were analyzed with MEME (v4.11.2) with the following parameters: distribution of motif was set to one per sequence, minimum and maximum motif lengths were set to 5 and 12 and the maximum number of motifs was set to 30. (Bailey et al. 2009). Phylogenetic reconstructions were performed with MrBayes (v3.2.2), using a mixed amino acid substitution model (Ronquist and Huelsenbeck 2003). The protein sequence alignment was run over 3,000,000 generations with a sampling frequency 100 and two independent runs, each containing four Markov Chain Monte Carlo (MCMC) chains and MCMC left at default settings to estimate posterior probabilities. Convergence was assessed using Tracer (v1.6) (http://beast.bio.ed.ac.uk/Tracer) over the trace files generated by the Bayesian MCMC runs. A consensus tree was generated and posterior probabilities were estimated by removing the 20% burn-in generations using the Sumt function of MrBayes. Alternatively, we have also reconstructed phylogenies with RAxML (version 8.2.9, best fit model, 100 bootstrap samples) (Stamatakis 2014). Best fit protein model was automatically calculated using the PROTGAMMAAUTO option. Gene structure analysis was performed using GSDS (v2) (Hu et al. 2015). The aligned proteins were used to guide the conversion of cDNA into codon alignments by PAL2NAL (Suyama et al. 2006). Ks calculations were performed with CODEML (PAML version 4.9b) (Yang 1997) using the Goldman and Yang ML method and the F3x4 model. For microsynteny, DAGchainer (v.r02062008) (Haas et al. 2004) was used to identify the homeologous regions in the genome using results from an all-vs-all bidirectional BlastP searches (e-value ≥ 1e-10, identity ≥35% and bit score ≥ 50). We required at least four aligned gene pairs to call a chromosome block as syntenic. Colinear genes surrounding GID1 genes were extracted and visualized in R using genoPlotR (Guy et al. 2010).

2.2.3 Functional divergence, *in silico* mutagenesis and docking

We used three different programs to infer functionally divergent sites in the GID1ac and GID1b subfamilies (with default parameters): FunDi (Gaston et al. 2011), GroupSim (Capra and Singh 2008) and Sequence Harmony (Feenstra et al. 2007). We used a threshold of 0.5 (Gaston et al. 2011) to filter the sites identified by FunDi and GroupSim, and the default threshold in Sequence harmony. Only those sites identified by all three programs were considered. An *in silico* mutagenesis approach was performed with FoldX (Schymkowitz et al. 2005), using the wild type crystal structures 2ZSH (GA_3–GID1a–DELLA) and 2ZSI (GA_4–GID1a–DELLA), excluding GA_3/GA_4 and DELLA. Crystal structures 2ZSH and 2ZSI were downloaded from the PDB database (Berman et al. 2000). Protein-ligand docking was performed using SwissDock (Grosdidier et al. 2011) and ligands (GA_3 and GA_4) were selected from the ZINC database (Irwin and Shoichet 2006). All structures were visualized with PyMOL (http://www.pymol.org/).

2.2.4 Gene expression data

Gene expression data of GID1 genes were obtained from publicly available sources, as following. Soybean: Soybase (http://soybase.org/soyseq/) (Severin et al. 2010) and from a recent manuscript from our group (Bellieny-Rabelo et al. 2016). *Ph. vulgaris*: Common bean gene expression atlas (http://plantgrn.noble.org/PvGEA) (O'Rourke et al. 2014). *Me. truncatula*: *Medicago truncatula* Gene Expression Atlas (MtGEA) (Benedito et al. 2008). *Ar. thaliana*: AtGenExpress (Schmid et al. 2005). Rice: Rice Express ion Database; (http://expression.ic4r.org). Maize and *Se. moellendorffii* gene expression data were obtained from two recent publications (Stelpflug et al. 2016; Zhu et al. 2017).

2.3 RESULTS AND DISCUSSION

2.3.1 Expansion and diversification of GID1 receptors in major groups of land plants

A total of 54 diverse plant genomes, including angiosperms, gymnosperms and basal land plants (i.e. a lycophyte and three bryophytes) (Table 2.1), were screened for GID1 proteins (see methods for details). Due to their high sequence similarity to HSLs, GID1s were separated with the aid of a phylogenetic reconstruction strategy (see methods for details) (Figure 2.1). We identified a total of 141 GID1 genes, with a median of two GID1s per genome (Table 2.2) and ~81% of the angiosperms containing 2-3 GID1 genes (Figure 2.2). All eudicots except the early-branching *Aquilegia coerulea* have more than one GID1; these proteins were classified using BLAST searches against *Ar. thaliana* GID1s and phylogenetic reconstructions by Bayesian and Maximum Likelihood approaches (see methods for details) (Figure 2.3, APPENDIX A.1 and Table 2.2). The species with the greatest number of GID1s are *Gossypium hirsutum* (nine), *Go. raimondii* (six) and *Glycine soja* (six) in eudicots, and *Musa acuminata* (six) in monocots (Figure 2.2). The 141 GID1s can be divided in four statistically supported groups (Voegele et al. 2011): group I (GID1ac) and II (GID1b), both with eudicot sequences; group III, with monocot GID1s and; group IV, containing GID1s from gymnosperms and basal plants. While GID1s from basal land plants and gymnosperms formed a separate small group, angiosperm GID1s diversified in the three former groups (Figure 2.3). Our results support the monophyly of all angiosperm GID1s, with the probable root of the tree lying within the paraphyletic group IV. The only GID1 from *Amborella trichopoda* (a basal angiosperm) is a sister group of the monocot and eudicot clades (Figure 2.3), supporting the expansion and divergence of GID1s after the emergence of angiosperms. More precisely, this diversification process happened after the split of Ranuncales, since *Aq. coerulea* has only one GID1 that is an early branch in the GID1ac clade (Figure 2.3). Our results also indicate that GID1b originated in eudicots after the separation of monocots, probably via the *gamma* polyploidy, a whole genome

triplication event shared by all core eudicots (Jiao et al. 2012). Because *Vi. vinifera* has not undergone other WGDs after the gamma polyploidy event (Jaillon et al. 2007), we used its genome as a reference to further investigate this hypothesis and found two genomic locations in *Vi. vinifera* (harboring *Vvini.GID1b* and *Vvini.GID1c*) that are collinear (i.e. syntenic) to a single region in *Ac. coerulea* (Figure 2.4A). Conversely, synteny is more eroded between *Vi. vinifera* and *Ar. thaliana*, probably due to two additional WGD and recombination events in the latter lineage (Figure 2.4A).

Next, we sought to explore the evolutionary history of the eudicot GID1ac and GID1b subfamilies. It is clear from our results that there is at least one GID1ac and one GID1b in every core eudicot, implying that these subfamilies acquired important non-redundant roles early in the evolution of eudicots (Voegele et al. 2011). Although GID1ac and GID1b subfamilies have comparable total sizes, their distribution is not uniform across lineages (Figure 2.2). Asterids and many rosids have a single GID1ac, although some independent lineage-specific expansions happened after the separation of these two large groups (Figures 2.2 and 2.3). In Malvaceae, the two *Gossypium* species experienced a more recent GID1ac expansion, after the split from *Theobroma cacao*. Microsynteny analysis indicates that *Graim.GID1c1* and *Graim.GID1c2* probably originated via a recent WGD or segmental duplication event (Figure 2.5A). In the order Brassicales, the GID1ac subfamily expanded after the separation of Brassicaceae and Caricaceae, with the emergence of a well-defined clade (harboring proteins related to *Ar. thaliana* GID1a) in the former, whereas *Carica papaya* preserved a single GID1c, outside of the GID1a clade (Figure 2.3). In fact, all GID1a proteins belong to a Brassicaceae-specific monophyletic clade nested inside GID1c; this GID1a clade could have emerged at the WGD events that took place after the split of Brassicaceae and Caricaceae (Schranz 2006), which is at least partially supported by microsynteny analysis between GID1a and GID1c genes in *Ar. thaliana* (Figure 2.4B). On the other hand, microsynteny between GID1a and GID1c is much lower in *Br. rapa*, possibly due to

rearrangements following the *Brassica* whole genome triplication (WGT) (Figure 2.5B). Interestingly, with the exception of *Capsella grandiflora* and *Capsella rubella* (Figure 2.3), Brassicaceae species retained both GID1a and GID1c genes, indicating that they also play non-redundant roles (Suzuki et al. 2009). Nevertheless, it has been shown that GID1a and GID1c can compensate the absence of each other during *Ar. thaliana* seed germination (Voegele et al. 2011), suggesting that the non-overlapping roles are performed in other conditions/tissues (Griffiths et al. 2006). *Capsella species* are the only core eudicots without a classical GID1c, suggesting a displacement of GID1c by GID1a in this genus. Therefore, these species would be good models to study the recent functional diversification within the GID1ac clade. Other GID1ac duplications that could be attributed to WGD events were also found in Salicaceae (*Populus trichocarpa*), *Glycine*, *Manihot esculenta* and in the most recent ancestor of *Malus domestica* and *Pyrus x bretschneideri* (Figure 2.3). Further, all the Fabaceae species except *Gl. max* and *Gl. soja* have a single GID1ac. Our phylogenetic analysis indicates that one of the GID1ac paralogs was rapidly lost after the legume WGD and the remaining GID1ac gene was later duplicated at the *Glycine* WGD (Figure 2.3). This scenario is also supported by synteny analysis (Figures 2.3, 2.4C and 2.4E) and by the presence of GID1 pairs with low Ks in *Gl. max* and *Gl. soja* (Table 2.3).

GID1b is mainly expanded in legumes, most likely due to the WGD events that happened at the base of Papilionoideae and *Glycine* (Figure 2.4C). Except for *Lotus japonicus* and *Cajanus cajan* (which independently lost one GID1b paralog), all other legumes retained duplicated GID1b sets, with two duplication rounds accounting for the 3-4 GID1b genes found in soybeans (Figure 2.2 and 2.3). Similarly to what was observed for GID1ac, there is also a soybean GID1b pair (*Gmax.GID1b1* and *Gmax.GID1b2*) that probably originated in the *Glycine* WGD event (Figure 2.3). This gene pair has a low Ks value (i.e. 0.187) that is compatible with the Glycine WGD age. Although these genes are not located in large homeologous genomic segments (Severin et al. 2011), they show a high

level of conservation in their genomic neighborhood (Figure 2.4D). Further, this scenario implies a loss of one GID1b in *Gl. max* after the separation from *Gl. soja*; this hypothesis is supported by phylogenetic reconstructions (Figure 2.3) and by the low Ks values of the respective surviving *Gl. soja* paralogous pair (*Gsoja.GID1b3* and *Gsoja.GID1b4*; Ks = 0.127; Figure 2.3, Table 2.3). Other expansions of GID1b genes can also be found in *Manihot esculenta*, *Fragaria vesca*, *Populus*, *Solanum* and *Gossypium*, for which polyploidization events have been documented or predicted (Mühlhausen and Kollmar 2013; Sato et al. 2012; Tuskan et al. 2006; Xu et al. 2011; Zhang et al. 2015). We have also found collinearity between *Go. raimondii* GID1b genes, indicating a role for WGD in the expansion of these genes (Figure 2.5C). Conversely to what was observed in GID1ac, the GID1b subfamily size is constrained in Brassicales, in which only *Br. rapa* has more than one member, which may have originated by a *Brassica* WGT event (Figures 2.2, 2.3 and 2.5D). Remarkably, even after several independent WGDs, almost all Brassicaceae species have reverted to a single GID1b, indicating that the retention of GID1b duplicates is peculiar to a few clades, particularly legumes.

We have not found diversified GID1 subgroups in monocots (Figures 2.2 and 2.3), in spite of multiple recent duplications in various lineages (e.g. maize and wheat). Nevertheless, a remarkable expansion resulted in six GID1 genes in banana (Figures 2.2 and 2.3). Interestingly, although three recent WGDs have been identified in the banana genome (D'Hont et al. 2012), the Ks values of these GID1 pairs are far greater than expected for duplicates generated in these WGDs (Table 2.3). The only banana GID1 pair with low Ks, *Macum.GID1_2* and *Macum.GID1_3,* is separated by less than 20 kb, with a single intervening gene (Figure 2.5E), supporting an origin via proximal (i.e. tandem) duplication. Furthermore, these banana GID1 pairs are outside of the homeologous blocks identified in the banana genome project (D'Hont et al. 2012). Nevertheless, we found conserved collinearity between several banana GID1s (Figures 2.5F and

2.5G), suggesting that these genes may have been originated by WGD followed by accelerated mutation rates.

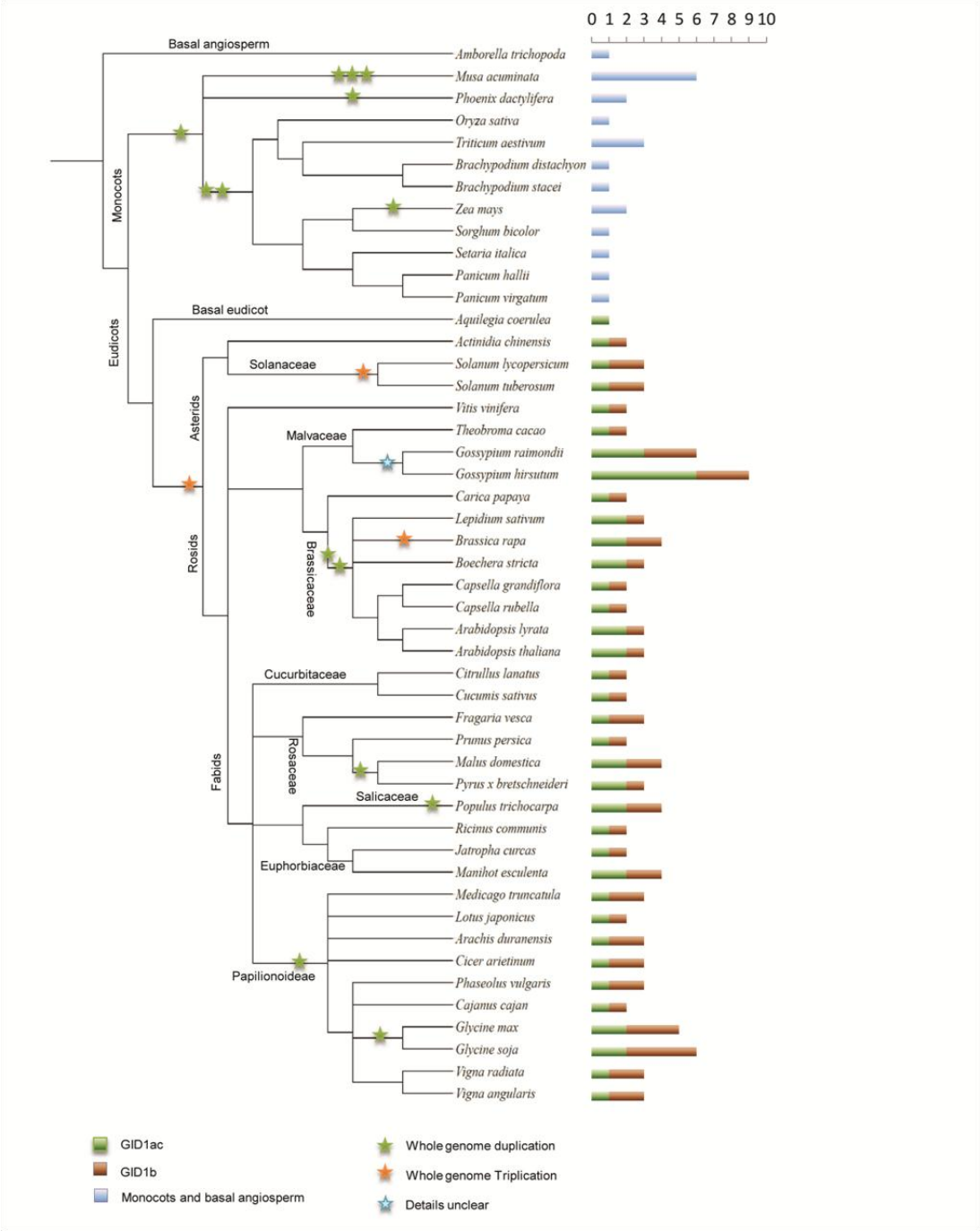


Figure 2.2 Number of GID1 genes in angiosperm species.
Numbers of GID1 genes in each species are represented as horizontal bars, colored according to subfamily. Polyploidization events are marked with colored stars. Species tree was generated using PhyloT (http://phylot.biobyte.de/). Branch lengths do not represent evolutionary time.

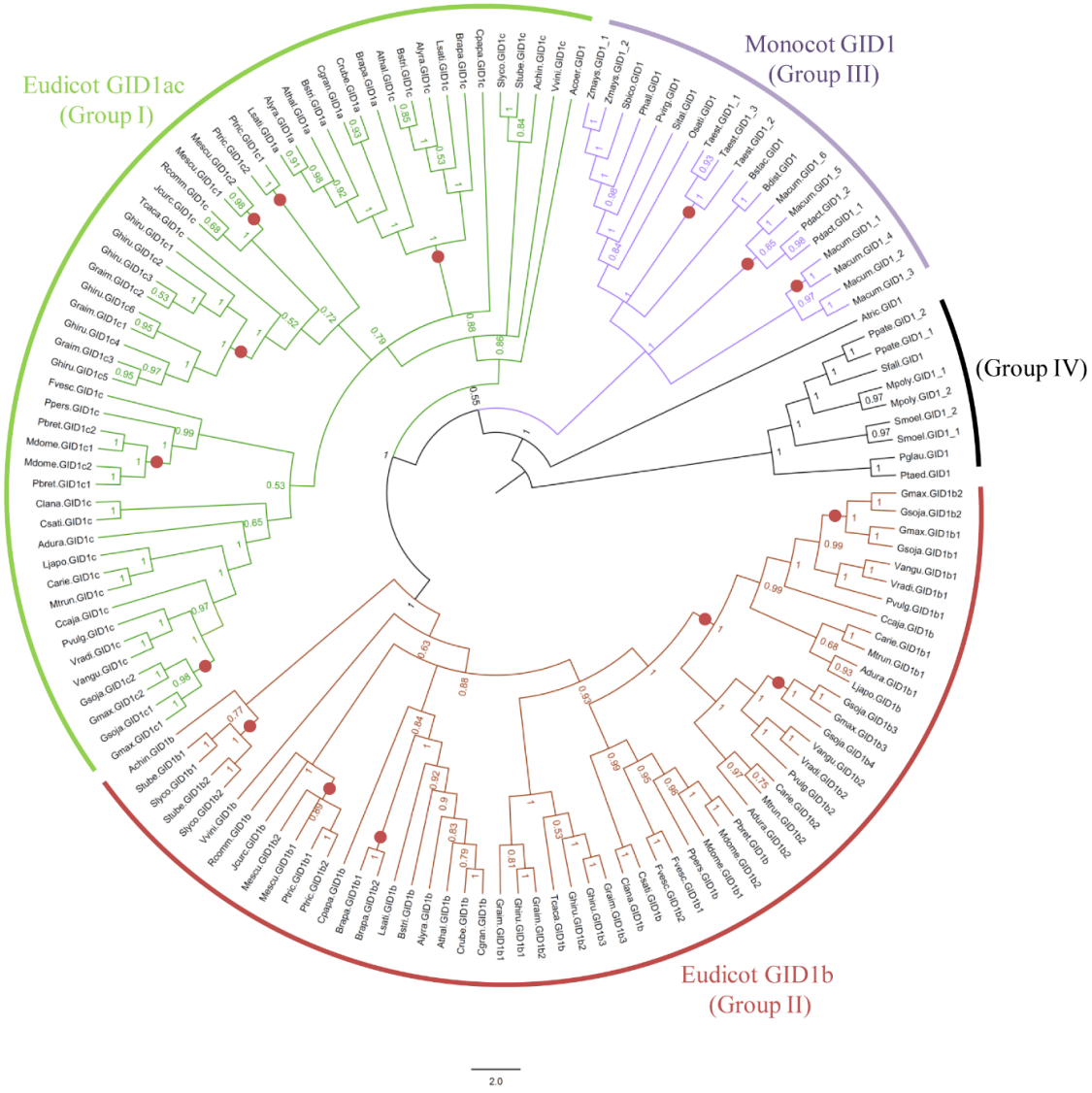


Figure 2.3 Phylogenetic reconstruction of the 141 GID1 proteins identified in 54 plant species.

Multiple sequence alignment was carried with PROMALS3D and phylogenetic reconstruction performed with MrBayes. GID1 proteins were classified in four groups, which are represented in different colors. Red circles show branches potentially originating from polyploidization events. Internal node labels represent Bayesian posterior probabilities.

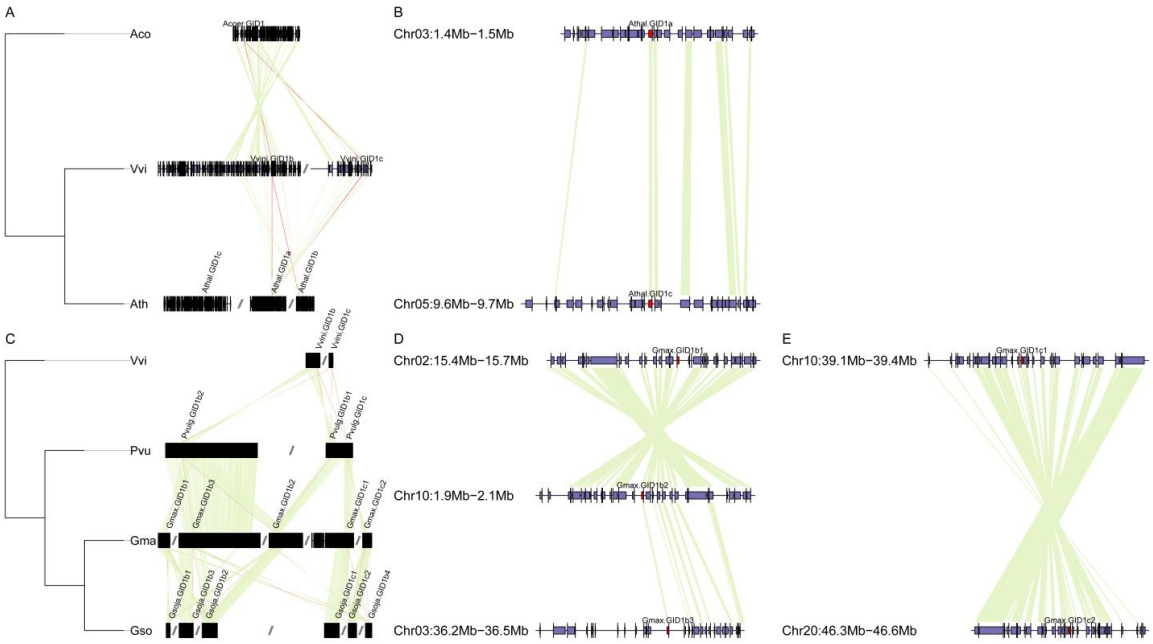


Figure 2.4 Interspecies and intraspecies microsynteny analysis of eudicot GID1s.

Microsynteny (i.e. collinearity) between: *Aq. coerulea* and *Vi. vinifera* and between *Vi. vinifera* and *Ar. Thaliana* GID1s (A); *Ar. thaliana* GID1a and GID1c (B); *Vi. vinifera* and *Ph. Vulgaris* (C); *Glycine* GID1s (D and E). GID1 genes are represented by red arrows and syntenic neighboring genes are connected by green lines. Syntenic GID1 pairs are connected by red lines.

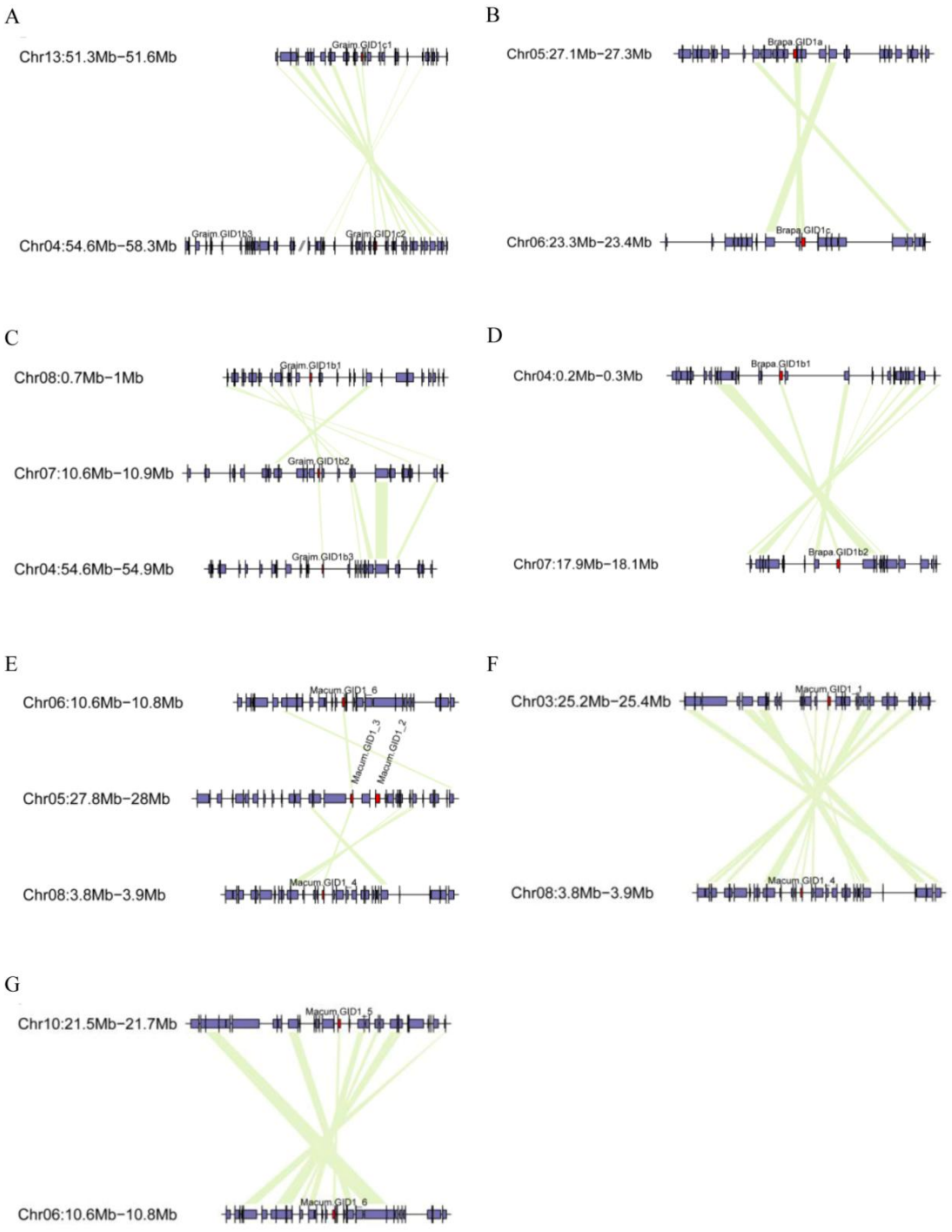


Figure 2.5 Intraspecies microsynteny analysis of GID1s in eudicots and monocots.

Collinearity between *Go. raimondii* GID1s (A and B), *Br. rapa* (C and D) and *Mu. acuminata* (E, F and G). GID1 genes are represented in red and syntenic genes are connected by green lines.

Table 2.3 Ks values of G. soja, G. max and M. acuminata GID1 genes.

Species	Duplicated GID gene pairs	K_s
G. soja	GID1b1/GID1b2	0.192
	GID1b1/GID1b3	1.224
	GID1b1/GID1b4	1.250
	GID1b2/GID1b3	1.366
	GID1b2/GID1b4	1.353
	GID1b3/GID1b4	0.127
	GID1c1/GID1c2	0.143
G. max	GID1b1/GID1b2	0.186
	GID1b1/GID1b3	1.260
	GID1b2/GID1b3	1.381
	GID1c1/GID1c2	0.142
M. acuminata	GID1_1/GID1_2	1.007
	GID1_1/GID1_3	0.941
	GID1_1/GID1_4	1.109
	GID1_1/GID1_5	1.196
	GID1_1/GID1_6	1.725
	GID1_2/GID1_3	0.038
	GID1_2/GID1_4	1.100

	GID1_2/GID1_5	1.246
	GID1_2/GID1_6	1.341
	GID1_3/GID1_4	1.11
	GID1_3/GID1_5	1.254
	GID1_3/GID1_6	1.390
	GID1_4/GID1_5	1.343
	GID1_4/GID1_6	1.548
	GID1_5/GID1_6	0.726

2.3.2 GID1 intron-exon structure is largely conserved throughout the evolution of land plants

In addition to genomic locations and phylogenetic reconstructions, we also investigated the GID1 gene architectures (i.e. intron-exon structures) and intron phases (Figures 2.6 and 2.7). There are three possible intron phases: phase 0, in which an intron is located between two codons; phase 1 and 2, with introns between the first and second codon nucleotides, and between the second and third codon nucleotides, respectively. We found that 104 out of 126 angiosperm GID1s (~82.5 %) with available gene structure have the same basic exon-intron structure, comprising a short and a long exon (average length of 42 bp and 990 bp, respectively) separated by an intervening phase 0 intron of ~610 bp (Figures 2.6 and 2.7). Gene structure conservation is even greater in eudicots, which have 95 out of 108 genes (86.11%) with the canonical architecture (Figure 2.6). Remarkably, gene structure conservation in eudicots is independent of subfamily division, strongly supporting the evolution of eudicot GID1 subgroups from a single ancestor, most likely with the gene structure similar to that of *Acoer.GID1*. The canonical GID1 gene structure is also largely preserved in monocots, although three different architectures are found in banana GID1s (Figure 2.6).

Importantly, the lycophyte GID1s resemble this architecture, indicating that it represents an ancestral state that has been widely conserved throughout angiosperms. This architecture is also found in some gymnosperm and bryophyte GID1s.

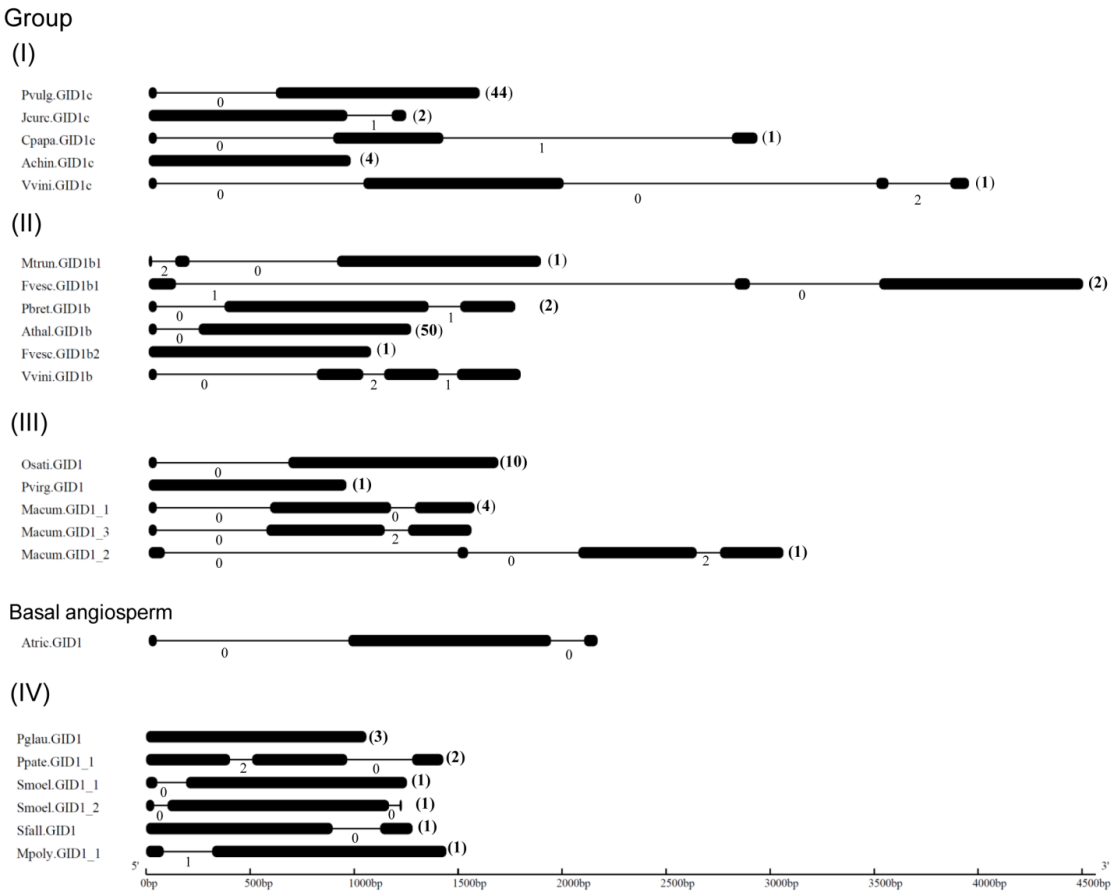
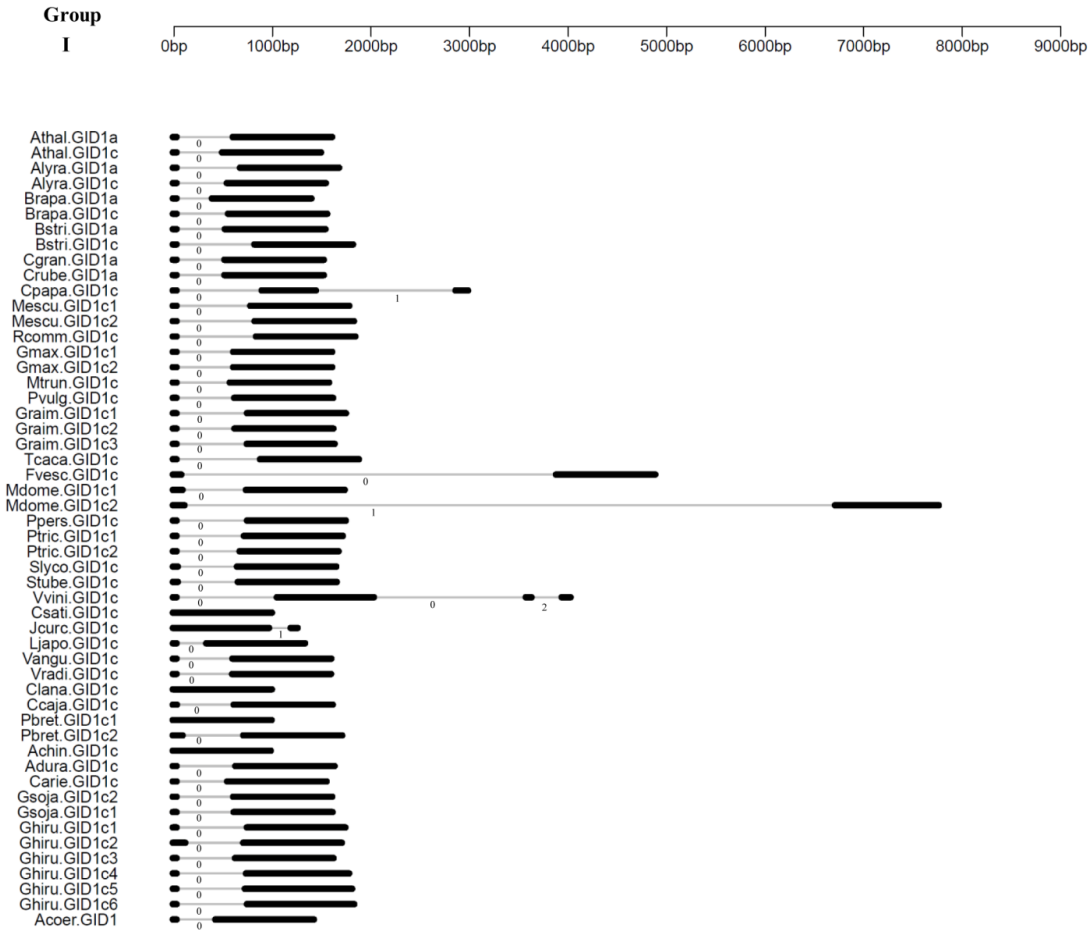


Figure 2.6 Representative GID1 intron-exon architectures.
Thin lines and thick bars represent introns and exons, respectively. Numbers below introns and at the right side of the gene architectures represent intron phases and number of occurrences of each structure, respectively. For comparison purposes, the intron-exon structure of the *Am. trichopoda* GID1, a basal angiosperm, is shown below the gene structures of monocot GID1s.


Group
I
0bp
1000bp
2000bp
3000bp
4000bp
5000bp
6000bp
7000bp
8000bp
9000bp
Athal.GID1a
Athal.GID1c
Alyra.GID1a
Alyra.GID1c
Brapa.GID1a
Brapa.GID1c
Bstri.GID1a
Bstri.GID1c
Cgran.GID1a
Crube.GID1a
Cpapa.GID1c
Mescu.GID1c1
Mescu.GID1c2
Rcomm.GID1c
Gmax.GID1c1
Gmax.GID1c2
Mtrun.GID1c
Pvulg.GID1c
Graim.GID1c1
Graim.GID1c2
Graim.GID1c3
Tcaca.GID1c
Fvesc.GID1c
Mdome.GID1c1
Mdome.GID1c2
Ppers.GID1c
Ptric.GID1c1
Ptric.GID1c2
Slyco.GID1c
Stube.GID1c
Vvini.GID1c
Csati.GID1c
Jcurc.GID1c
Ljapo.GID1c
Vangu.GID1c
Vradi.GID1c
Clana.GID1c
Ccaja.GID1c
Pbret.GID1c1
Pbret.GID1c2
Achin.GID1c
Adura.GID1c
Carie.GID1c
Gsoja.GID1c2
Gsoja.GID1c1
Ghiru.GID1c1
Ghiru.GID1c2
Ghiru.GID1c3
Ghiru.GID1c4
Ghiru.GID1c5
Ghiru.GID1c6
Acoer.GID1

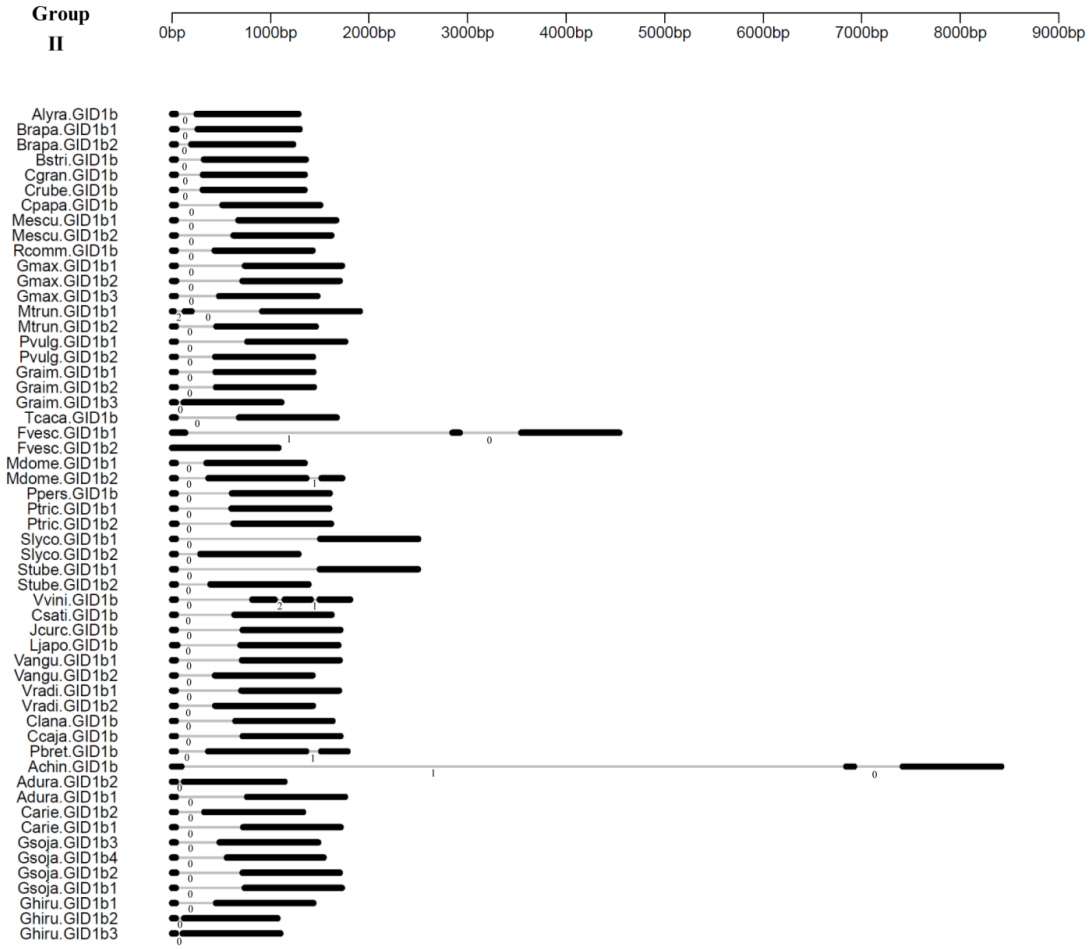

Group II
0bp
1000bp
2000bp
3000bp
4000bp
5000bp
6000bp
7000bp
8000bp
9000bp
Alyra.GID1b
Brapa.GID1b1
Brapa.GID1b2
Bstri.GID1b
Cgran.GID1b
Crube.GID1b
Cpapa.GID1b
Mescu.GID1b1
Mescu.GID1b2
Rcomm.GID1b
Gmax.GID1b1
Gmax.GID1b2
Gmax.GID1b3
Mtrun.GID1b1
Mtrun.GID1b2
Pvulg.GID1b1
Pvulg.GID1b2
Graim.GID1b1
Graim.GID1b2
Graim.GID1b3
Tcaca.GID1b
Fvesc.GID1b1
Fvesc.GID1b2
Mdome.GID1b1
Mdome.GID1b2
Ppers.GID1b
Ptric.GID1b1
Ptric.GID1b2
Slyco.GID1b1
Slyco.GID1b2
Stube.GID1b1
Stube.GID1b2
Vvini.GID1b
Csati.GID1b
Jcurc.GID1b
Ljapo.GID1b
Vangu.GID1b1
Vangu.GID1b2
Vradi.GID1b1
Vradi.GID1b2
Clana.GID1b
Ccaja.GID1b
Pbret.GID1b
Achin.GID1b
Adura.GID1b2
Adura.GID1b1
Carie.GID1b2
Carie.GID1b1
Gsoja.GID1b3
Gsoja.GID1b4
Gsoja.GID1b2
Gsoja.GID1b1
Ghiru.GID1b1
Ghiru.GID1b2
Ghiru.GID1b3

Figure 2.7 Gene structures of 135 GID1 genes.
Thin lines and thick bars represent introns and exons, respectively. Numbers below introns stand for intron phases. GID1 genes of *Tr. aestivum* and *Le. sativum* were not included in this analysis due to the absence of complete gene sequences.

2.3.3 Shared and specific structural features of GID1 subfamilies

Two critical steps in the evolution of GID1s from the HSL family were the loss of catalytic activity and the emergence of GA-binding properties (Hirano et al. 2012). We performed extensive sequence comparisons to better understand the conservation and divergence of GID1 subfamilies. Notably, the characteristic motifs HGGS and GDSSG are conserved in all analyzed GID1s, except for the presence of HGGG instead of HGGS and GDSAG instead of GDSSG in bryophytes (Figure 2.8). Moreover, a SUMO-Interaction Motif (SIM; amino acids W[V/I]LI), that is important for the recognition of SUMOylated DELLA proteins (Conti et al. 2014; Nelis et al. 2015), is also conserved across GID1s, again except in bryophytes. Five GID1s have amino acid substitutions in the first position of the SIM: Met (in Bdist.GID1 and Bstac.GID1), Tyr (in Mtrun.GID1b) and Phe (in Pgalu.GID1 and Ptaed.GID1). Furthermore, the His from the HSL

catalytic triad (Ser-Asp-His) is conserved in Bryophytes and replaced by Val or Ile in all other GID1s (Figures 2.8 and 2.9). A hydrophobic surface in the N-terminal lid (Leu^{18}, Trp^{21}, Val^{29}, Ile^{33}, Leu^{45} and Tyr^{48} in *Ar. thaliana* GID1a) forms a DELLA-binding surface (Murase et al. 2008; Shimada et al. 2008) and is also highly conserved in almost all GID1s (Figures 2.8 and 2.9); we mapped these hydrophobic residues in the alignment and found that Leu^{45} is fully conserved, whereas the remaining positions tolerate substitutions by other hydrophobic residues (Figures 2.8 and 2.9). For example, instead of Ile^{33}, Met^{33} is present in all monocots (except rice) and in very few eudicots, mainly Solanaceae (i.e. Slyco.GID1b1, Slyco.GID1.b2, Stube.GID1.b1, Stube.GID1b2 and Achin.GID1b). Met^{33} seems to be the ancestral state, as it is also present in *Am. trichopoda* and *Aq. coerulea*. Further, because these species have only a single GID1, we infer that Met^{33} can be part of GID1 DELLA binding surfaces. Unexpectedly, some GA interacting residues (i.e. Asn^{218}, Phe^{238}, Val^{239}, Asp^{243}, Arg^{244}, Tyr^{247} in *Ar. thaliana* GID1a) are missing in all banana GID1s (Figure 2.9C). The functional impact of these mutations in the banana GID1s warrants further investigation, for example by expressing banana GID1s in rice GID1 mutants.

Although GID1s display an overall high level of sequence similarity, we were able to clearly define four major groups (one is paraphyletic, see above) (Figure 2.2), supporting some level of functional divergence between them. To better understand the conservation patterns in the family, we sought to analyze conserved motifs that influence GID1-GA interaction (Figure 2.9). There are five motifs conserved in groups I, II and III that contain GA-interacting residues. Three of these motifs are well known: Motif 1, which encompasses the SIM, GA- and DELLA-interacting residues; Motif 3, which contains the HGGS motif and; Motif 4 harboring the GDSSG domain and GA interacting residues. The remaining two motifs (i.e. Motif 5 and 6) have other GA-binding residues (Figure 2.9). We also identified motifs specific to the GID1ac (Motif 2), GID1b (Motif 7) and monocot GID1s (Motif 8) sub-groups (Figures 2.8 and 2.9), which

correspond to the same alignment region. Their within-group conservation patterns suggest that they might play important subfamily-specific roles.

To further explore the mechanistic differences of GID1ac and GID1b, we have also predicted functionally divergent sites using three different programs (see methods for details). A total of nine alignment positions were predicted to be functionally divergent between GID1ac and GID1b groups (Table 2.4). We mapped these residues on the tertiary structure of the wild type Athal.GID1a (Figure 2.10A) and modeled this structure with *in silico* mutations reflecting the divergent sites with respect to GID1b (Figure 2.10B). Two sites, Asp^{102} and Gly^{103} in Athal.GID1a (Ser^{102} and Thr^{103} in Athal.GID1b) are inside specific motifs discussed above (Figures 2.8 and 2.9). Interestingly, the positions 102 and 103 are much more conserved in the GID1b (Ser^{102} and Thr^{103} in Athal.GID1b) than in the GID1ac subfamily (Figures 2.8 and 2.9), supporting that these sites are under type I functional divergence (Gu 2001, 1999) (Table 2.4). Four other functionally divergent sites were highly conserved within GID1ac and GID1b subgroups but with important amino acid changes (e.g. Leu^{323} in GID1ac and Phe^{323} in GID1b) between them, suggesting type II functional divergence (Table 2.4) (Gu 2001, 1999).

Intriguingly, one of the functionally divergent sites, Leu^{323} (in GID1ac, corresponding to Phe^{323} and Leu^{330} in GID1b and rice GID1, respectively), is involved in hydrophobic interactions with GA (Murase et al. 2008). Previous studies in rice demonstrated that mutation of GA interacting residues, including the substitution of Leu^{330} for Ile^{330} or Ala^{330}, reduced the GID1 affinity and specificity for GAs (Hirano et al. 2007; Shimada et al. 2008; Xiang et al. 2011). We performed *in silico* mutagenesis with FoldX to estimate the effects of converting Leu^{323} into Phe^{323} on the GA binding pocket of Athal.GID1a (PDB: 2ZSH and 2ZSI), followed by docking analysis of mutated 2ZSH with GA_3 and mutated 2ZSI with GA_4. The native and mutant docked GID1-GAs had similar hydrogen bond lengths (Murase et al. 2008). In previously reported structures,

the O7-2 atom of GA_3/GA_4 formed a hydrogen bond to the Oγ atom of Ser^{191} with a distance of 2.9 Å (for GA_3) and 3.2 Å (for GA_4) (Figures 2.11A and 2.11C). These distances became longer in the mutated structures (3.5 Å for both GA_3 and GA_4), although still within the range of hydrogen bonds (Figures 2.11B and 2.11D). Interestingly, we found that Phe^{323} is closer to GA_3/GA_4 than Leu^{323} with a significant difference of ~1Å, suggesting that Phe^{323} in GID1b confers a tighter binding pocket that could be related with the higher affinity of GID1b for GA_3/GA_4. Interestingly, the higher affinity of GID1b has been attributed to a partially closed configuration of the N-terminal lid (Yamamoto et al. 2010). We hypothesize that Phe^{323} may also contribute to this phenomenon.

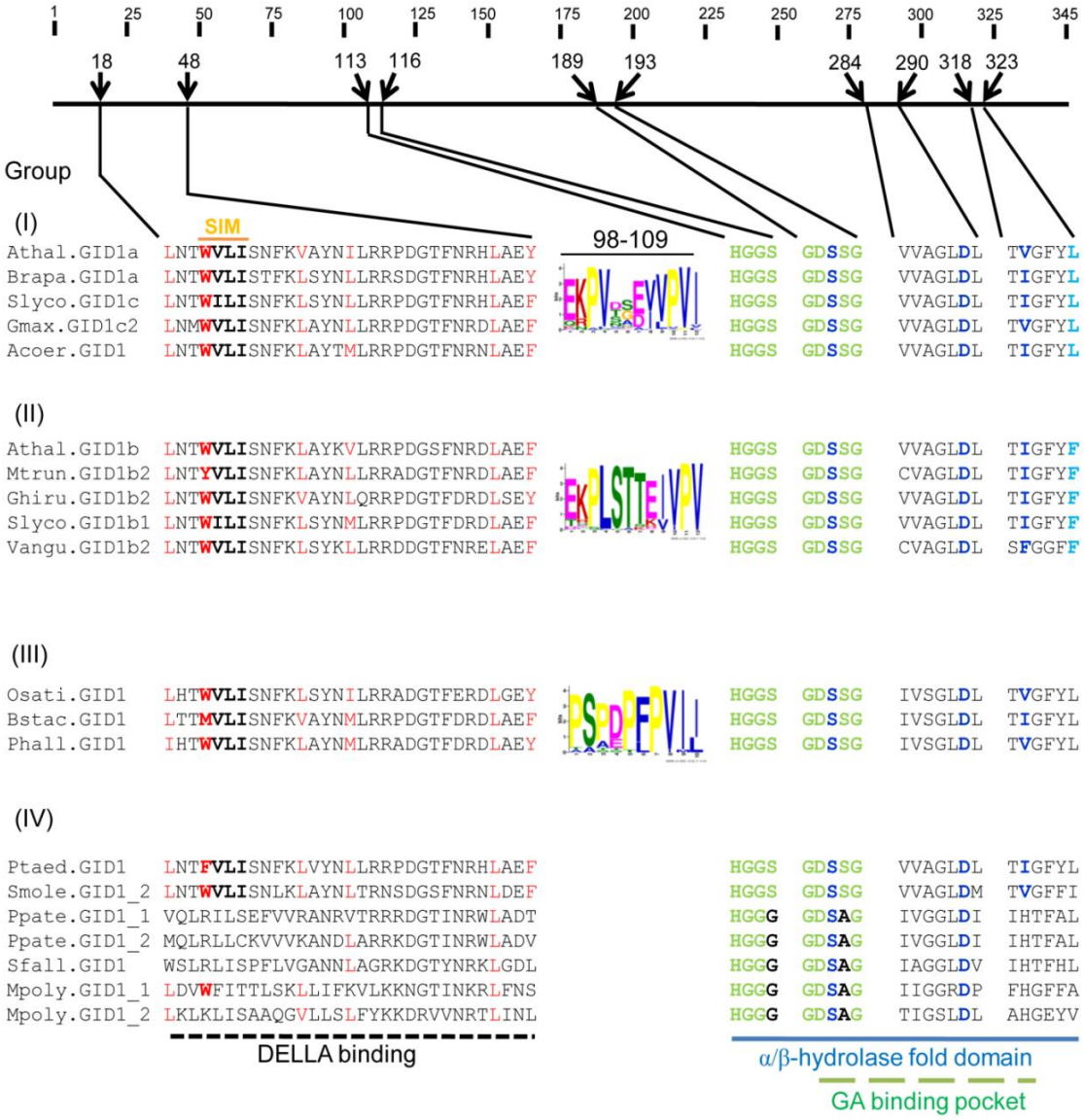
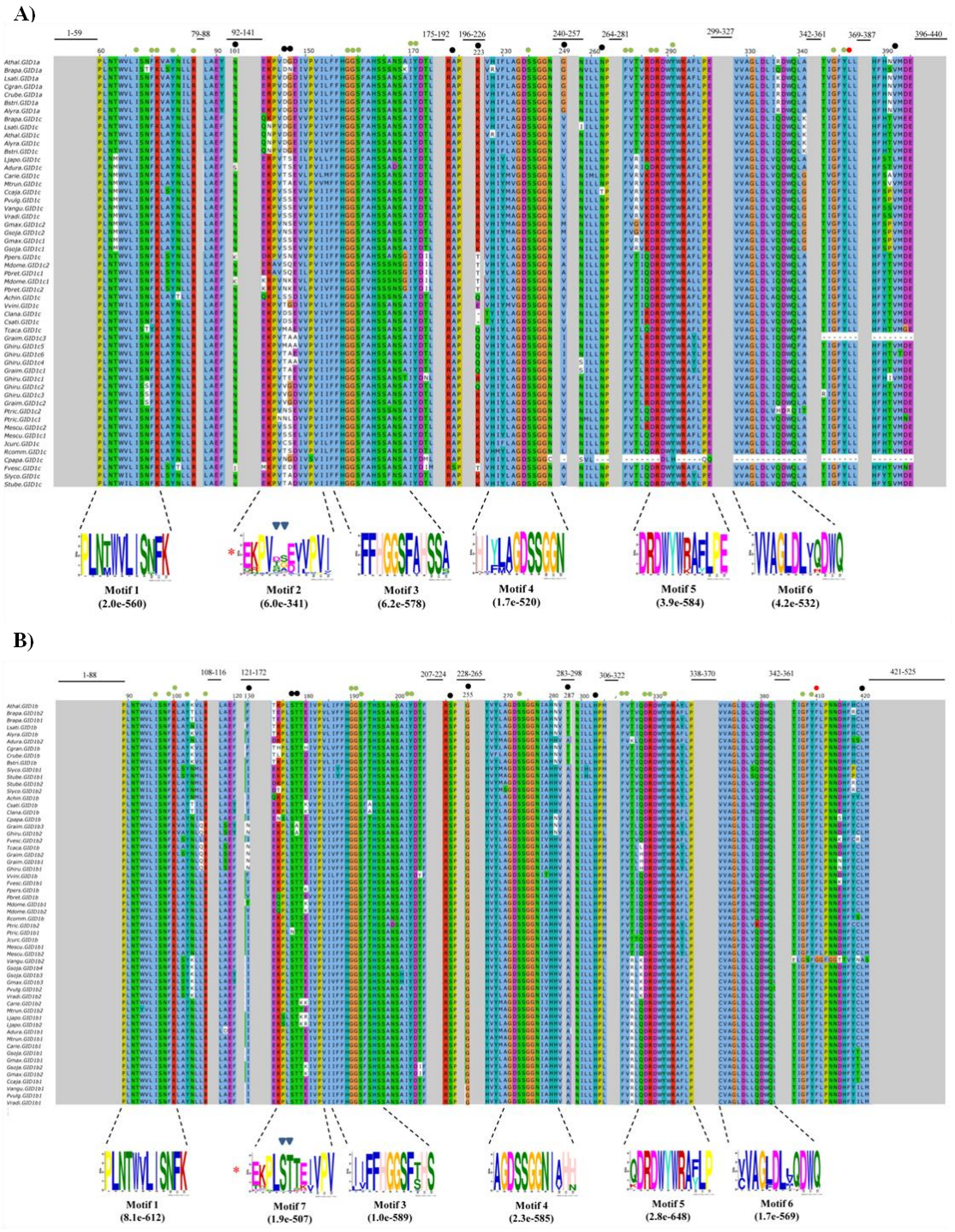


Figure 2.8 Schematic representative GID1s.
The thick line with coordinates represents the positions of important GID1 features, using the *Ar. thaliana* GID1a as a reference. HGGS and GDSSG motifs are shown in green; the SIM motif (W[V/I]LI) is also marked. DELLA binding residues are shown in red and the 'catalytic triad' involved in GA binding (Ser, Asp, and Val/Ile) are in dark blue. Subgroup specific motifs are represented with logos. One functionally divergent site between GID1ac and GID1b, which is also a GA interacting residue, is represented in sky blue. A complete representation of all functionally divergent sites is available in Table 2.4 and Figure 2.10.


A)
Motif 1
(2.0e-560)
Motif 2
(6.0e-341)
Motif 3
(6.2e-578)
Motif 4
(1.7e-520)
Motif 5
(3.9e-584)
Motif 6
(4.2e-532)
B)
Motif 1
(8.1e-612)
Motif 7
(1.9e-507)
Motif 3
(1.0e-589)
Motif 4
(2.3e-585)
Motif 5
(2.8e-648)
Motif 6
(1.7e-569)

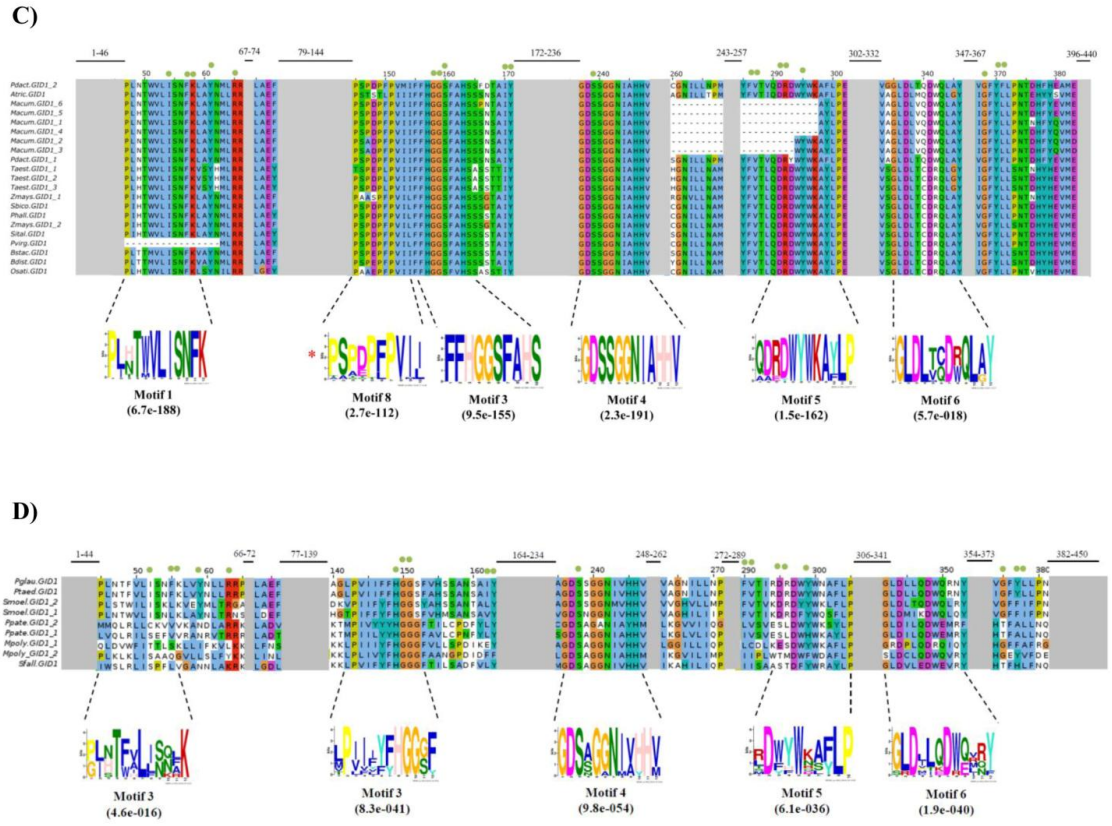


Figure 2.9 Group-wise multiple sequence alignment of GID1 proteins.
Proteins were aligned with PROMALS3D and visualized with Jalview. Panels represent different GID1 groups, as follows: group I (A, GID1ac), group II (B, GID1b), group III (C, monocot GID1s) and group IV (D). GA-interacting and functionally divergent residues are marked with green and black circles, respectively. One of the functionally divergent sites, is also a GA-interacting residue is marked with a red circle. The conserved motifs and group-specific motifs (indicated by red asterisks) are shown below the alignments. Motifs were predicted with MEME.

Table 2.4 Functionally divergent sites in GID1ac and GID1b groups.

Position in GID1a crystal structure	GID1ac group*	GID1b group*	Functional Divergence Type
Loop	Asn^{58}	Phe^{58}	Type I
Loop	Asp^{102}	Ser^{102}	Type I
Loop	Gly^{103}	Thr^{103}	Type I

Loop	Ala^{150}	Ser^{150}	Type II
Loop	Lys^{178}	Gly^{178}	Type I
α3 helix	Gly^{205}	Thr^{205}	Type U
Loop	Asn^{218}	His^{218}	Type II
η2 helix	Leu^{323}	Phe^{323}	Type II
α7 helix	Val^{333}	Leu^{333}	Type II

* Athal.GID1a and Athal.GID1b were used as references to define positions in GID1ac and GID1b, respectively. Type I represents amino acid configurations that are highly conserved in one sub-family but highly variable in another sub-family, or vice versa. Type II represents amino acid configurations that are highly conserved within sub-families but but differ between them. Type U represents amino acid configurations at many residues are not so clear-cut, and they are regarded as unclassified.

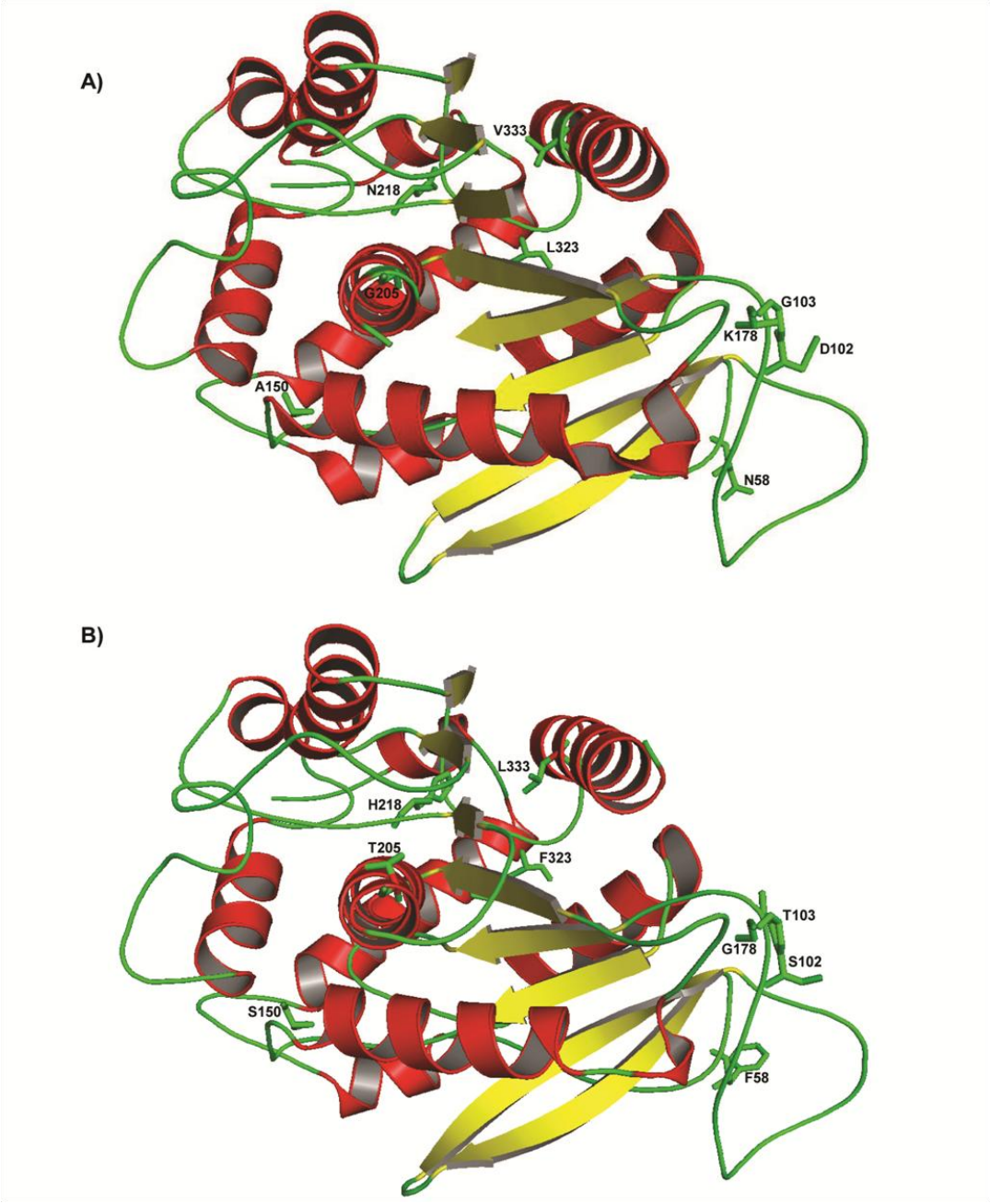


Figure 2.10 Localization of critical amino acids in the 3D structures.
Functionally divergent sites were mapped on the crystal structure (as sticks) of the native *Ar. thaliana* GID1a (2ZSH) **(A)** and of a mutated GID1a structure with *in silico* mutations reflecting the divergent sites with respect to GID1b **(B).**

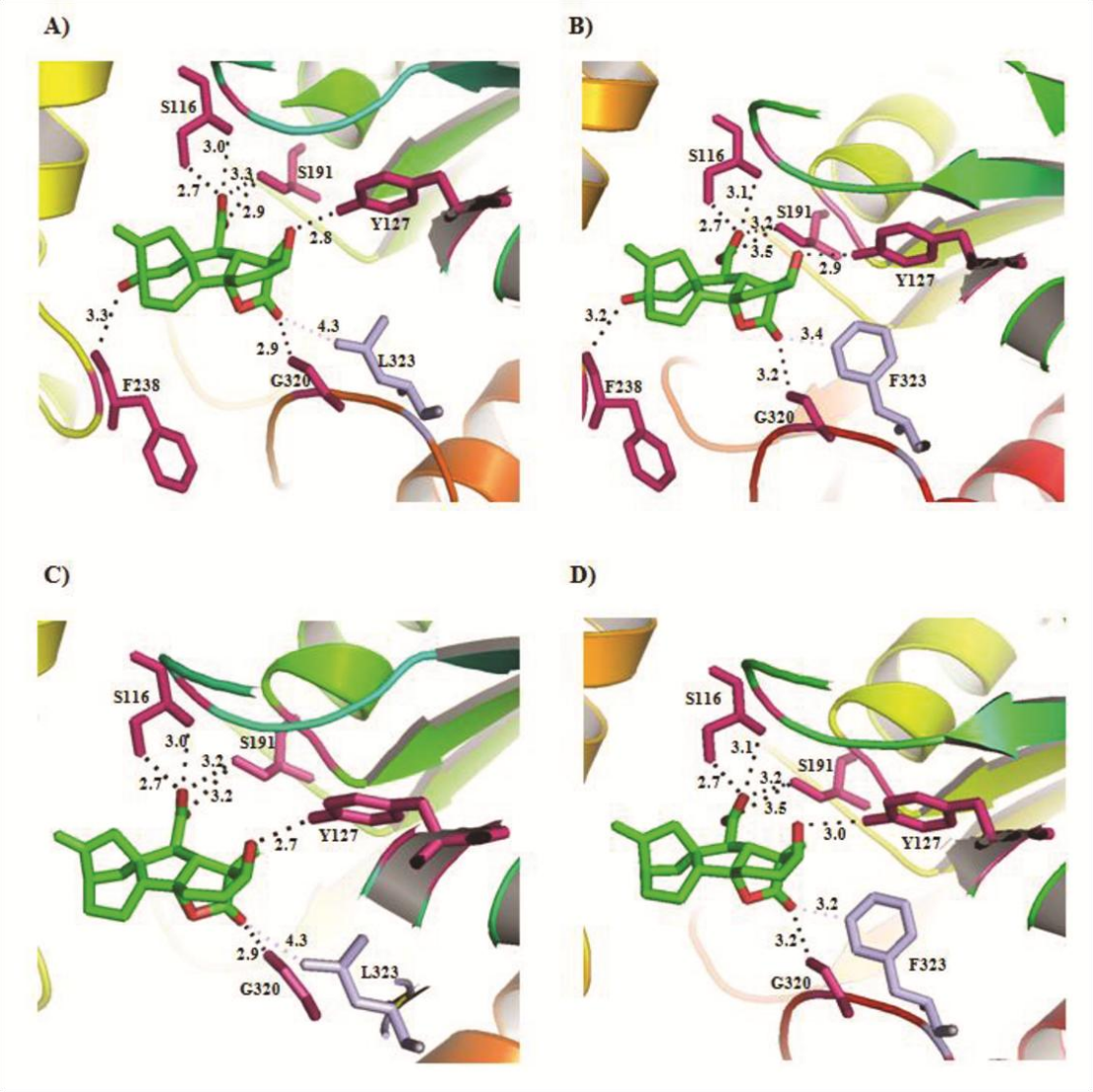


Figure 2.11 Comparison of GID1a-GA in the native versus mutated GID1a-GA. **A)** native structure of 2ZSH-GA_3; **B)** mutated structure of 2ZSH-GA_3; **C)** native structure of 2ZSI-GA_4 and; **D)** mutated structure of 2ZSH-GA_4. GA_3/GA_4 are shown in green, with oxygen atoms in red. Leu^{323} of GID1a and its corresponding amino acid in GID1b (Phe^{323}) are shown in slate light blue; distances of these amino acids to GA_3/GA_4 are also represented in dotted slate light blue. Hydrogen bonds are shown in dotted black lines. Mutations were generated *in silico* by changing amino acids in the GID1a structure at functionally divergent sites, with respect to GID1b.

2.3.4 GID1 subfamilies have substantial divergence in their expression patterns

Given the expansion and diversification of GID1 subfamilies, we sought to study their expression profiles as a means to understand their functional specialization.

In *Phaseolus vulgaris*, we found a generally high GID1 expression, particularly of GID1b, in underground tissues (i.e. roots and nodules) (Figure 2.12A). In soybean, we also found a remarkable activation of a GID1b paralog (*Gmax.GID1b3*) in roots and nodules, in addition to a conspicuous expression peak in flowers (not observed in common bean) (Figure 2.12B). Interestingly, the homeolog of *Gmax.GID1b3* that originated in the *Glycine* WGD was lost in *Gl. max* (but not in *Gl. soja*) (Figure 2.3). We speculate that the specialized expression profile of *Gmax.GID1b3* and the lack of a close homeolog may be involved in the selection of traits of agricultural interest. Interestingly, our group has shown that soybean GID1b genes are highly expressed in the embryonic axes of dry seeds and down-regulated as germination proceeds, as opposed to GID1c genes (Bellieny-Rabelo et al. 2016) (Figure 2.12C). This scenario can be part of a system to detect low GA levels and trigger important signaling processes until the canonical GID1c-mediated GA signaling pathway is activated in the onset of germination (Bellieny-Rabelo et al. 2016; Griffiths et al. 2006; Hauvermale et al. 2015; Nakajima et al. 2006). We have also analyzed the expression of GID1 genes in a third legume species, *Me. truncatula* (Figure 2.12D). Similarly to what was observed in soybean and common bean, GID1b is also more expressed than GID1c in most *Me. truncatula* tissues and at least one GID1b gene is highly expressed in roots and nodules (Figure 2.12D). Interestingly, *Mtrun.GID1b1* transcripts accumulate during seed maturation, whereas *Mtrun.GID1c1* transcription is reduced, similarly to what was observed in soybean (Bellieny-Rabelo et al. 2016), but not in *Ar. thaliana* (Voegele et al. 2011). Taken together, these results indicate that the expression divergence of GID1b and GID1ac in seed development and germination predates the divergence of soybean and *Medicago* [~52 MYA (Kumar et al. 2017)]. Nevertheless, this scenario will be clearer only when more gene expression data of GID1 genes during seed development and germination become available for other species. In particular, the relationship between transcriptional divergence and GID1b subfamily expansion in legumes remain to be directly addressed.

As found in the other species discussed above, *Athal.GID1b* is more expressed than *Athal.GID1a* and *Athal.GID1c* in roots, whereas GID1ac expression is dominant in leaves, flowers and developing seeds (Figure 2.12E). Thus, our results indicate that the specialization of GID1b towards roots, nodules and dry seeds supports the scenario where GID1b, probably because of its higher affinity for GA, is important under low GA concentrations and/or at tissues with high GA sensitivity (Tanimoto 1987, 1994). It has been shown that GA regulates root elongation and thickening (Tanimoto and Hirano 2013). In root elongation, GA action specifically takes place at the endodermis (Ubeda-Tomás et al. 2008). In addition, GA also influences the number and length of root meristems (Tanimoto and Hirano 2013; Ubeda-Tomás et al. 2009). Thus, GID1b probably specialized to mediate GA signaling in eudicot roots in the presence of low hormone concentrations.

We have also investigated GID1 expression in monocots and in the lycophyte *Se. moellendorffii*, in which there are often fewer GID1 genes and no family subdivision (Figures 2.12F to 2.12H). Interestingly, we found that GID1 is highly expressed in all tissues, with at least one GID1 gene expressed in high levels in roots. Collectively, our results show that the high expression of GID1 in roots dates back to the origin of the canonical GA perception system in lycophytes (Hirano et al. 2007; Nelson and Steber 2016), far earlier than the emergence of seed plants. In species without GID1 subfamilies (e.g. monocots and lycophytes), all tissues have at least one expressed GID1 and roots have high GID1 transcriptional level. We hypothesize that after the divergence of GID1ac and GID1b subfamilies, the former retained roles more related to the ancestral GA perception system (already present in lycophytes), and was later recruited to more modern features like seed germination. On the other hand, GID1b specialized in conditions of low GA concentrations (e.g. roots and germinating legume seeds) through biased gene expression and mutations that increased its affinity for GA (Nakajima et al. 2006). Further, with GID1ac mediating canonical GA signaling, GID1b was also free to integrate alternative

GA perception mechanisms, such as GA-independent DELLA binding and non-proteolytic GA signaling (Fuentes et al. 2012; Yamamoto et al. 2010).

Important aspects regarding the origin of GA perception system remain to be elucidated. While the lycophyte *Se. moellendorffii* and the bryophyte *Physcomitrella patens* have some of the key components of the canonical GA perception machinery, several lines of evidence indicate the absence of a functional GA signaling pathway in the bryophytes (Hayashi et al. 2010; Hirano et al. 2007; Nelson and Steber 2016; Vesty et al. 2016; Yasumura et al. 2007), such as: 1) *Ph. patens* GID1 and DELLA do not interact; 2) *Ph. patens* GID1 does not interact with GA; 3) DELLA-deficient *Ph. patens* strains do not exhibit derepressed growth like that observed in DELLA-deficient angiosperms; 4) *Ph. patens* DELLA does not suppress GA response in rice, although it can repress growth in *Ar. thaliana*. On the other hand, certain bioactive diterpene hormones from early steps of the GA biosynthesis pathway (e.g. *ent*-kaurene) promote spore germination in *Ph. patens* (Hayashi et al. 2010; Vesty et al. 2016). Interestingly, ABA can inhibit *Ph. patens* spore germination, strongly supporting the existence of a diterpene/ABA signaling module before the emergence of vascular plants, although apparently not as prominent as that found in seed plants (Hayashi et al. 2010). The key genes involved in diterpene perception in *Ph. patens* remain to be elucidated and could involve direct diterpene recognition by GRAS domain proteins (e.g. DELLA), which were already diversified early in the evolution of land plants (Zhang et al. 2012).

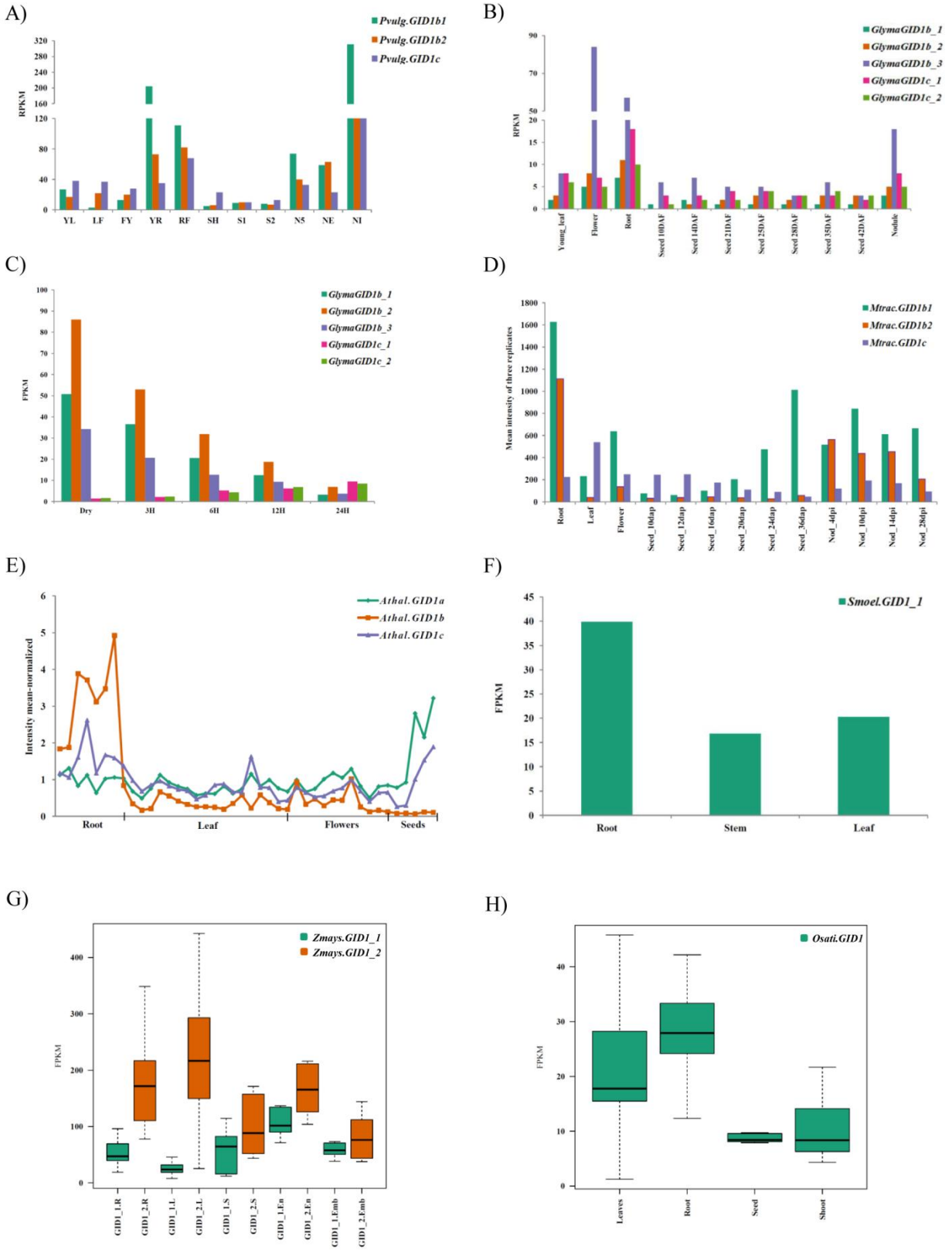

A)
Pvulg.GID1b1
Pvulg.GID1b2
Pvulg.GID1c
RPKM
YL LF FY YR RF SH S1 S2 N5 NE NI
B)
GlymaGID1b_1
GlymaGID1b_2
GlymaGID1b_3
GlymaGID1c_1
GlymaGID1c_2
RPKM
Young_leaf Flower Root Seed 10DAF Seed 14DAF Seed 21DAF Seed 25DAF Seed 28DAF Seed 35DAF Seed 42DAF Nodule
C)
GlymaGID1b_1
GlymaGID1b_2
GlymaGID1b_3
GlymaGID1c_1
GlymaGID1c_2
FPKM
Dry 3H 6H 12H 24H
D)
Mtrac.GID1b1
Mtrac.GID1b2
Mtrac.GID1c
Mean intensity of three replicates
Root Leaf Flower Seed_10dap Seed_12dap Seed_16dap Seed_20dap Seed_24dap Seed_36dap Nod_4dpi Nod_10dpi Nod_14dpi Nod_28dpi
E)
Athal.GID1a
Athal.GID1b
Athal.GID1c
Intensity mean-normalized
Root Leaf Flowers Seeds
F)
SmoeI.GID1_1
FPKM
Root Stem Leaf
G)
Zmays.GID1_1
Zmays.GID1_2
FPKM
GID1_1R GID1_2-R GID1_1-L GID1_2-L GID1_1-S GID1_2-S GID1_1-En GID1_2-En GID1_1-Emb GID1_2-Emb
H)
Osati.GID1
FPKM
Leaves Root Seed Shoot

Figure 2.12 Expression analysis of GID1 genes in different species.
A) *Phaseolus vulgaris*: fully expanded 2nd trifoliate leaf tissue (YL), developing leaf tissue (LF), young flowers prior to floral emergence (FY), whole roots at the 2nd trifoliate stage (YR), whole roots (RF), heart stage seeds with 3 to 4 mm across and ~7 mg (SH), stage 1 seeds with 6 to 7 mm across and ~50 mg (S1), stage 2 seeds with 8 to 10 mm across and 140 to 150 mg (S2), pre-fixing (effective) nodules at 5 days after inoculation (N5), effectively fixing nodules at 21 days after inoculation (NE), ineffectively fixing nodules at 21 days after inoculation (NI). These abbreviations were defined by Common bean gene expression atlas (O'Rourke et al. 2014). **B)** *Medicago truncatula*: roots, leaves, flowers, seeds (at 10, 12, 16, 20, 24 and 36 days after pollination, DAP) and nodules (at 4, 10, 14 and 28 days old plant after inoculation with *Sinorhizobium meliloti*, DPI); **C)** *Glycine max*: young leaves, flowers, root, nodules and developing seeds (at 10, 14, 21, 25, 28, 35 and 42 days after flowering, DAF); **D)** *Arabidopsis thaliana*: root, leaf, flowers, seeds; **E)** *Gl. max* during seed germination: embryonic axes at 0, 3, 6, 12 and 24 hours after imbibition (H); **F)** *Selaginella moellendorffii*: root, stem and leaf; **G)** *Zea mays*: roots (R), leaves (L), seeds (S), endosperm (En) and embryo (Emb). Gene expression information of *Smoel.GID1_2* was not found; **H)** *Oryza sativa*: leaves, root, seed, shoot.

2.4 REFERENCES

Aleman L, et al. 2008. Functional analysis of cotton orthologs of GA signal transduction factors GID1 and SLR1. Plant Molecular Biology 68: 1-16. doi: 10.1007/s11103-008-9347-z

Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25: 3389-3402.

Bailey TL, et al. 2009. MEME Suite: Tools for motif discovery and searching. Nucleic Acids Research 37: 202-208. doi: 10.1093/nar/gkp335

Bellieny-Rabelo D, et al. 2016. Transcriptome analysis uncovers key regulatory and metabolic aspects of soybean embryonic axes during germination. Scientific Reports 6: 36009. doi: 10.1038/srep36009

Benedito VA, et al. 2008. A gene expression atlas of the model legume Medicago truncatula. Plant Journal 55. doi: 10.1111/j.1365-313X.2008.03519.x

Berman HM, et al. 2000. The protein data bank. Nucleic Acids Research 28. doi: 10.1093/nar/28.1.235

Capra JA, Singh M 2008. Characterization and prediction of residues determining protein functional specificity. Bioinformatics 24. doi: 10.1093/bioinformatics/btn214

Chandler PM, et al. 2008. Characterization of gibberellin receptor mutants of barley (Hordeum vulgare L.). Molecular Plant 1: 285-294. doi: 10.1093/mp/ssn002

Conti L, et al. 2014. Small Ubiquitin-like Modifier Protein SUMO Enables Plants to Control Growth Independently of the Phytohormone Gibberellin. Developmental Cell 28: 102-110. doi: 10.1016/j.devcel.2013.12.004

D'Hont A, et al. 2012. The banana (Musa acuminata) genome and the evolution of monocotyledonous plants. Nature 488: 213-217. doi: 10.1038/nature11241

Dill A, Thomas SG, Hu J, Steber CM, Sun T-P 2004. The Arabidopsis F-box protein SLEEPY1 targets gibberellin signaling repressors for gibberellin-induced degradation. The Plant Cell 16: 1392-1405. doi: 10.1105/tpc.020958

Feenstra KA, Pirovano W, Krab K, Heringa J 2007. Sequence harmony: detecting functional specificity from alignments. Nucleic Acids Research 35. doi: 10.1093/nar/gkm406

Fleet CM, Sun TP 2005. A DELLAcate balance: The role of gibberellin in plant morphogenesis. Current Opinion in Plant Biology 8: 77-85. doi: 10.1016/j.pbi.2004.11.015

Fu X, et al. 2004. The Arabidopsis mutant sleepy1 gar2-1 protein promotes plant growth by increasing the affinity of the SCF SLY1 E3 ubiquitin ligase for DELLA protein substrates. The Plant Cell 16: 1406-1418. doi: 10.1105/tpc.021386.dependent

Fuentes S, et al. 2012. Fruit Growth in Arabidopsis Occurs via DELLA-Dependent and DELLA-Independent Gibberellin Responses. The Plant Cell 24: 3982-3996. doi: 10.1105/tpc.112.103192

Gaston D, Susko E, Roger AJ 2011. A phylogenetic mixture model for the identification of functionally divergent protein residues. Bioinformatics 27: 2655-2663. doi: 10.1093/bioinformatics/btr470

Gomi K, et al. 2004. GID2, an F-box subunit of the SCF E3 complex, specifically interacts with phosphorylated SLR1 protein and regulates the gibberellin-dependent degradation of SLR1 in rice. Plant Journal 37: 626-634. doi: 10.1111/j.1365-313X.2003.01990.x

Griffiths J, et al. 2006. Genetic characterization and functional analysis of the GID1 gibberellin receptors in Arabidopsis. The Plant Cell 18: 3399-3414. doi: 10.1105/tpc.106.047415

Grosdidier Al, Zoete V, Michielin O 2011. SwissDock, a protein-small molecule docking web service based on EADock DSS. Nucleic Acids Research 39: 270-277. doi: 10.1093/nar/gkr366

Gu X 2001. Maximum-likelihood approach for gene family evolution under functional divergence. Molecular Biology and Evolution 18.

Gu X 1999. Statistical methods for testing functional divergence after gene duplication. Molecular Biology and Evolution 16.

Guy L, Roat Kultima J, Andersson SGE 2010. genoPlotR: comparative gene and genome visualization in R. Bioinformatics 26: 2334-2335. doi: 10.1093/bioinformatics/btq413

Haas BJ, Delcher AL, Wortman JR, Salzberg SL 2004. DAGchainer: a tool for mining segmental genome duplications and synteny. Bioinformatics 20: 3643-3646. doi: 10.1093/bioinformatics/bth397

Hauvermale AL, Tuttle KM, Takebayashi Y, Seo M, Steber CM 2015. Loss of Arabidopsis thaliana Seed Dormancy is Associated with Increased Accumulation of the GID1 GA Hormone Receptors. Plant and Cell Physiology. doi: 10.1093/pcp/pcv084

Hayashi K-i, et al. 2010. Endogenous Diterpenes Derived from ent-Kaurene, a Common Gibberellin Precursor, Regulate Protonema Differentiation of the Moss Physcomitrella patens. PLANT PHYSIOLOGY 153: 1085-1097. doi: 10.1104/pp.110.157909

Hirano K, Aya K, Matsuoka M, Ueguchi-Tanaka M 2012. Molecular Determinants that Convert Hormone Sensitive Lipase into Gibberellin Receptor. Protein and Peptide Letters 19: 180-185. doi: 10.2174/092986612799080248

Hirano K, et al. 2007. The GID1-mediated gibberellin perception mechanism is conserved in the Lycophyte Selaginella moellendorffii but not in the Bryophyte Physcomitrella patens. The Plant Cell 19: 3058-3079. doi: 10.1105/tpc.107.051524

Hu B, et al. 2015. GSDS 2.0: An upgraded gene feature visualization server. Bioinformatics 31: 1296-1297. doi: 10.1093/bioinformatics/btu817

Irwin JJ, Shoichet BK 2006. ZINC – A Free Database of Commercially Available Compounds for Virtual Screening. J. Chem. Inf. Model 45: 177-182.

Iuchi S, et al. 2007. Multiple loss-of-function of Arabidopsis gibberellin receptor AtGID1s completely shuts down a gibberellin signal. Plant Journal 50: 958-966. doi: 10.1111/j.1365-313X.2007.03098.x

Jaillon O, et al. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449: 463-467. doi: 10.1038/nature06148

Jiao Y, et al. 2012. A genome triplication associated with early diversification of the core eudicots. Genome Biology 13: R3. doi: 10.1186/gb-2012-13-1-r3

Kumar S, Stecher G, Suleski M, Hedges SB 2017. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. Molecular Biology and Evolution 34: 1812-1819. doi: 10.1093/molbev/msx116

Li A, et al. 2013. Molecular characterization of three GIBBERELLIN-INSENSITIVE DWARF1 homologous genes in hexaploid wheat. Journal of Plant Physiology 170: 432-443. doi: 10.1016/j.jplph.2012.11.010

Livne S, Weiss D 2014. Cytosolic activity of the gibberellin receptor GIBBERELLIN INSENSITIVE DWARF1A. Plant and Cell Physiology 55: 1727-1733. doi: 10.1093/pcp/pcu104

McGinnis KM, et al. 2003. The Arabidopsis SLEEPY1 gene encodes a putative F-box subunit of an SCF E3 ubiquitin ligase. The Plant Cell 15: 1120-1130. doi: 10.1105/tpc.010827.increased

Mühlhausen S, Kollmar M 2013. Whole genome duplication events in plant evolution reconstructed and predicted using myosin motor proteins. BMC Evolutionary Biology 13. doi: 10.1186/1471-2148-13-202

Murase K, Hirano Y, Sun T-p, Hakoshima T 2008. Gibberellin-induced DELLA recognition by the gibberellin receptor GID1. Nature 456: 459-463. doi: 10.1038/nature07519

Nakajima M, et al. 2006. Identification and characterization of Arabidopsis gibberellin receptors. Plant Journal 46: 880-889. doi: 10.1111/j.1365-313X.2006.02748.x

Nelis S, Conti L, Zhang C, Sadanandom A 2015. A functional Small Ubiquitin-like Modifier (SUMO) interacting motif (SIM) in the gibberellin hormone receptor GID1 is conserved in cereal crops and disrupting this motif does not abolish hormone dependency of the DELLA-GID1 interaction. Plant signaling & behavior 10: e987528. doi: 10.4161/15592324.2014.987528

Nelson SK, Steber CM. 2016. Gibberellin hormone signal perception: down-regulating DELLA repressors of plant growth and development. In. Annual Plant Reviews, Volume 49: John Wiley & Sons, Ltd. p. 153-188.

O'Rourke Ja, et al. 2014. An RNA-Seq based gene expression atlas of the common bean. BMC Genomics 15. doi: 10.1186/1471-2164-15-866

Olszewski N, Sun T-P, Gubler F 2002. Gibberellin signaling: biosynthesis, catabolism, and response pathways. The Plant Cell 14 Suppl: S61-S80. doi: 10.1105/tpc.010476.GAs

Pei J, Kim B-H, Grishin NV 2008. PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Research 36: 2295-2300. doi: 10.1093/nar/gkn072
Peng J, et al. 1997. The Arabidopsis GAI gene defines a signaling pathway that negatively regulates gibberellin responses. Genes and Development 11: 3194-3205. doi: 10.1101/gad.11.23.3194
Price MN, Dehal PS, Arkin AP 2010. FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS ONE 5. doi: 10.1371/journal.pone.0009490
Ronquist F, Huelsenbeck JP 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572-1574. doi: 10.1093/bioinformatics/btg180
Sato S, et al. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485: 635-641. doi: 10.1038/nature11119
Schmid M, et al. 2005. A gene expression map of Arabidopsis thaliana development. Nature Genetics 37: 501–506. . doi: 10.1038/ng1543
Schranz ME 2006. Independent Ancient Polyploidy Events in the Sister Families Brassicaceae and Cleomaceae. THE PLANT CELL ONLINE 18: 1152-1165. doi: 10.1105/tpc.106.041111
Schymkowitz J, et al. 2005. The FoldX web server: an online force field. Nucleic Acids Research 33: 382-388. doi: 10.1093/nar/gki387
Severin AJ, Cannon SB, Graham MM, Grant D, Shoemaker RC 2011. Changes in twelve homoeologous genomic regions in soybean following three rounds of polyploidy. The Plant Cell 23: 3129-3136. doi: 10.1105/tpc.111.089573
Severin AJ, et al. 2010. RNA-Seq Atlas of Glycine max: a guide to the soybean transcriptome. BMC plant biology 10. doi: 10.1186/1471-2229-10-160
Shimada A, et al. 2008. Structural basis for gibberellin recognition by its receptor GID1. Nature 456: 520-523. doi: 10.1038/nature07546
Stamatakis A 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312-1313. doi: 10.1093/bioinformatics/btu033
Stelpflug SC, et al. 2016. An Expanded Maize Gene Expression Atlas based on RNA Sequencing and its Use to Explore Root Development. plant genome. doi: 10.3835/plantgenome2015.04.0025
Suyama M, Torrents D, Bork P 2006. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Research 34: 609-612. doi: 10.1093/nar/gkl315
Suzuki H, et al. 2009. Differential expression and affinities of Arabidopsis gibberellin receptors can explain variation in phenotypes of multiple knock-out mutants. Plant Journal 60: 48-55. doi: 10.1111/j.1365-313X.2009.03936.x
Tanimoto E 1987. Gibberellin-Dependent Root Elongation in Lactuca sativa Recovery from Growth Retardant-Suppressed Elongation with Thickening by Low concentration of GA3. Plant and Cell Physiology 28: 963–973. doi: 10.1093/oxfordjournals.pcp.a077399
Tanimoto E 1994. Interaction of Gibberellin A3 and Ancymidol in the Growth and Cell-Wall Extensibility of Dwarf Pea Roots. Plant and Cell Physiology 35: 1019–1028. doi: 10.1093/oxfordjournals.pcp.a078689
Tanimoto E, Hirano K. 2013. Role of Gibberellins in Root Growth. In. Plant Roots: The Hidden Half, Fourth Edition. Boca Raton, FL, USA: CRC Press. p. 13-11--13-14.
Tuskan GA, et al. 2006. The Genome of Black Cottonwood, Populus trichocarpa (Torr. & Gray). Science 313: 1596-1604. doi: 10.1126/science.1128691
Ubeda-Tomás S, et al. 2009. Gibberellin Signaling in the Endodermis Controls Arabidopsis Root Meristem Size. Current Biology 19: 1194-1199. doi: 10.1016/j.cub.2009.06.023
Ubeda-Tomás S, et al. 2008. Root growth in Arabidopsis requires gibberellin/DELLA signalling in the endodermis. Nature Cell Biology 10: 625-628. doi: 10.1038/ncb1726
Ueguchi-Tanaka M, et al. 2005. GIBBERELLIN INSENSITIVE DWARF1 encodes a soluble receptor for gibberellin. Nature 437: 693-698. doi: 10.1038/nature04028
Vesty EF, et al. 2016. The decision to germinate is regulated by divergent molecular networks in spores and seeds. New Phytologist 211. doi: 10.1111/nph.14018

Voegele A, Linkies A, Muller K, Leubner-Metzger G 2011. Members of the gibberellin receptor gene family GID1 (GIBBERELLIN INSENSITIVE DWARF1) play distinct roles during Lepidium sativum and Arabidopsis thaliana seed germination. Journal of Experimental Botany 62: 5131-5147. doi: 10.1093/jxb/err214

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ 2009. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. Bioinformatics 25: 1189-1191. doi: 10.1093/bioinformatics/btp033

Willige BC, et al. 2007. The DELLA domain of GA INSENSITIVE mediates the interaction with the GA INSENSITIVE DWARF1A gibberellin receptor of Arabidopsis. The Plant Cell 19: 1209-1220. doi: 10.1105/tpc.107.051441

Xiang H, et al. 2011. Thermodynamic characterization of OsGID1-gibberellin binding using calorimetry and docking simulations. J Mol Recognit: 275-282. doi: 10.1002/jmr.1049

Xu X, et al. 2011. Genome sequence and analysis of the tuber crop potato. Nature 475: 189-195. doi: 10.1038/nature10158

Yamamoto Y, et al. 2010. A rice gid1 suppressor mutant reveals that gibberellin is not always required for interaction between its receptor, GID1, and DELLA proteins. The Plant Cell 22: 3589-3602. doi: 10.1105/tpc.110.074542

Yang Z 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Bioinformatics 13: 555-556. doi: 10.1093/bioinformatics/13.5.555

Yasumura Y, Crumpton-Taylor M, Fuentes S, Harberd NP 2007. Step-by-Step Acquisition of the Gibberellin-DELLA Growth-Regulatory Mechanism during Land-Plant Evolution. Current Biology 17. doi: 10.1016/j.cub.2007.06.037

Zhang D, Iyer LM, Aravind L 2012. Bacterial GRAS domain proteins throw new light on gibberellic acid response mechanisms. Bioinformatics 28: 2407-2411. doi: 10.1093/bioinformatics/bts464

Zhang T, et al. 2015. Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. Nature Biotechnology 33: 531-537. doi: 10.1038/nbt.3207

Zhu Y, et al. 2017. Global transcriptome analysis reveals extensive gene remodeling, alternative splicing and differential transcription profiles in non-seed vascular plant Selaginella moellendorffii. BMC Genomics 18: 1-15. doi: 10.1186/s12864-016-3266-1

Chapter 3 Transcriptional landscape of soybean (*Glycine max*) embryonic axes during germination in the presence of paclobutrazol, a gibberellin biosynthesis inhibitor

3.1 INTRODUCTION

Gibberellins (GAs) constitute a large family of diterpenoid compounds that are ubiquitous in higher plants. Some GAs regulate processes such as seed germination, root and stem elongation, leaf expansion, flower and fruit development (Olszewski et al. 2002, Tanimoto and Hirano 2013). Seed germination typically starts with imbibition and ends with testa rupture, followed by emergence of the embryonic axis (Bewley 1997). During this relatively short period, metabolic activity resumes, mitochondria and DNA damaged during desiccation are repaired, stored mRNAs are translated or degraded and new transcriptional programs are activated. This complex series of interconnected events is fueled by the mobilization of stored reserves and gradually shifts towards photosynthesis and autotrophic growth (Bewley et al. 2013, Rajjou et al. 2012, Weitbrecht et al. 2011).

Over the past decades, seminal studies unequivocally demonstrated the role of GA in promoting seed germination (Urbanova and Leubner-Metzger 2016), in particular because GA-deficient mutants (e.g. *ga1-3* and *ga2-1*) often require exogenous GA to germinate (Groot et al. 1987, Koornneef and van der Veen 1980). Further, the inhibition of radicle emergence in the presence of GA biosynthesis inhibitors (e.g. uniconazole and paclobutrazol, PBZ) indicates that GA is essential for seed germination (Jacobsen and Olszewski 1993, Karssen et al. 1989, Nambara et al. 1991). PBZ is a plant growth retardant that blocks GA biosynthesis by inhibiting kaurene oxidase (Hedden and Graebe 1985). Other key GA biosynthesis enzymes are GA20- and GA3-oxidases (GA20ox and GA3ox, respectively), whereas GA2-oxidases (GA2ox) inactivate GA. During late germination, GA is synthesized at the radicle, hypocotyls and micropylar endosperm (Ogawa et al. 2003). GA is recognized by soluble receptors of the

GIBBERELLIN INSENSITIVE DWARF1 (GID1) family (Ueguchi-Tanaka et al. 2005), which comprises the subfamilies GID1ac and GID1b in eudicots. Although very similar at the primary sequence level, different lines of evidence indicate that these subfamilies are functionally divergent (Gazara et al. 2018, Tanimoto and Hirano 2013, Voegele et al. 2011). The GA–GID1 complex promotes the degradation of DELLA transcriptional repressors via the 26S proteasome pathway (Murase et al. 2008). Further, enhanced germination has been reported in loss-of-function DELLA-mutants (Kucera et al. 2005). GA is also notorious for its antagonistic interactions with ABA, a well-known seed germination inhibitor. In addition, GA has also been proposed to positively interact with brassinosteroids (BRs) and ethylene, which are ABA antagonists during seed germination (Holdsworth et al. 2008, Kucera et al. 2005, Linkies et al. 2009).

During seed germination, GA enhances embryo growth by promoting cell elongation and weakening of the surrounding tissues (Kucera et al. 2005, Ogawa et al. 2003). Several genes regulated by GA or DELLA have been identified during *Arabidopsis* seed germination, seedling and floral development (Cao et al. 2006, Nemhauser et al. 2006, Ogawa et al. 2003, Zentella et al. 2007). In addition, various genes related to hormone pathways and cell wall metabolism were modulated by GA (Cao et al. 2006, Ogawa et al. 2003). Despite the valuable information accumulated on the biochemical details of GA signaling and interactions with other hormones, little is known about the transcriptional programs driven by GA in germinating seeds of species other than *A. thaliana*. To date, only one report investigated the transcriptome of embryonic axes during soybean (*Glycine max*) germination (Bellieny-Rabelo et al. 2016). Although this study showed a conspicuous activation of GA biosynthesis genes, it does not allow one to distinguish GA-driven transcriptional alterations. In the present work, we report the transcriptome of soybean embryonic axes during seed germination in the presence of the GA biosynthesis inhibitor PBZ, aiming to uncover the genes that are regulated by GA. We show that PBZ: 1) up-regulates several photosynthesis genes; 2) modulates the expression of numerous genes involved

in the biosynthesis, signaling and transport of other hormones, suggesting an intensive hormonal cross-talk during germination; 3) modulates the expression of several genes encoding cell wall modifying enzymes, supporting their roles in embryo cell expansion during germination and; 4) represses several transcription factors (TFs) in a time-specific fashion, indicating that these TFs might drive the transcriptional reprogramming mediated by GA during germination.

3.2 MATERIAL AND METHODS

3.2.1 Plant material and growth conditions

G. max seeds (BRS-284, from EMBRAPA, Brazil) were used in this study. Seeds were surface sterilized with 70% ethanol for 1 minute and with commercial bleach (1% v/v) for 3 minutes, followed by three washes with sterile distilled water (30 seconds per wash). Seeds were germinated in 15 cm Petri dishes with 2 g of sterile cotton in two conditions: in the presence of 30 ml of sterile water (control) or sterile water with 200 µM paclobutrazol (Sigma Aldrich). Seeds were allowed to germinate in an incubation chamber at 28°C and 12/12h photoperiod (dark/light). We used three plates per sample, with 20 seeds per plate. Embryonic axes from dry seeds were also collected. For total RNA extraction, seeds were harvested at 12, 24 and 36 HAI in control and PBZ treated conditions. Embryonic axes were separated from cotyledons and immediately placed in RNA*later*™ (Qiagen) until RNA extraction. RNA was extracted from harvested embryonic axes using RNeasy Plant Mini Kit (Qiagen) according to manufacturer instructions. Three independent biological replicates of each condition were used.

3.2.2 RNA purification, sequencing and analysis

RNA-Seq libraries were prepared using the TruSeq RNA Sample Preparation Kit v2 and submitted to 1x100bp single-end sequencing on a HiSeq 2500 instrument at LaCTAD (UNICAMP, Campinas, Brazil). Read quality was assessed by FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were

aligned on *G. max* cv. Williams 82 reference genome version 2 (Wm82.a2.v1) using novoalign (V3.06.05; http://www.novocraft.com). Gene expression levels were calculated with cufflinks v2.1.1 (Trapnell et al. 2012) and normalized by reads per kilobase of transcript per million mapped reads (RPKM). Genes with RPKM greater than or equal to one were considered expressed. The differential expression between Control vs PBZ at 12 HAI, 24 HAI and 36 HAI were determined by cuffdiff v2.2.1 (Trapnell et al. 2012). Genes with at least two-fold difference in expression and q-value ≤ 0.05 were considered differentially expressed. Enrichment of Gene Ontology (GO) term was performed using agriGO (v2.0) with hypergeometric test, corrected by the Hochberg FDR method (FDR ≤ 0.05) (Tian et al. 2017). Redundant GO terms were removed with REViGO (Supek et al. 2011). KOBAS 3.0 (Wu et al. 2006) was used to assess the enrichment of DEGs in KEGG pathways (Fisher's exact test, $P < 0.05$). The list of expressed genes (i.e. RPKM ≥ 1) were used as the background set for GO and KEGG enrichment analyses. *G. max* TFs were obtained from the Plant Transcription Factor Database (PlantTFDB) (Jin et al. 2017). Figure 3.1 represents the workflow for RNA-seq data analysis. The datasets generated in this study have been deposited in the NCBI Gene Expression Omnibus database, under the accession number GSE112872.

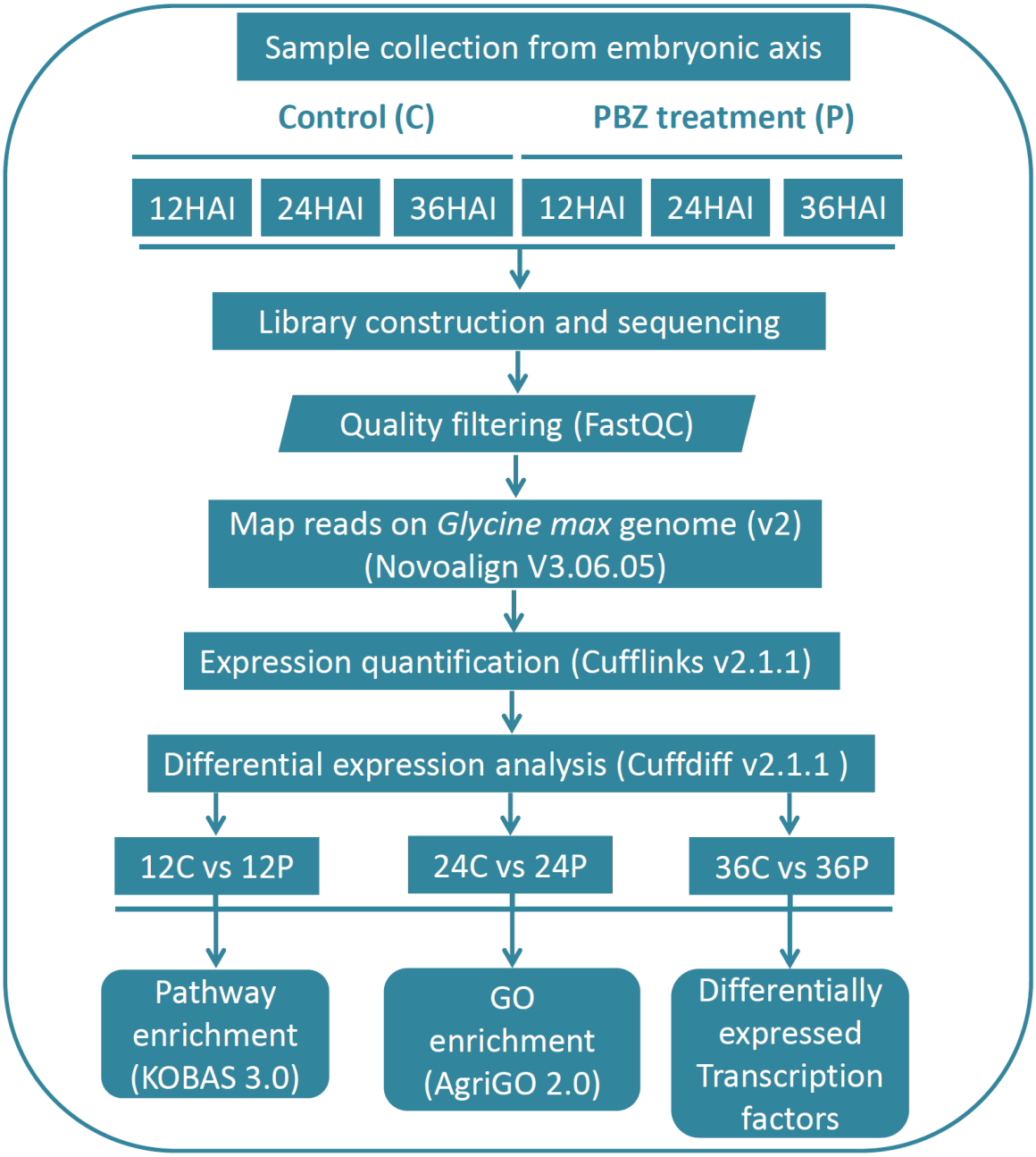


Figure 3.1 Workflow for RNA-Seq data analysis

3.3 RESULTS AND DISCUSSION

3.3.1 Transcriptome sequencing and functional analysis of differentially expressed genes

We conducted an initial assay to investigate the effects of PBZ on soybean seed germination. As expected, PBZ administration reduced embryonic axes length, fresh weight and dry weight, resulting in a delay in germination (Figure 3.2). Embryonic axes at 12, 24 and 36 hours after imbibition (HAI) were carefully separated from the cotyledons and submitted to RNA extraction, library preparation and sequencing on an Illumina HiSeq 2500 instrument (see methods for details). A total of 18 libraries (three biological replicates, with or without PBZ) were sequenced, resulting in a total of 14 to 67 million reads per sample (Table 3.1). High-quality reads were mapped to the soybean reference genome (Wm82.a2.v1) and used for downstream analysis. Overall, 97.2% of the reads mapped to the reference genome (Table 3.1). In general, we found good correspondence between the biological replicates (Figure 3.3) and high pair-wise correlations (0.95 to 0.99) (APPENDIX B1). Genes with RPKM (Reads Per Kilobase per Million mapped reads) greater than or equal to 1 were considered expressed. In total, 29,204, 29,467, 31,065, 30,887, 32,636 and 32,466 genes were found to be expressed in 12C (control), 12P (PBZ), 24C, 24P, 36C and 36P, respectively (Figure 3.4A). Approximately 62.43% of the soybean protein-coding genes (34,990 genes) were expressed in at least at one time point (Figure 3.2A; APPENDIX B2), which is comparable to a previously published soybean germination transcriptome (with 33,305 expressed genes) (Bellieny-Rabelo et al. 2016).

We compared the transcriptional profiles of PBZ-treated seeds at each time point with their respective controls and found a total of 85, 486 and 307 differentially expressed genes (DEGs) at 12, 24 and 36 HAI, respectively (APPENDIX B3). Because PBZ is a GA antagonist, PBZ down- and up-regulated genes (i.e. PBZ-down and PBZ-up, respectively) are likely those induced and

repressed by GA. The absolute number of genes down-regulated by PBZ and their ratios to up-regulated genes increased along germination (Figure 3.4B**;** APPENDIX B3). In DEG counts, 24 HAI was the most notable time point (297 and 189 up- and down-regulated genes, respectively; Figure 3.4B). On the other hand, 12 HAI had the lowest number of DEGs (58 and 27 up- and down-regulated genes), indicating that GA transcriptional programs are mostly activated between 12 and 24 HAI and decrease afterwards, when most seeds had completed germination (Figure 3.2). Notably, we found 63 genes that are significantly up-regulated at 24 HAI and down-regulated at 36 HAI by PBZ (APPENDIX B4**,** Figure 3.5). About 41% (26 out of 63) of these genes are related with photosynthesis and their up-regulation by PBZ at 24 HAI followed by a down-regulation at 36 HAI might be a strategy to anticipate the transition to autotrophic growth in the absence of energetic resources resulting from proper GA signaling. This gene set encodes chloroplast ATP synthase subunits, RuBisCO, chloroplast ribosomal proteins, DNA-directed RNA polymerase subunit beta (*rpoC1*), YCF3 and several photosystem I and II subunits. Decrease in the expression of plastidial RNA polymerases (i.e. *rpoB* and *rpoC1)* caused aberrant chloroplast development and diminish photoautotrophic growth in *A. thaliana* (Hricova 2006). Chloroplast *YCF3* encodes a thylakoid protein that is essential for photosystem I complex biogenesis in tobacco (Ruf et al. 1997) and *Chlamydomonas reinhardtii* (Boudreau 1997). The regulation of photosynthesis genes by GA has also been recently demonstrated in rice seedlings under submergence (Xiang et al. 2017). Interestingly all these 26 genes are nuclear encoded copies of genes that are located in the soybean chloroplast (Reference Sequence: NC_007942). Most of these copies seem to be functional, as they encode proteins with high sequence coverage (68 to 100%) and similarity (78 to 100%) with their plastidial counterparts (Table 3.2). Similarly, 17 out of 63 genes encode proteins similar to those encoded by mitochondrial genes (Reference sequence: JX463295) (Table 3.2). Collectively, these genes might integrate a system to reduce the dependence on cotyledonary reserves and optimize ATP production. Out of these 43 genes with organellar copies, 41 have been assigned

to soybean reference chromosomes, suggesting that they are not annotated as nuclear genes due to contamination of organelle DNA fragments.

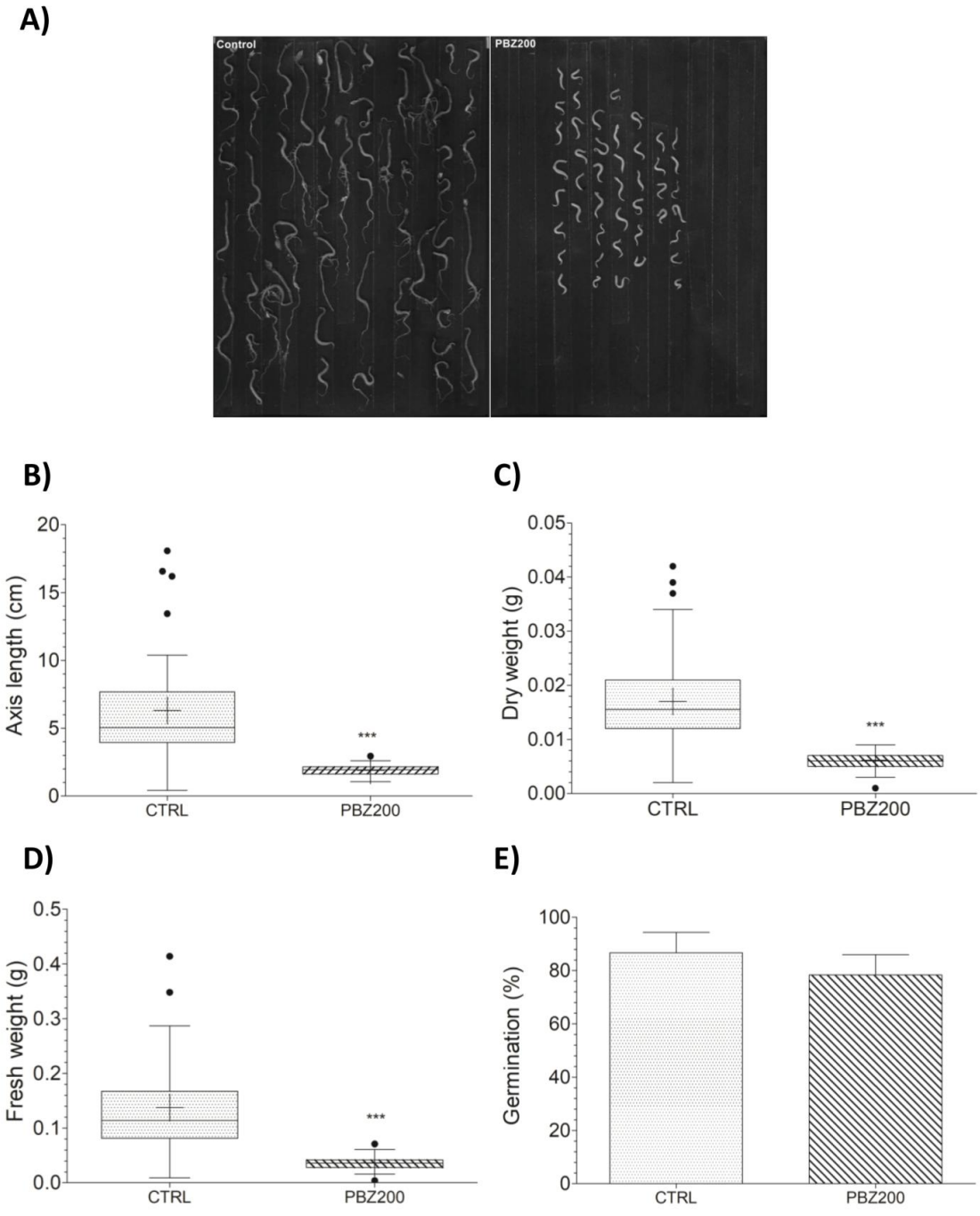


Figure 3.2 PBZ delays soybean seed germination.
Soybean seeds (BRS-284) were allowed to germinate and grow in an incubation chamber under 28°C temperature, and 12/12h photoperiod (dark/light) for 7 days. Seeds were germinated in the presence of 30 ml of sterile water (control) or sterile water with 200 µM

paclobutrazol (PBZ). **(A)** Photographs of embryonic axis submitted or not to PBZ treatment. **(B)** Embryonic axis length, **(C)** Dry weight, **(D)** Fresh weight and **(E)** Seed germination. Asterisks above boxplot demonstrate significant difference ($p<0.0001$, Student´s T-test).

Table 3.1 Read mapping of RNA-Seq reads to the soybean reference genome (Wm82.a2.v1).

Sample_ID	Total Number of reads	Total mapped (%)	Multiple mapped (%)	Unique mapped (%)
12CR1	55,357,068	97.8	62.5	35.3
12CR2	53,215,247	97.6	13.6	84
12CR3	31,799,519	97.4	49.2	48.2
12PR1	50,137,072	96.5	54.1	42.4
12PR2	62,557,238	96.5	43.1	53.4
12PR3	67,178,211	96.7	38	58.7
24CR1	17,541,588	97.6	8.7	88.9
24CR2	21,773,672	97.5	7.8	89.7
24CR3	30,037,674	97.1	8.2	88.9
24PR1	28,242,334	98.2	38.2	60
24PR2	42,910,390	97.8	35	62.8
24PR3	33,643,809	97.7	48.9	48.8
36CR1	18,034,685	96.7	7.1	89.6
36CR2	17,877,074	96.7	7.3	89.4
36CR3	33,954,443	97.5	33.5	64
36PR1	16,691,755	96	7.2	88.8

36PR2	14,374,866	96.7	8.2	88.5
36PR3	25,399,244	96.8	8.3	88.5

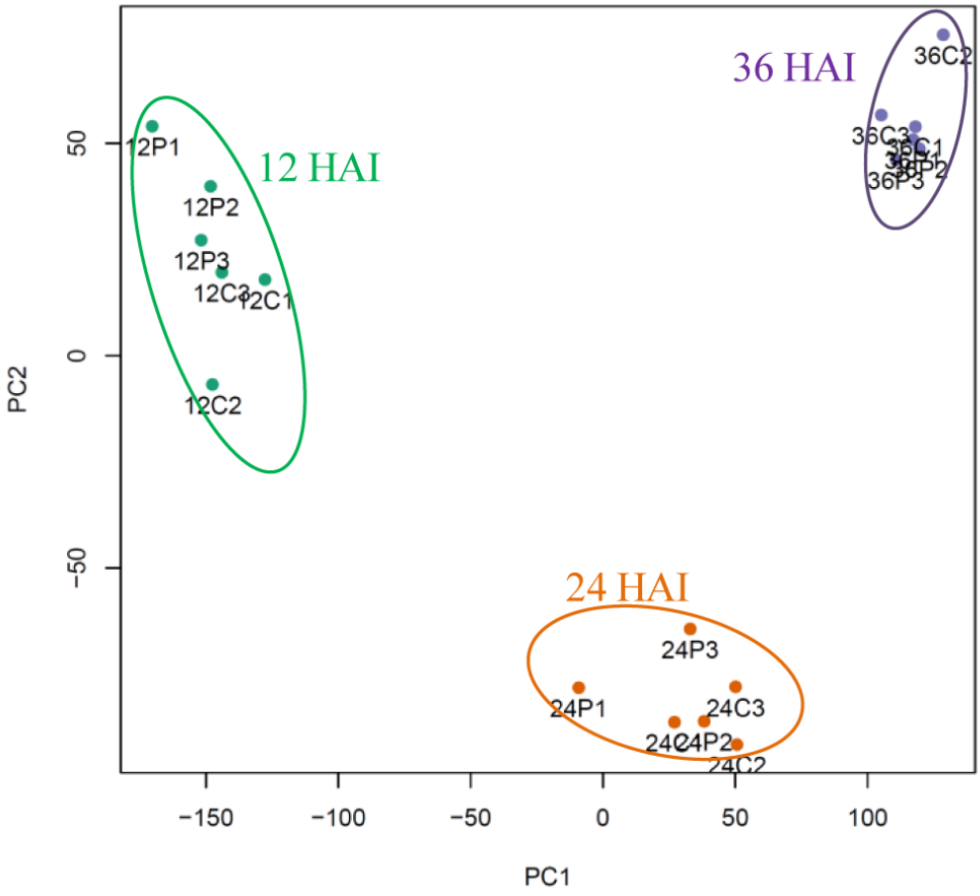


Figure 3.3 Principal Component Analysis (PCA) of expressed genes under control and PBZ at 12 HAI, 24 HAI and 36 HAI.
Three distinct groups can be observed: 12 HAI (green), 24 HAI (orange) and 36 HAI (purple).

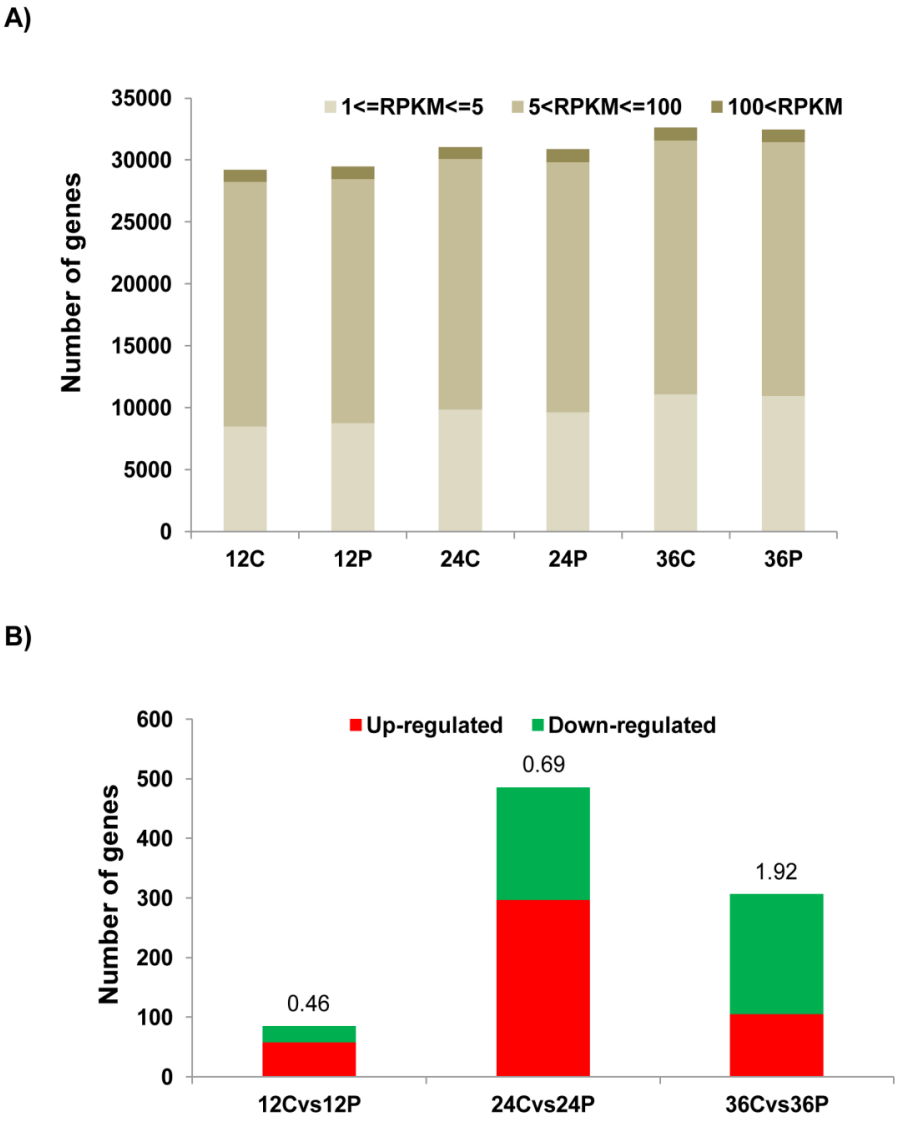


Figure 3.4 Gene expression profiling during seed germination.
A) Number of expressed genes (RPKM ≥ 1) and their estimated expression levels in each sample. **B)** Number of DEGs at 12, 24 and 36 HAI. Numbers above the vertical bars stand for the ratio between down- and up-regulated genes. In the x-axes labels, C and P stand for control and PBZ, respectively.

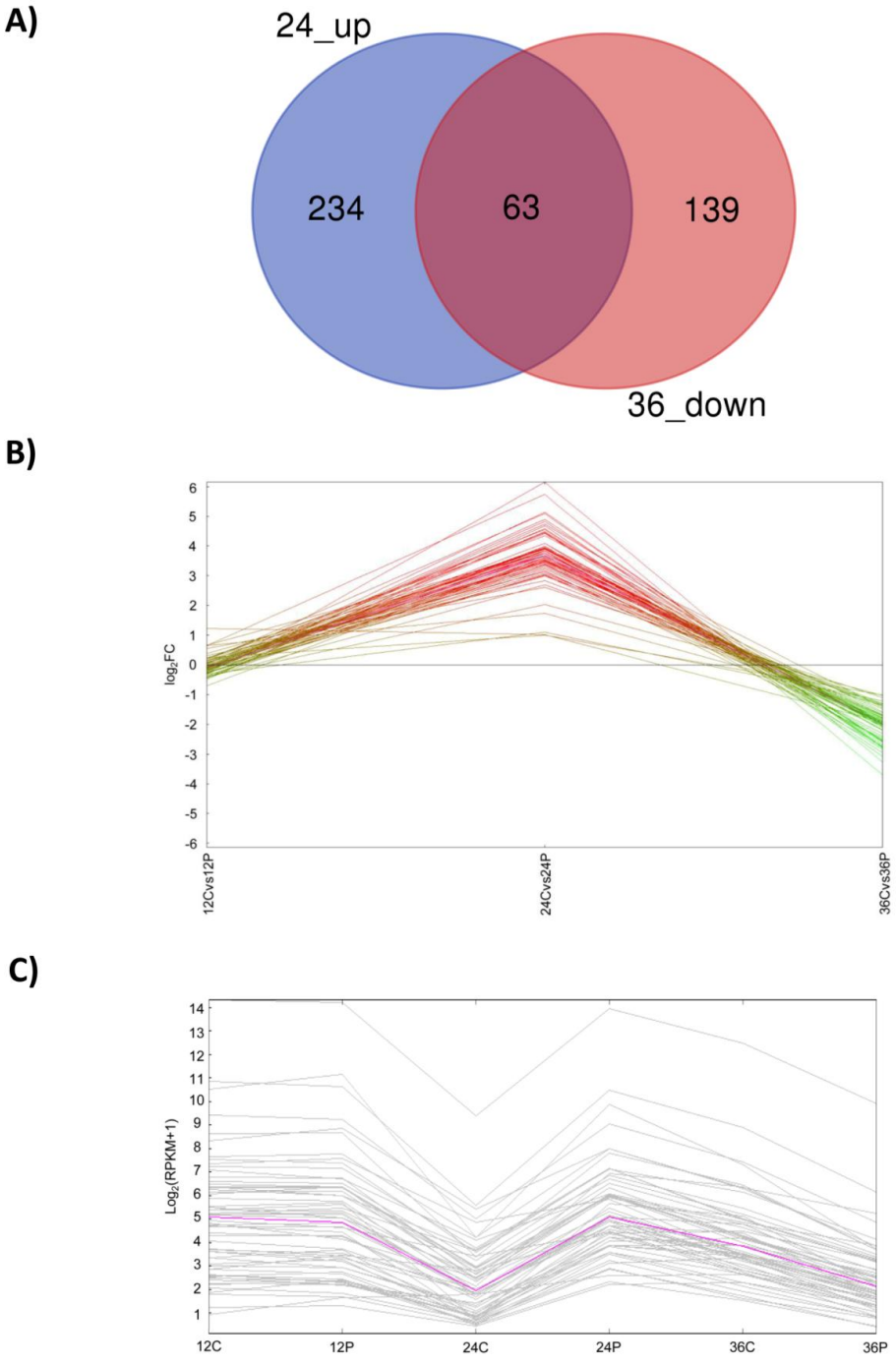


Figure 3.5 Overlap between 24-up- and 36-down-regulated genes.
(A) Venn diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/) shows 63 overlapping genes between 24 PBZ up- and 36 down-regulated genes. **(B)** Line plot of $\log_2$(fold change, FC) of the 63 genes at each time point. **(C)** Expression profile of the 63 genes in $\log_2$(RPKM+1) at each time point. MeV was used to generate graphs B and C. The pink line represents the median gene expression.

Table 3.2 Similarity of the protein products of nuclear genes that are up-regulated at 24 HAI and down-regulated at 36 HAI (APPENDIX B4) with organelle-encoded proteins.

BLASTP of the proteins encoded by the genes from APPENDIX B4 against chloroplast-encoded proteins.				
Gmax.protein.id (query)	**Gmax.chloroplast.protein.id (Reference Sequence: NC_007942)**	**Gene name**	**Sequence similarity (%)**	**Query coverage (%)**
Glyma.01G058600.1.p	lcl|NC_007942.1_prot_YP_538795.1_51	petB	97.39	99
Glyma.01G076000.1.p	lcl|NC_007942.1_prot_YP_538764.1_21	rpoC1	97.14	99
Glyma.01G201600.1.p	lcl|NC_007942.1_prot_YP_538754.1_11	ycf3	88.89	100
Glyma.03G031100.1.p	lcl|NC_007942.1_prot_YP_538766.1_23	rps2	98.41	89
Glyma.04G095000.1.p	lcl|NC_007942.1_prot_YP_538760.1_17	psbD	94.63	100
Glyma.05G073600.1.p	lcl|NC_007942.1_prot_YP_538745.1_2	psbA	78.16	79
Glyma.05G074900.1.p	lcl|NC_007942.1_prot_YP_538759.1_16	psbC	100	84
Glyma.06G217900.1.p	lcl|NC_007942.1_prot_YP_538760.1_17	psbD	100	89
Glyma.06G229000.1.p	lcl|NC_007942.1_prot_YP_538768.1_25	atpH	93.55	97
Glyma.08G281300.1.p	lcl|NC_007942.1_prot_YP_538791.1_47	psbB	94.62	87

Glyma.09G090100.1.p	lcl|NC_007942.1_prot_YP_538756.1_13	psaB	88.89	93
Glyma.09G171300.1.p	lcl|NC_007942.1_prot_YP_538762.1_19	petN	89.66	100
Glyma.10G078900.1.p	lcl|NC_007942.1_prot_YP_538827.1_83	rpl2	92.16	98
Glyma.11G114700.1.p	lcl|NC_007942.1_prot_YP_538759.1_16	psbC	99.06	79
Glyma.12G061600.1.p	lcl|NC_007942.1_prot_YP_538747.1_4	rbcL	90.14	99
Glyma.12G231900.1.p	lcl|NC_007942.1_prot_YP_538769.1_26	atpF	100	100
Glyma.12G232000.1.p	lcl|NC_007942.1_prot_YP_538770.1_27	atpA	100	100
Glyma.12G232100.1.p	lcl|NC_007942.1_prot_YP_538787.1_44	rpl20	100	100
Glyma.12G232300.1.p	lcl|NC_007942.1_prot_YP_538785.1_42	rpl33	100	100
Glyma.12G232700.1.p	lcl|NC_007942.1_prot_YP_538781.1_38	psbE	98.77	68
Glyma.13G068600.1.p	lcl|NC_007942.1_prot_YP_538750.1_7	ndhC	100	100
Glyma.14G213500.1.p	lcl|NC_007942.1_prot_YP_538788.1_45	rps12	100	83
Glyma.16G065300.1.p	lcl|NC_007942.1_prot_YP_538825.1_81	ycf2	97.11	80
Glyma.20G048700.1.p	lcl|NC_007942.1_prot_YP_53879	clpP	98.18	95

	0.1_46			
Glyma.U009000.1.p	lcl|NC_007942.1_prot_YP_538826.1_82	rpl23	100	100
Glyma.U016600.1.p	lcl|NC_007942.1_prot_YP_538755.1_12	psaA	96.69	98
BLASTP of the proteins encoded by the genes from APPENDIX B4 against mitochondria-encoded proteins.				
Gmax.protein.id (query)	**Gmax.mitochondrial.protein.id (Genbank accession number: JX463295)**	**Gene name**	**Sequence similarity (%)**	**Query coverage (%)**
Glyma.04G145900.1.p	lcl|JX463295.1_prot_AFR34329.1_82	atp8	91.09	79
Glyma.05G091900.1.p	lcl|JX463295.1_prot_AFR34300.1_7	rp15	71.35	76
Glyma.05G092000.1.p	lcl|JX463295.1_prot_AFR34302.1_9	cob	91.29	99
Glyma.05G092200.1.p	lcl|JX463295.1_prot_AFR34378.1_79	orf172-3	96.4	95
Glyma.05G092300.1.p	lcl|JX463295.1_prot_AFR34327.1_78	atp1-3	92.59	95
Glyma.07G122200.1.p	lcl|JX463295.1_prot_AFR34302.1_9	cob	97.08	99
Glyma.08G300900.1.p	lcl|JX463295.1_prot_AFR34348.1_33	orf189	100	71
Glyma.08G301000.1.p	lcl|JX463295.1_prot_AFR34314.1_32	nad6	100	99

Glyma.10G098500.1.p	lcl|JX463295.1_prot_AFR34311.1_28	nad4	92.86	64
Glyma.11G235100.1.p	lcl|JX463295.1_prot_AFR34303.1_15	nad3	92.59	60
Glyma.12G155300.1.p	lcl|JX463295.1_prot_AFR34298.1_5	cox3	100	99
Glyma.14G103300.1.p	lcl|JX463295.1_prot_AFR34316.1_37	nad7	84.31	100
Glyma.16G133700.1.p	lcl|JX463295.1_prot_AFR34328.1_80	atp6-2	95.89	99
Glyma.17G185500.1.p	lcl|JX463295.1_prot_AFR34307.1_18	rps10	95.59	76
Glyma.17G185600.1.p	lcl|JX463295.1_prot_AFR34305.1_17	atp4	96.92	99
Glyma.17G185800.1.p	lcl|JX463295.1_prot_AFR34299.1_53	nad5	83.1	55
Glyma.17G186100.1.p	lcl|JX463295.1_prot_AFR34331.1_45	nad4L-2	100	66

3.3.2 Gene Ontology and KEGG pathway enrichment analysis

Aiming to unravel major trends in the DEG lists, we conducted Gene Ontology (GO) and KEGG pathway enrichment analyses. There was no enrichment of GO terms or KEGG pathways at 12 HAI. In up-regulated genes at 24 HAI, we found a total of 19 enriched GO terms, including terms related with photosynthesis and translation (Table 3.3). Three of the GO terms enriched in the genes up-regulated at 24 HAI were also found enriched in the genes down-regulated at 36 HAI, namely “generation of precursor metabolites and energy”, “photosynthesis”

and “thylakoid” (Table 3.3), providing further support to the results discussed in the previous section.

KEGG pathway enrichment analysis revealed that ‘plant hormone signal transduction’ was enriched in down-regulated genes at 24 HAI and 36 HAI, supporting the regulation of other hormonal pathways by GA, and possibly their cross-talk, during germination (Table 3.4). These genes are involved in BR, auxin, jasmonic acid, ABA and cytokinin signaling or biosynthesis. Given their indispensable roles in regulating seed germination, genes related with hormone signaling and biosynthesis are discussed in more detail in the next section. In down-regulated genes at 36 HAI, a number of genes encoding chaperones resulted in the enrichment of the pathway ‘protein processing in endoplasmic reticulum’. Phenylpropanoid biosynthesis genes were enriched in PBZ-down (7 genes) and PBZ-up genes (6 genes) at 24 HAI and 36 HAI, respectively. These genes include β-glucosidases, peroxidases and spermidine hydroxycinnamoyl transferases that might be involved in cell wall modification or oxidative stress response (Table 3.4). ‘Biosynthesis of secondary metabolites’ genes were enriched in PBZ-down (15 genes) at 24 HAI and, both in PBZ-up (23 genes) and PBZ-down genes (15 genes) at 36 HAI. Most of these PBZ-up genes encode UDP-glycosyltransferases, cytochrome P450 proteins and brassinosteroid-6-oxidases, whereas PBZ-down genes encode 3-ketoacyl-CoA synthases, 1-amino-cyclopropane-1-carboxylate synthases (ACS) and peroxidases (Table 3.4). ‘Glutathione metabolism’, ‘RNA polymerase’, ‘purine metabolism’, ‘nucleotide excision repair’, ‘pyrimidine metabolism’ and spliceosome pathways were only enriched in up-regulated genes at 24 HAI (Table 3.4). All ’glutathione metabolism’ DEGs encode glutathione-S-transferases (GSTs) and their up-regulation is related to an increased antioxidant capacity (Roxas et al. 1997). Increased antioxidant capacity and DNA repair mechanisms at 24 HAI in response to PBZ might be part of a tolerance mechanism to cope with the germination delay, which is in line with a recent study that proposed a link

between DNA repair and antioxidant activity in *Medicago truncatula* seed germination and seedling establishment (Pagano et al. 2017).

Table 3.3 Enrichment analysis of Gene Ontology (GO) terms among differentially expressed genes.

			Up-regulated (24 HAI)		Down-regulated (36 HAI)	
GO term	**Level**	**Description**	**FDR**	**Num**	**FDR**	**Num**
GO:0006412	P	translation	1.00E-10	30		
GO:0006091	P	generation of precursor metabolites and energy	1.60E-08	14	1.8E-07	12
GO:0015979	P	photosynthesis	1.60E-08	13	2.30E-06	10
GO:0044249	P	cellular biosynthetic process	5.50E-05	51		
GO:0010467	P	gene expression	5.50E-05	46		
GO:0009058	P	biosynthetic process	1.20E-04	52		
GO:0006807	P	nitrogen compound metabolic process	1.30E-04	56		
GO:0044237	P	cellular metabolic process	7.80E-03	77		
GO:0008152	P	metabolic process	3.20E-02	101		
GO:0009987	P	cellular process	5.00E-02	85		
GO:0005840	C	ribosome	2.10E-13	29		
GO:0032991	C	macromolecular complex	3.60E-12	50		
GO:0009579	C	thylakoid	6.00E-09	11	1.50E-04	7
GO:0044444	C	cytoplasmic part	2.00E-07	30		

GO:0044424	C	intracellular part	1.70E-06	54		
GO:0005622	C	intracellular	2.30E-06	55		
GO:0005623	C	Cell	5.90E-06	55		
GO:0043226	C	organelle	2.00E-04	39		
GO:0005198	F	structural molecule activity	8.60E-13	29		

Table 3.4 Enrichment of KEGG pathways among differentially expressed genes.

	24 HAI				**36 HAI**			
	Up-regulated		**Down-regulated**		**Up-regulated**		**Down-regulated**	
Pathway Name	**Num**	**P-value**	**Num**	**P-value**	**Num**	**P-value**	**Num**	**P-value**
Metabolic pathways	55	1.18E-14			14	0.025	44	6.8E-13
Biosynthesis of secondary metabolites			15	0.042	15	0.016	22	3.09E-10
Plant hormone signal transduction			16	2.95E-08			9	0.002
Phenylpropanoid biosynthesis			7	0.002	6	0.0003		
Photosynthesis	20	4.56E-23					16	1.05E-18
Ribosome	34	1.38E-24					10	0.0003

Oxidative phosphorylation	24	4.56E-23					15	2.37E-13
Brassinosteroid biosynthesis	6	1.61E-08			8	8.85E-15		
Linoleic acid metabolism			4	0.0005				
Spliceosome	5	0.049						
Glutathione metabolism	4	0.02						
Nucleotide excision repair	4	0.01						
Pyrimidine metabolism	6	0.002						
RNA polymerase	5	2.97E-05						
Purine metabolism	6	0.009						

3.3.3 Feedback regulation and cross-talk with other hormones

GA biosynthesis can be divided into early (*CPS*, *KS*, *KO* and *KAO*) and late (e.g. *GA20ox* and *GA3ox*) stages (Olszewski et al. 2002). While early GA biosynthesis genes are generally not affected by GA (Helliwell et al. 1998), a negative GA-mediated feedback mechanism involving the down-regulation of late GA biosynthesis genes and up-regulation of the GA-deactivating *GA2ox* has been proposed as a system to keep balanced GA levels (Olszewski et al. 2002). Although not included by our statistical thresholds, we found *GA3ox* and *GA20ox* genes with greater expression in the presence of PBZ at 24 HAI and 36 HAI (Figure 3.6**,** APPENDIX B5), which might indicate a compensating mechanism in

response to PBZ. Long- and short-distance GA movement are also critical for developmental processes such as seed germination (Binenbaum et al. 2018). Recently, some transporters from the NPF and SWEET families are identified and shown to transport GA *in planta* (Kanno et al. 2016, Tal et al. 2016). Curiously, NPF3 transports GA and ABA in *A. thaliana* (Tal et al. 2016). We found two *NPF3* genes strongly up-regulated by PBZ at 36 HAI, which is in accordance with the GA-mediated repression of *NPF3* expression (Tal et al. 2016). The spatiotemporal expression pattern of NPF3 has been proposed as a key aspect of its functionality (Tal et al. 2016). In line with this, recent elegant works in *A. thaliana* showed that GA gradients correlate with cell length in dark-grown hypocotyls (Rizza and Jones 2018, Rizza et al. 2017). We hypothesize that this might be the case in soybean embryonic axes, particularly in the context of the recently described radicle-derived growth pattern in germinating soybean embryos (Souza et al. 2017).

In addition to biosynthesis and transport, we have also investigated GA signaling genes. We found 11 DELLA genes (one PBZ-down at 24 HAI) and all 5 GID1s (Gazara et al. 2018) expressed in at least one time point (Figure 3.6**,** APPENDIX B5). Almost all DELLAs showed greater expression in the absence of PBZ (Figure 3.6**,** APPENDIX B5). The expression levels of GID1b1, GID1b2 and GID1b3 were greater in PBZ than in controls (except GID1b1 and GID1b3 at 24 HAI), supporting that GID1b is particularly important under low GA concentrations, as previously hypothesized by us and others (Gazara et al. 2018, Tanimoto and Hirano 2013). Collectively, our results support that the low GA production resulting from PBZ administration activates an intricate system involving GA biosynthesis, signaling and transport genes, probably to minimize the effects of impaired GA production to allow germination to occur.

3.3.4 Other phytohormones

ABA is the most notorious GA antagonist for its inhibitory effect on seed germination (Kucera et al. 2005, Olszewski et al. 2002). The regulatory step in

ABA biosynthesis is catalyzed by 9-cis-epoxycarotenoid dioxygenase (NCED), which is transcriptionally regulated by positive and negative feedback loops in different species (Espasandin et al. 2014, Liu et al. 2016). The ABA receptor (PYL) inhibits the protein phosphatase 2C (PP2C) in the presence of ABA (Umezawa et al. 2009). We found one *NCED3* (*Glyma.08G176300*) and two *PP2Cs* as PBZ-down and one *PYL5* as PBZ-up (Figures 3.6 and 3.7, APPENDIX B5). In addition, two ABA transporters, *ABCG40* (up-regulated, *Glyma.19G169400*) and *NRT1.2* (down-regulated, *Glyma.08G296000*) were also differentially expressed upon PBZ treatment (APPENDIX B5). Collectively, these results show that GA modulate different genes involved in ABA biosynthesis, signaling and transport, which might directly interfere with a gradient of GA:ABA ratios along germinating soybean embryonic axes. This GA:ABA dynamics might be involved in the differential cell expansion patterns observed in germinating soybean embryos (Souza et al. 2017).

GA and ethylene positively interact with each other, promoting seed germination in several species (Corbineau et al. 2014). Multiple lines of evidence, including PBZ administration, support the positive regulation of ethylene biosynthesis and signaling by GA (Calvo et al. 2004, Calvo et al. 2004, Hua et al. 1998, Lehman et al. 1996, Ogawa et al. 2003). Further, several ethylene biosynthesis genes are expressed in soybean embryonic axes during germination (Bellieny-Rabelo et al. 2016). Accordingly, we found three PBZ-down 1-amino-cyclopropane-1-carboxylate synthase (ACS) genes (Figure 3.6**,** APPENDIX B5). ACS catalyzes the first committed and rate-limiting step in ethylene biosynthesis (Yang and Hoffman 1984). Our results suggest that up-regulation of ACS by GA is likely a key part of the synergy between GA and ethylene during soybean germination.

Several studies have shown that auxin inhibits or delays seed germination in wheat (Ramaih et al. 2003), *Arabidopsis* (Park et al. 2011) and soybean (Shuai et al. 2017). On the other hand, exogenous GA_4 up-regulated auxin biosynthesis

and carrier genes in germinating *Arabidopsis* seeds (Ogawa et al. 2003), supporting a complex GA-auxin cross-talk during soybean germination. There are multiple tryptophan-dependent IAA biosynthesis pathways in plants (Mashiguchi et al. 2011). The tryptophan aminotransferases *TAR1* and *TAR2* convert trp to indole-3-pyruvate (IPA), which is converted to indole acetic acid (IAA) by the YUCCA flavin monooxygenase (Zhao 2014). Further, *superroot2* (*SUR2*) encodes the cytochrome P450 monooxygenase CYP83B1, involved in glucosinolate biosynthesis and auxin homeostasis (Bak et al. 2001, Barlier et al. 2000). We found two PBZ-up *SUR2* at 12 HAI and one PBZ-down *TAR2* at 24 HAI, indicating that GA promotes IAA production at these time points. We also found one auxin transporter (*PIN*; PBZ-up) and eleven auxin-responsive genes, including seven PBZ-down Auxin/Indole-3-Acetic Acid (Aux/IAA) repressors, small auxin upregulated RNA (SAUR), and the auxin-responsive Gretchen Hagen3 (GH3) family were differentially expressed at least at one of the time-point (Figures 3.6 and 3.7, APPENDIX B5). Although apparently conflicting with the promotion of IAA biosynthesis at 12 and 24 HAI, the down-regulation of several *AUX/IAA* genes by PBZ at 24 HAI and 36 HAI suggests that GA represses auxin signaling during late germination. Accordingly, three AUX/IAA genes have been recently demonstrated to promote hypocotyl elongation in *A. thaliana* (Reed et al. 2018).

BRs typically induce seed germination and BR biosynthesis genes (*DET2*, *DWF4*, *DWF3*, *BR6ox1*, and *ROT3*) are up-regulated when endogenous BR concentrations are reduced (Tanaka et al. 2005). Interestingly, six and eight BR biosynthesis genes were PBZ-up at 24 and 36 HAI, respectively (Figure 3.6**,** APPENDIX B5). BR promotes GA biosynthesis by regulating *GA20ox1* and *GA3ox1* expression in *A. thaliana* (Unterholzner et al. 2015). Further, GA partially rescued hypocotyl elongation defects resulting from BR deficiency (Unterholzner et al. 2015). Our group has proposed that BR signaling regulates cell expansion during soybean germination (Bellieny-Rabelo et al. 2016). Taken together, the up-regulation of BR biosynthesis upon PBZ treatment might be involved in the

activation of late GA biosynthesis genes to counter PBZ effects on GA production. This hypothesis also fits the observation that PBZ delays germination without a clear effect on germination rates (Figure 3.2E). Finally, since BR also promotes GA biosynthesis in rice (Tong et al. 2014), the emergence of this regulatory module probably predates the diversification of monocotyledonous and dicotyledonous species.

Antagonistic interactions between GA and cytokinin (CK) have been reported in different plants (Fleishon et al. 2011, Fonouni-Farde et al. 2017, Greenboim-Wainberg 2005). Type-A response regulators negatively regulate CK signaling by competing with type-B response regulators for phosphoryl transfer from the upstream *Arabidopsis* Hpt proteins or by interacting with other pathway components (To et al. 2007). We found four and three PBZ-down type-A response regulators at 24 HAI and 36 HAI, respectively (Figure 3.7**,** APPENDIX B5). Since CK biosynthesis genes were not differentially expressed, our results indicate GA antagonizes CK by the up-regulation of negative CK signaling regulators during soybean germination.

In the canonical Jasmonic Acid (JA) signaling pathway, the receptor CORONATINE INSENSITIVE 1 (COI1) interacts with JA and promotes the proteasomal degradation of JASMONATE ZIM-domain (JAZ) repressors (Wasternack and Hause 2013). JAZ represses the transcription of JA-responsive genes through interaction with the MYC2 TF and other regulatory proteins (Song et al. 2014, Wasternack and Hause 2013). JA and GA perform antagonistic roles in regulating hypocotyl elongation via physical interactions between JAZ and DELLA repressors. In summary, JA-mediated JAZ degradation releases DELLA to repress GA signaling (and hypocotyl elongation), whereas GA-mediated DELLA degradation releases JAZ to inhibit JA responses (Song et al. 2014, Yang et al. 2012). We found two PBZ-down *JAZ* genes at 24 HAI (APPENDIX B5), indicating that GA represses JA signaling during germination. Interestingly, *JAZ* up-regulation might constitute an additional layer of JA repression, as GA-

promoted DELLA degradation would release JAZ proteins to repress JA signaling, as discussed above. Figure 3.8 is a schematic diagram of hormone cross-talk during seed germination in *G. max*.

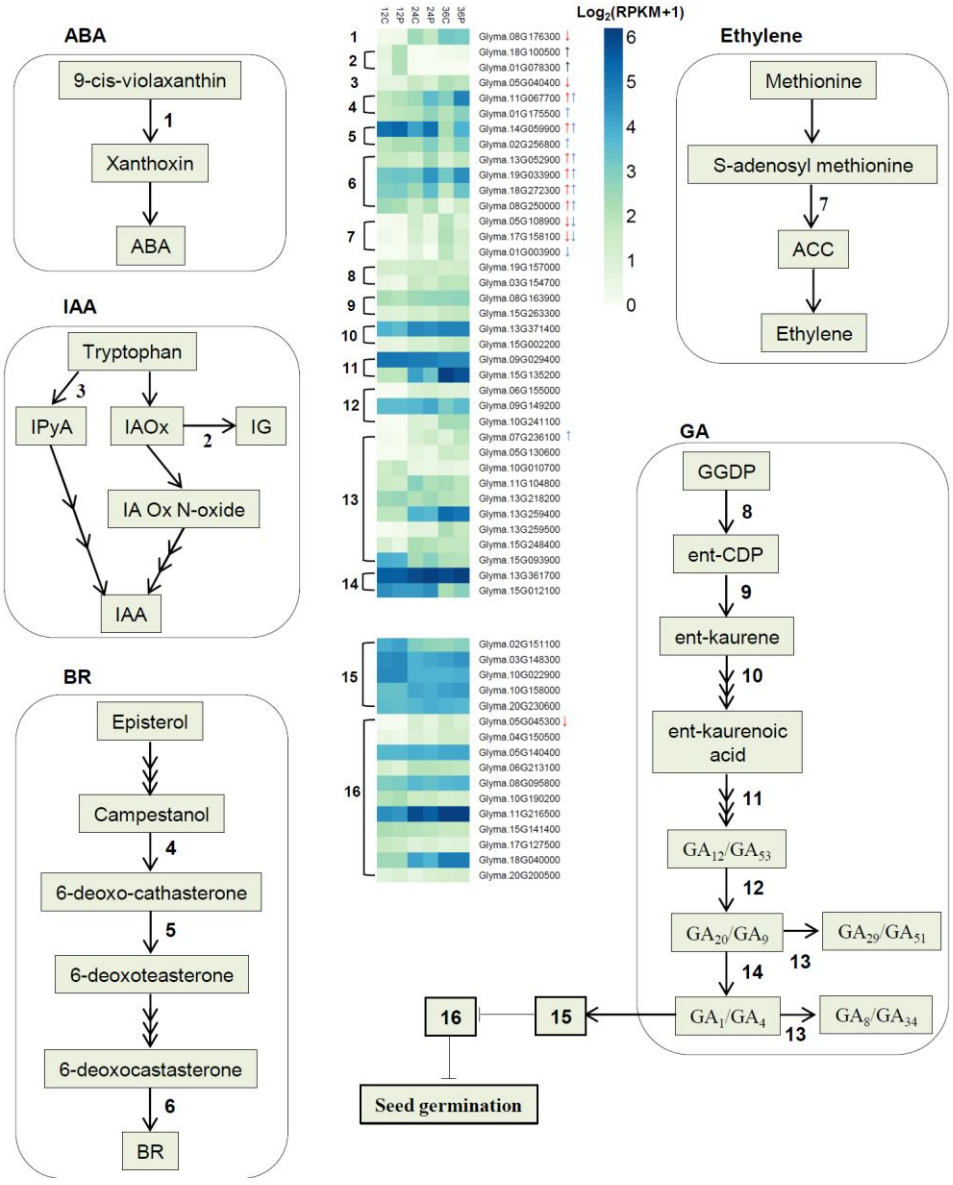


Figure 3.6 Hormone biosynthesis pathways.
Some GA deactivation and signaling genes discussed are also included. Up- and down-regulated genes are shown with up and down arrows. Black, red and blue arrows represent differential expression at 12, 24 and 36 HAI, respectively. Genes without arrows are expressed in at least one condition, although not included by our statistical thresholds. Genes are numbered as follows: 1) nine-cis-epoxycarotenoid dioxygenase 3 (*NCED3*); 2) *SUR2*; 3) tryptophan aminotransferase related 2 (*TAR2*); 4) DWARF4 (*DWF4*); 5) DWARF3 (*DWF3*); 6) brassinosteroid-6-oxidase 2 (*BR6ox2*); 7) 1-amino-cyclopropane-1-carboxylate

synthase (*ACS*); 8) ent-copalyl diphosphate synthase (*CPS*); 9) ent-kaurene synthase (*KS*); 10) ent-kaurene oxidase (*KO*); 11) ent-kaurenoic acid oxidase (*KAO*); 12) GA 20-oxidase (*GA20ox*); 13) GA 2-oxidase (*GA2ox*) ; 14) GA 3-oxidase (*GA3ox*); 15) GIBBERELLIN INSENSITIVE DWARF1 (*GID1*) [*Glyma.02G151100* (GID1b1), *Glyma.10G022900* (GID1b2), *Glyma.03G148300* (GID1b3), *Glyma.10G158000* (GID1c1) and *Glyma.20G230600* (GID1c2); 16) DELLA. Abbreviations: Abscisic Acid (ABA), indole-3-pyruvic acid (IPyA), Indol-3-acetaldoxime (IAOx), Indol-3-acetaldoxime N-oxide (IA Ox N-oxide), indole glucosinolates (IG), Indole-3-acetic acid (IAA), Brassinosteroid (BR), 1-aminocyclopropane-1-carboxylic acid (ACC), geranyl geranyl diphosphate (GGDP), ent-copalyl diphosphate (ent-CDP).

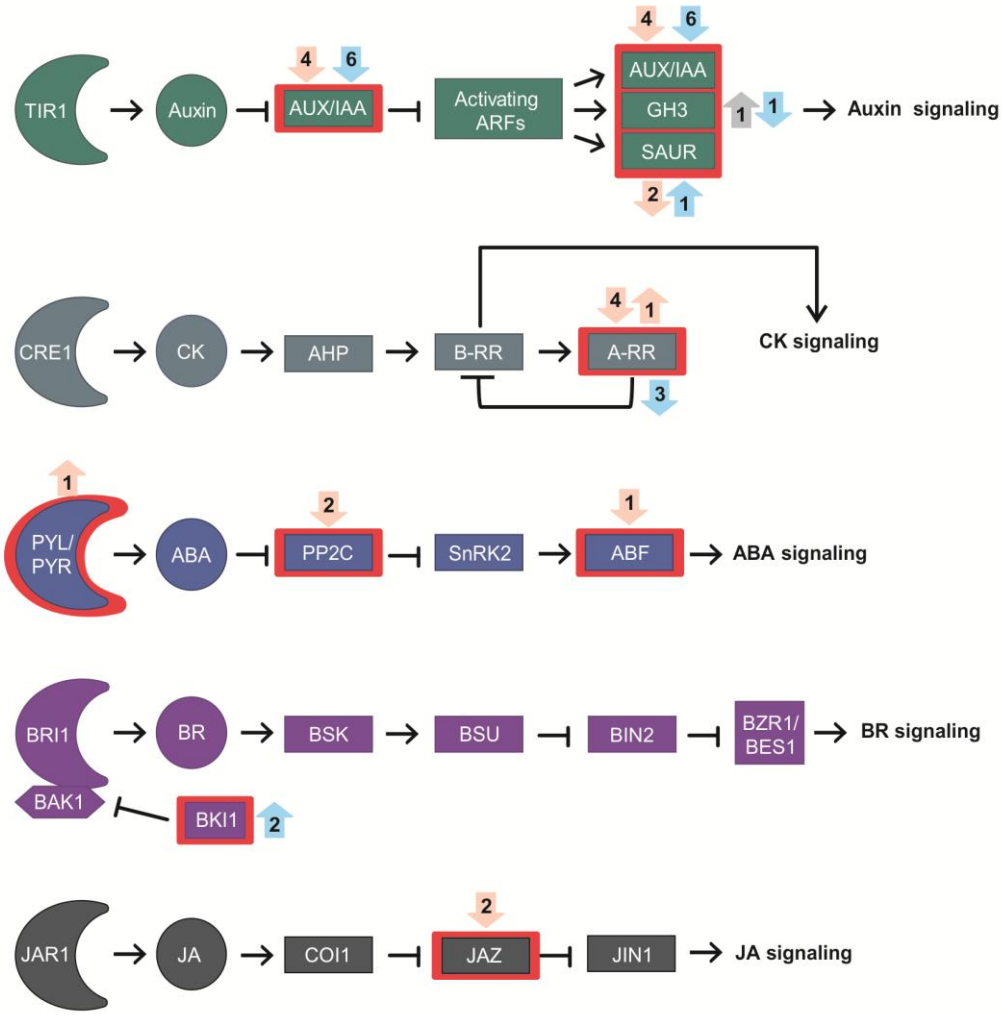


Figure 3.7 Hormone signal transduction.
Rectangles with red lines represent gene families with at least one DEG. Up and down arrows represent PBZ up- and down-regulated genes. Number of DEGs are shown in circles adjacent to the red rectangles. Grey, light orange and light blue arrows represent DEGs at 12, 24 and 36 HAI, respectively**.** Abbreviations: transport inhibitor response 1 (TIR1); Auxin/Indole-3-Acetic Acid (Aux/IAA); auxin-responsive Gretchen Hagen3 (GH3); small auxin upregulated RNA (SAUR); CYTOKININ RESPONSE 1 (CRE 1); Cytokinin (CK); His-

containing phosphotransfer protein (AHP) ;Type-B response regulator (B-RR); Type-A response regulator (A-RR); Pyrabactin Resistance (PYR); PYR-like (PYL); Abscisic acid (ABA); Protein Phosphatase 2C (PP2C); Sucrose non-fermenting 1-related protein kinases subfamily 2 (SnRK2s); Abscisic acid responsive element-binding factor (ABF); Brassinosteroid-insensitive 1 (BRI1); BRI1-associated receptor kinase 1 (BAK1); Brassinosteroid (BR); BRI1 kinase inhibitor (BKI1); Brassinosteroid signaling kinases (BSK); BRI1-suppressor (BSU); brassinosteroid-insensitive 2 (BIN2); Brassinazole-resistant 1 (BZR1); BRI1-ethyl methanesulfonate-suppressor 1 (BES1); JASMONATE RESISTANT1 (JAR1); Jasmonic acid (JA); Coronatine Insensitive1 (COI1); JASMONATE ZIM DOMAIN (JAZ); JASMONATE INSENSITIVE 1 (JIN1).

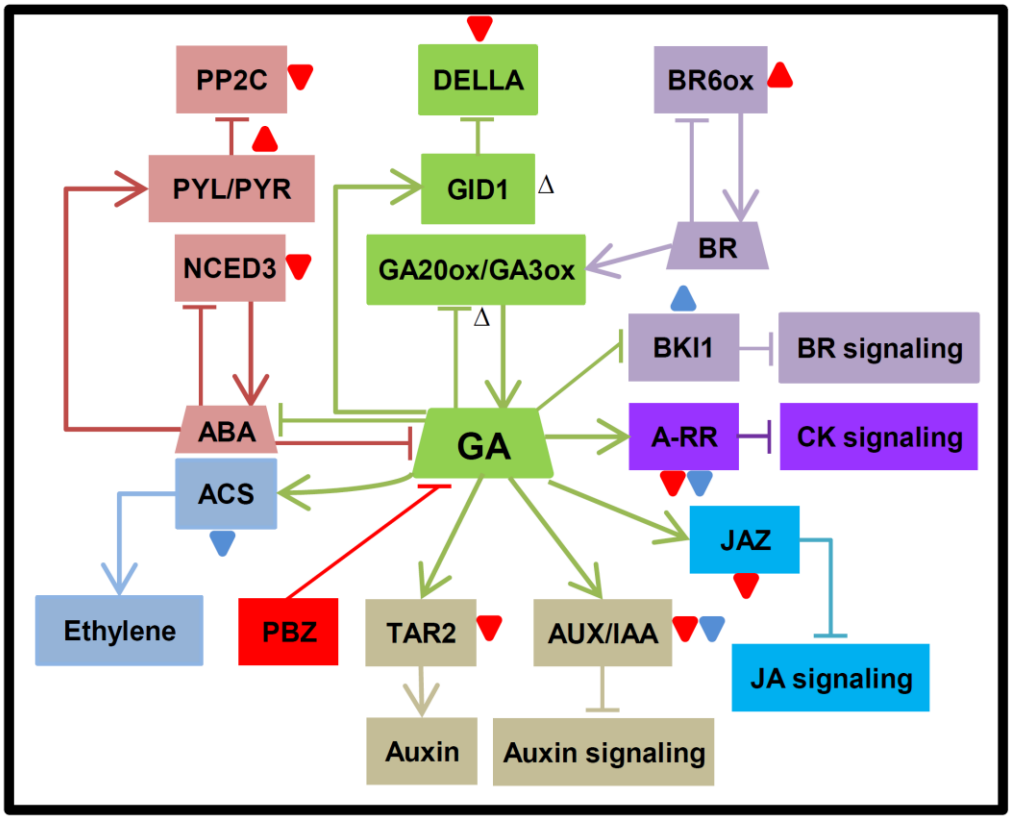


Figure 3.8 Schematic model of hormonal crosstalk with gibberellin during *G. max* seed germination.

The model was derived from a careful literature curation based on differentially expressed genes discussed along the manuscript. Positive interactions are indicated by arrows and T bars indicate repression. Red and blue fill triangles represent DEGs at 24 and 36 HAI, respectively. Up- and down-regulated genes are shown with up and down fill triangles. Genes that are not differentially expressed but have greater expressions in PBZ condition shown with open triangles. Abbreviations: Pyrabactin Resistance (PYR); PYR-like (PYL); Protein Phosphatase 2C (PP2C); Nine-cis-epoxycarotenoid dioxygenase 3 (NCED3); Abscisic acid (ABA); Aminocyclopropane-1-carboxylic acid synthase (ACS); Paclobutrazol (PBZ); Gibberellin (GA); GIBBERELLIN INSENSITIVE DRAWF 1 (GID1), particularly representing GID1bs of *G. max*; GA 20-oxidase (GA20ox); GA 3-oxidase (GA3ox); Tryptophan aminotransferases 2 (TAR2); Auxin/Indole-3-Acetic Acid (AUX/IAA);

Brassinosteroid (BR); BR 6-oxidase (BR6ox); BRI1 kinase inhibitor (BKI1); Type-A response regulator (A-RR); JASMONATE ZIM DOMAIN (JAZ); Jasmonic acid (JA).

3.3.5 Gibberellins regulate cell wall remodeling enzymes

Several genes encoding cell elongation and cell wall remodeling enzymes such as xyloglucan endotransglycosylase/hydrolases (XTH), pectin methylesterases (PME), expansins, pectin lyases, aquaporin and others are induced by GA in *Arabidopsis* and tomato seed germination (Cao et al. 2006, Chen 2002, Chen and Bradford 2000, Ogawa et al. 2003, Park et al. 2017). We found a number of these cell wall remodeling genes as differentially expressed (Figure 3.9A). Peroxidases and glycosyl hydrolases (GHs) also play active role in cell wall loosening (Liszkay et al. 2003, Minic and Jouanin 2006). Accordingly, nine and eight peroxidases and GHs were differentially expressed, respectively. Genes involved in pectin metabolism were also modulated by PBZ (Figure 3.9A), suggesting that this process is also under GA regulation during germination. We also found other cell wall related DEGs, such as arabinogalactan-proteins, fasciclin-like AGPs, hydroxyproline (Hyp)-rich glycoproteins, and proline- or glycine-rich proteins, which play important roles in cell proliferation (Langan and Nothnagel 1997, Ringli et al. 2001, Serpe and Nothnagel 1994) and expansion (Willats and Knox 1996). Several of those genes are also GA-responsive in cucumber, maize and barley (Liu et al. 2013, Park et al. 2003, Suzuki et al. 2002). Importantly, 30 out of 44 cell wall DEGs were PBZ-down, supporting that the notorious effect of GA in promoting cell elongation.

3.3.6 Transcription factor genes modulated by paclobutrazol are likely drivers of GA-mediated transcriptional reprogramming

Because seed germination is mainly regulated by the embryonic axis, we have specifically investigated the differential expression of TFs in this tissue, as they might be major drivers of the GA transcriptional programs. A total of 45 TFs were differentially expressed upon PBZ treatment. Strikingly, one, 18 and 23 TFs were differentially expressed exclusively at 12, 24 and 36 HAI, respectively (Figure 3.9B). This pattern indicates that differentially expressed TFs play specific roles

at different germination times. Further, most of the differentially expressed TFs (66.7%) were down-regulated by PBZ and likely comprise regulators that are downstream of GA (Figure 3.9B). The TF families with the greatest number of down-regulated members were MyB (myeloblastosis; 10 down), bHLH (basic helix-loop-helix; 8 down) and bZIP (basic leucine zipper domain; 3 down), which is in line with previous studies in soybean (Bellieny-Rabelo et al. 2016) and *A. thaliana* (Cao et al. 2006), which showed that MyB and bHLH are among the mostly activated TF families during germination. Interestingly, five and six of the PBZ-down MYB and bHLH genes, respectively, were also differentially expressed in a time-dependent manner during soybean germination (Bellieny-Rabelo et al. 2016), further supporting that GA coordinate the transcription of specific TFs at different HAI. Conversely, S1Fa-like (4 up) and WRKY (3 up) were the families that were most represented among PBZ-up TFs (Figure 3.9B). S1Fa-like is a poorly-studied TF family that has been associated with photomorphogenesis (Zhou et al. 1995). Remarkably, all four soybean S1Fa-like TFs were strongly up-regulated by PBZ at 24 HAI, indicating that they might be part of the regulatory system to activate photosynthetic growth in response to low GA concentrations, as discussed above. Photomorphogenesis is regulated by a complex pathway involving GA and light in *A. thaliana* seedlings (Alabadı´ et al. 2004, Alabadí et al. 2007). Nevertheless, no PIF or HY5 genes, which encode important regulators of photomorphogenesis, were modulated by PBZ.

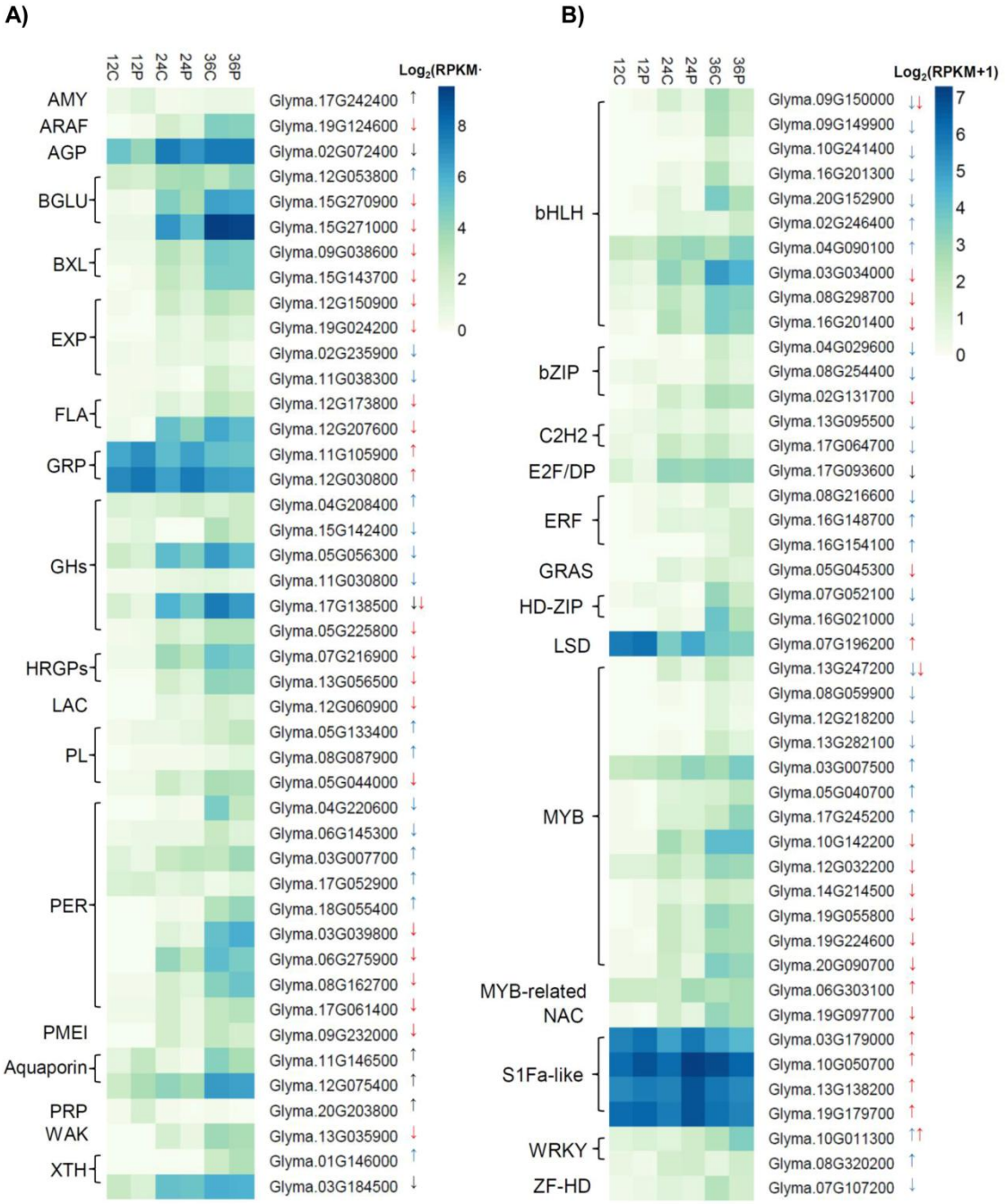


Figure 3.9 Genes encoding differentially expressed cell-wall remodeling enzymes (A) and transcription factors (B).

Up- and down-regulated genes are shown with up and down arrows. Black, red and blue arrows represent DEGs at 12, 24 and 36 HAI, respectively. Abbreviations: alpha amylase-like (AMY); alpha-L-arabinofuranosidase (ARAF); arabinogalactan protein (AGP); beta glucosidase (BGLU); beta-xylosidase (BXL); expansin (EXP); FASCICLIN-like arabinogalactan-protein (FLA); glycine-rich protein (GRP); Glycosyl hydrolase family protein (GH); hydroxyproline-rich glycoprotein family protein (HRGP); laccase (LAC); Pectin lyase-like superfamily protein (PL); Peroxidase superfamily protein (PER); pectin methylesterase inhibitor superfamily protein (PMEI); proline-rich protein (PRP); wall associated kinase (WAK); xyloglucan endotransglucosylase/hydrolase (XTH); basic helix-loop-helix (bHLH);

Basic Leucine Zipper (bZIP) ; C2H2 zinc finger (C2H2); Ethylene response factor (ERF); GRAS (gibberellin insensitive (GAI), Repressor of ga1-3 (RGA), SCARECROW-LIKE 3 (SCR) gene family; Homeodomain-leucine zipper (HD-ZIP); LESION SIMULATING DISEASE (LSD); Myelobastosis (MYB); Zinc finger Homeodomain (ZF-HD); No apical meristem (NAM), ATAF, and CUC (cup-shaped cotyledon) (NAC) family.

3.3.7 Comparison with *A. thaliana* GA-responsive genes

Ogawa *et al* identified a total of 230 and 127 up- and down-regulated genes during germination of *A. thaliana ga1-3* seeds upon GA treatment (Ogawa et al. 2003). Other study, also in *A. thaliana*, reported DEGs in imbibed seeds and developing flowers of wild type, *ga1-3*, and a quintuple DELLA null mutant (*ga1 rga gai rgl1 rgl2*) (Cao et al. 2006). This latter study identified 541 and 571 up- and down-regulated GA-responsive genes in imbibed seeds. It is important to mention that Ogawa et al. used a microarray platform representing ~8,200 genes, while Cao et al. used one covering ~23,000 genes. This difference is likely an important factor accounting for the differences in DEG numbers between these studies. Overall, these studies have an overlap of 109 GA-up genes and 90 GA-down genes. Importantly, a significant fraction of these genes are also regulated by DELLA (Cao et al. 2006).

Although *A. thaliana* and soybean are distantly related and their seeds are remarkably different, we investigated the conservation of the DEGs identified in *A. thaliana* described above with the ones reported here using BLASTP (minimum query coverage and similarity of 50%). We found 178 and 124 differentially expressed soybean orthologs for 122 and 84 *A. thaliana* GA-up and GA-down genes, respectively. These soybean gene sets were named GA-up-orthologs and GA-down-orthologs, respectively. Curiously, a significant part (47.19% and 55.66% of the GA-up-orthologs and GA-down-orthologs, respectively) of these genes is modulated in opposite directions in the two species (APPENDIX B6). Nevertheless, most of the genes related with cell-wall modification, GSTs, auxin responsive genes (AUX/IAA and SAUR), oxidoreductases (aldo-ketoreductases), and transferases are modulated in same

directions in soybean and *A. thaliana*, whereas genes modulated in opposite directions between the species encode HSPs, cytochrome p450, serine carboxypeptides, late embryogenesis proteins and flavonol synthase/flavanone 3-hydroxylase (APPENDIX B6). Proportionally and in absolute numbers, 24 HAI is the stage with the most conserved DEG profile between the two species. Further, 351 out of the 468 soybean DEGs without a DEG ortholog in *A. thaliana* do have orthologs in the *A. thaliana* genome, indicating that a several orthologous genes are differentially regulated in the two species. Finally, in addition to the evolutionary distance, there are also important technical aspects that require consideration. The *A. thaliana* studies used microarrays to investigate modulated genes in *ga1-3* mutants either upon treatment with exogenous GA (Ogawa et al. 2003) or in contrast with wild type seeds during germination (Cao et al. 2006). Here we analyzed an RNA-Seq transcriptome of embryonic axes of germinating soybean seeds treated with PBZ. Both experimental designs have limitations; even the *A. thaliana ga1-3* dry seeds have bioactive GA from the GA treatment used to rescue parental fertility of mutant plants (Ogawa et al. 2003). In addition, administration of exogenous GA may have unintended effects due to locations and concentrations different from those found under natural conditions. On the other hand, while allowing the investigation with more natural GA concentrations and locations, chemical inhibition of GA biosynthesis probably does not shutdown GA signaling completely. Further, it is not unreasonable to expect that the inhibitor effects might be overcome after some time, for example by an increase in the levels of GA biosynthesis enzymes. A more detailed picture of the interspecies conservation of GA-driven transcriptional programs will be clearer when more species are studied using state-of-the-art RNA-Seq technologies.

3.4 REFERENCES

Alabadı´ D, Gil J, Bla´zquez MA, García-Martínez JL. 2004. Gibberellins Repress Photomorphogenesis in Darkness. PLANT PHYSIOLOGY.134:1050-1057.

Alabadí D, Gallego-Bartolomé J, Orlando L, García-Cárcel L, Rubio V, Martínez C, Frigerio M, Iglesias-Pedraz JM, Espinosa A, Deng XW, et al. 2007. Gibberellins modulate light signaling pathways to prevent Arabidopsis seedling de-etiolation in darkness. The Plant Journal.53:324-335.

Bak S, Tax FE, Feldmann KA, Galbraith DW, Feyereisen R. 2001. CYP83B1, a cytochrome P450 at the metabolic branch point in auxin and indole glucosinolate biosynthesis in Arabidopsis. The Plant cell.13:101-111.

Barlier I, Kowalczyk M, Marchant A, Ljung K, Bhalerao R, Bennett M, Sandberg G, Bellini C. 2000. The SUR2 gene of Arabidopsis thaliana encodes the cytochrome P450 CYP83B1, a modulator of auxin homeostasis. Proc Natl Acad Sci U S A. Dec 19;97:14819-14824. Epub 2000/12/13.

Bellieny-Rabelo D, Alves Gamosa de Oliveira E, da Silva Ribeiro E, Pessoa Costa E, Elenir Amâncio Oliveira A, Motta Venancio T. 2016. Transcriptome analysis uncovers key regulatory and metabolic aspects of soybean embryonic axes during germination. Scientific Reports.6:36009.

Bewley JD. 1997. Seed Germination and Dormancy. THE PLANT CELL ONLINE.9:1055-1066.

Bewley JD, Bradford KJ, Hilhorst HWM, Nonogaki H. 2013. Seeds: Physiology of Development, Germination and Dormancy New York, NY: Springer New York.

Binenbaum J, Weinstain R, Shani E. 2018. Gibberellin Localization and Transport in Plants. Trends in Plant Science.23:410-421.

Boudreau E. 1997. The chloroplast ycf3 and ycf4 open reading frames of Chlamydomonas reinhardtii are required for the accumulation of the photosystem I complex. The EMBO Journal.16:6095-6104.

Calvo AP, Nicolas C, Lorenzo O, Nicolas G, Rodriguez D. 2004. Evidence for positive regulation by gibberellins and ethylene of ACC oxidase expression and activity during transition from dormancy to germination in Fagus sylvatica L. seeds. Journal of Plant Growth Regulation.23:44-53.

Calvo AP, Nicolas C, Nicolas G, Rodriguez D. 2004. Evidence of a cross-talk regulation of a GA 20-oxidase (FsGA20ox1) by gibberellins and ethylene during the breaking of dormancy in Fagus sylvatica seeds. Physiologia Plantarum.120:623-630.

Cao D, Cheng H, Wu W, Soo HM, Peng J. 2006. Gibberellin mobilizes distinct DELLA-dependent transcriptomes to regulate seed germination and floral development in Arabidopsis. PLANT PHYSIOLOGY.142:509-525.

Chen F. 2002. A gibberellin-regulated xyloglucan endotransglycosylase gene is expressed in the endosperm cap during tomato seed germination. Journal of Experimental Botany.53:215-223.

Chen F, Bradford KJ. 2000. Expression of an Expansin Is Associated with Endosperm Weakening during Tomato Seed Germination. PLANT PHYSIOLOGY.124:1265-1274.

Corbineau F, Xia Q, Bailly C, El-Maarouf-Bouteau H. 2014. Ethylene, a key factor in the regulation of seed dormancy. Front Plant Sci.5:539. Epub 2014/10/28.

Espasandin FD, Maiale SJ, Calzadilla P, Ruiz OA, Sansberro PA. 2014. Transcriptional regulation of 9-cis-epoxycarotenoid dioxygenase (NCED) gene by putrescine accumulation positively modulates ABA synthesis and drought tolerance in Lotus tenuis plants. Plant Physiology and Biochemistry.76:29-35.

Fleishon S, Shani E, Ori N, Weiss D. 2011. Negative reciprocal interactions between gibberellin and cytokinin in tomato. New Phytologist.190:609-617.

Fonouni-Farde C, Kisiala A, Brault M, Emery RJN, Diet A, Frugier F. 2017. DELLA1-Mediated Gibberellin Signaling Regulates Cytokinin-Dependent Symbiotic Nodulation. Plant Physiology.175:1795-1806.

Gazara RK, Moharana KC, Bellieny-Rabelo D, Venancio TM. 2018. Expansion and diversification of the gibberellin receptor GIBBERELLIN INSENSITIVE DWARF1 (GID1) family in land plants. Plant Molecular Biology.97:435-449.
Greenboim-Wainberg Y. 2005. Cross Talk between Gibberellin and Cytokinin: The Arabidopsis GA Response Inhibitor SPINDLY Plays a Positive Role in Cytokinin Signaling. THE PLANT CELL ONLINE.17:92-102.
Groot SPC, Bruinsma J, Karssen CM. 1987. The role of endogenous gibberellin in seed and fruit development of tomato: Studies with a gibberellin-deficient mutant. Physiologia Plantarum.71:184-190.
Hedden P, Graebe JE. 1985. Inhibition of gibberellin biosynthesis by paclobutrazol in cell-free homogenates of Cucurbita maxima endosperm and Malus pumila embryos. Journal of Plant Growth Regulation.10:111-122.
Helliwell CA, Sheldon CC, Olive MR, Walker AR, Zeevaart JA, Peacock WJ, Dennis ES. 1998. Cloning of the Arabidopsis ent-kaurene oxidase gene GA3. Proceedings of the National Academy of Sciences of the United States of America.95:9019-9024.
Holdsworth MJ, Bentsink L, Soppe WJJ. 2008. Molecular networks regulating Arabidopsis seed maturation, after-ripening, dormancy and germination. New Phytologist.179:33-54.
Hricova A. 2006. The SCABRA3 Nuclear Gene Encodes the Plastid RpoTp RNA Polymerase, Which Is Required for Chloroplast Biogenesis and Mesophyll Cell Proliferation in Arabidopsis. PLANT PHYSIOLOGY.141:942-956.
Hua J, Sakai H, Nourizadeh S, Chen QG, Bleecker AB, Ecker JR, Meyerowitz EM. 1998. EIN4 and ERS2 are members of the putative ethylene receptor gene family in Arabidopsis. The Plant cell.10:1321-1332.
Jacobsen SE, Olszewski NE. 1993. Mutations at the SPINDLY locus of Arabidopsis alter gibberellin signal transduction. The Plant cell.5:887-896.
Jin J, Tian F, Yang D-C, Meng Y-Q, Kong L, Luo J, Gao G. 2017. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Research.45:D1040-D1045.
Kanno Y, Oikawa T, Chiba Y, Ishimaru Y, Shimizu T, Sano N, Koshiba T, Kamiya Y, Ueda M, Seo M. 2016. AtSWEET13 and AtSWEET14 regulate gibberellin-mediated physiological processes. Nature Communications.7:13245.
Karssen CM, Zaorski S, Kepczynski J, Groot SPC. 1989. Key Role for Endogenous Gibberellins in the Control of Seed Germination. Ann Bot.63:71-80.
Koornneef M, van der Veen JH. 1980. Induction and analysis of gibberellin sensitive mutants in Arabidopsis thaliana (L.) heynh. Theoretical and Applied Genetics.58:257-263.
Kucera B, Cohn MA, Leubner-Metzger G. 2005. Plant hormone interactions during seed dormancy release and germination. Seed Science Research.15:281-307.
Langan KJ, Nothnagel EA. 1997. Cell surface arabinogalactan-proteins and their relation to cell proliferation and viability. Protoplasma.196:87-98.
Lehman A, Black R, Ecker JR. 1996. HOOKLESS1, an ethylene response gene, is required for differential cell elongation in the Arabidopsis hypocotyl. Cell.85:183-194.
Linkies A, Muller K, Morris K, Tureckova V, Wenk M, Cadman CSC, Corbineau F, Strnad M, Lynn JR, Finch-Savage WE, et al. 2009. Ethylene Interacts with Abscisic Acid to Regulate Endosperm Rupture during Germination: A Comparative Approach Using Lepidium sativum and Arabidopsis thaliana. The Plant Cell.21:3803-3822.
Liszkay A, Kenk B, Schopfer P. 2003. Evidence for the involvement of cell wall peroxidase in the generation of hydroxyl radicals mediating extension growth. Planta.217:658-667.
Liu S, Li M, Su L, Ge K, Li L, Li X, Liu X, Li L. 2016. Negative feedback regulation of ABA biosynthesis in peanut (Arachis hypogaea): a transcription factor complex inhibits AhNCED1 expression during water stress. Scientific Reports.6:37943.
Liu W-Y, Chang Y-M, Chen SC-C, Lu C-H, Wu Y-H, Lu M-YJ, Chen D-R, Shih AC-C, Sheue C-R, Huang H-C, et al. 2013. Anatomical and transcriptional dynamics of maize embryonic leaves during seed germination. Proceedings of the National Academy of Sciences.110:3979-3984.

Mashiguchi K, Tanaka K, Sakai T, Sugawara S, Kawaide H, Natsume M, Hanada A, Yaeno T, Shirasu K, Yao H, et al. 2011. The main auxin biosynthesis pathway in Arabidopsis. Proceedings of the National Academy of Sciences.108:18512-18517.
Minic Z, Jouanin L. 2006. Plant glycoside hydrolases involved in cell wall polysaccharide degradation. Plant Physiology and Biochemistry.44:435-449.
Murase K, Hirano Y, Sun T-p, Hakoshima T. 2008. Gibberellin-induced DELLA recognition by the gibberellin receptor GID1. Nature.456:459-463.
Nambara E, Akazawa T, McCourt P. 1991. Effects of the gibberellin biosynthetic inhibitor uniconazol on mutants of Arabidopsis. Plant physiology.97:736-738.
Nemhauser JL, Hong F, Chory J. 2006. Different Plant Hormones Regulate Similar Processes through Largely Nonoverlapping Transcriptional Responses. Cell.126:467-475.
Ogawa M, Hanada A, Yamauchi Y, Kuwahara A, Kamiya Y, Yamaguchi S. 2003. Gibberellin biosynthesis and response during Arabidopsis seed germination. The Plant cell.15:1591-1604.
Olszewski N, Sun T-P, Gubler F. 2002. Gibberellin signaling: biosynthesis, catabolism, and response pathways. The Plant cell.14 Suppl:S61-S80.
Pagano A, Araujo SS, Macovei A, Leonetti P, Balestrazzi A. 2017. The Seed Repair Response during Germination: Disclosing Correlations between DNA Repair, Antioxidant Response, and Chromatin Remodeling in Medicago truncatula. Front Plant Sci.8:1972. Epub 2017/12/01.
Park J, Kim Y-S, Kim S-G, Jung J-H, Woo J-C, Park C-M. 2011. Integration of Auxin and Salt Signals by the NAC Transcription Factor NTM2 during Seed Germination in Arabidopsis. PLANT PHYSIOLOGY.156:537-549.
Park J, Oh D-H, Dassanayake M, Nguyen KT, Ogas J, Choi G, Sun T-p. 2017. Gibberellin Signaling Requires Chromatin Remodeler PICKLE to Promote Vegetative Growth and Phase Transitions. Plant Physiology.173:1463-1474.
Park MH, Suzuki Y, Chono M, Knox JP, Yamaguchi I. 2003. CsAGP1, a Gibberellin-Responsive Gene from Cucumber Hypocotyls, Encodes a Classical Arabinogalactan Protein and Is Involved in Stem Elongation. PLANT PHYSIOLOGY.131:1450-1459.
Rajjou L, Duval M, Gallardo K, Catusse J, Bally J, Job C, Job D. 2012. Seed Germination and Vigor. Annual Review of Plant Biology.63:507-533.
Ramaih S, Guedira M, Paulsen GM. 2003. Relationship of indoleacetic acid and tryptophan to dormancy and preharvest sprouting of wheat. Functional Plant Biology.30:939.
Reed JW, Wu MF, Reeves PH, Hodgens C, Yadav V, Hayes S, Pierik R. 2018. Three Auxin Response Factors Promote Hypocotyl Elongation. Plant Physiol. Oct;178:864-875. Epub 2018/08/25.
Ringli C, Keller B, Ryser U. 2001. Glycine-rich proteins as structural components of plant cell walls. Cellular and molecular life sciences : CMLS.58:1430-1441.
Rizza A, Jones AM. 2018. The makings of a gradient: spatiotemporal distribution of gibberellins in plant development. Curr Opin Plant Biol. Aug 30;47:9-15. Epub 2018/09/03.
Rizza A, Walia A, Lanquar V, Frommer WB, Jones AM. 2017. In vivo gibberellin gradients visualized in rapidly elongating tissues. Nature plants. Oct;3:803-813. Epub 2017/10/04.
Roxas VP, Smith RK, Allen ER, Allen RD. 1997. Overexpression of glutathione S-transferase/glutathioneperoxidase enhances the growth of transgenic tobacco seedlings during stress. Nature Biotechnology.15:988-991.
Ruf S, Kössel H, Bock R. 1997. Targeted inactivation of a tobacco intron-containing open reading frame reveals a novel chloroplast-encoded photosystem I-related gene. The Journal of cell biology.139:95-102.
Serpe MD, Nothnagel EA. 1994. Effects of Yariv phenylglycosides on Rosa cell suspensions: Evidence for the involvement of arabinogalactan-proteins in cell proliferation. Planta.193:542-550.
Shuai H, Meng Y, Luo X, Chen F, Zhou W, Dai Y, Qi Y, Du J, Yang F, Liu J, et al. 2017. Exogenous auxin represses soybean seed germination through decreasing the gibberellin/abscisic acid (GA/ABA) ratio. Scientific Reports.7:12620.
Song S, Qi T, Wasternack C, Xie D. 2014. Jasmonate signaling and crosstalk with gibberellin and ethylene. Curr Opin Plant Biol. Oct;21:112-119. Epub 2014/07/30.

Souza NM, Topham AT, Bassel GW. 2017. Quantitative analysis of the 3D cell shape changes driving soybean germination. Journal of Experimental Botany.68:1531-1537.
Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. PloS one.7:e21800.
Suzuki Y, Kitagawa M, Knox JP, Yamaguchi I. 2002. A role for arabinogalactan proteins in gibberellin-induced alpha-amylase production in barley aleurone cells. The Plant journal : for cell and molecular biology.29:733-741.
Tal I, Zhang Y, Jørgensen ME, Pisanty O, Barbosa ICR, Zourelidou M, Regnault T, Crocoll C, Erik Olsen C, Weinstain R, et al. 2016. The Arabidopsis NPF3 protein is a GA transporter. Nature Communications.7:11486.
Tanaka K, Asami T, Yoshida S, Nakamura Y, Matsuo T, Okamoto S. 2005. Brassinosteroid homeostasis in Arabidopsis is ensured by feedback expressions of multiple genes involved in its metabolism. Plant physiology.138:1117-1125.
Tanimoto E, Hirano K. 2013. Role of Gibberellins in Root Growth. In: Plant Roots: The Hidden Half. CRC Press. p. 13-11--13-14.
Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z. 2017. agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Research.45:W122-W129.
To JPC, Deruere J, Maxwell BB, Morris VF, Hutchison CE, Ferreira FJ, Schaller GE, Kieber JJ. 2007. Cytokinin Regulates Type-A Arabidopsis Response Regulator Activity and Protein Stability via Two-Component Phosphorelay. THE PLANT CELL ONLINE.19:3901-3014.
Tong H, Xiao Y, Liu D, Gao S, Liu L, Yin Y, Jin Y, Qian Q, Chu C. 2014. Brassinosteroid regulates cell elongation by modulating gibberellin metabolism in rice. Plant Cell. Nov;26:4376-4393. Epub 2014/11/06.
Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols.7:562-578.
Ueguchi-Tanaka M, Ashikari M, Nakajima M, Itoh H, Katoh E, Kobayashi M, Chow T-y, Hsing Y-iC, Kitano H, Yamaguchi I, et al. 2005. GIBBERELLIN INSENSITIVE DWARF1 encodes a soluble receptor for gibberellin. Nature.437:693-698.
Umezawa T, Sugiyama N, Mizoguchi M, Hayashi S, Myouga F, Yamaguchi-Shinozaki K, Ishihama Y, Hirayama T, Shinozaki K. 2009. Type 2C protein phosphatases directly regulate abscisic acid-activated protein kinases in Arabidopsis. Proceedings of the National Academy of Sciences.106:17588-17593.
Unterholzner SJ, Rozhon W, Papacek M, Ciomas J, Lange T, Kugler KG, Mayer KF, Sieberer T, Poppenberger B. 2015. Brassinosteroids Are Master Regulators of Gibberellin Biosynthesis in Arabidopsis. Plant Cell. Aug;27:2261-2272. Epub 2015/08/06.
Urbanova T, Leubner-Metzger G. 2016. Gibberellins and seed germination. In: Annual Plant Reviews, Volume 49. p. 253-284.
Voegele A, Linkies A, Muller K, Leubner-Metzger G. 2011. Members of the gibberellin receptor gene family GID1 (GIBBERELLIN INSENSITIVE DWARF1) play distinct roles during Lepidium sativum and Arabidopsis thaliana seed germination. Journal of Experimental Botany.62:5131-5147.
Wasternack C, Hause B. 2013. Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in Annals of Botany. Annals of Botany.111:1021-1058.
Weitbrecht K, Müller K, Leubner-Metzger G. 2011. First off the mark: Early seed germination. Journal of Experimental Botany.62:3289-3309.
Willats WG, Knox JP. 1996. A role for arabinogalactan-proteins in plant cell expansion: evidence from studies on the interaction of beta-glucosyl Yariv reagent with seedlings of Arabidopsis thaliana. The Plant journal : for cell and molecular biology.9:919-925.
Wu J, Mao X, Cai T, Luo J, Wei L. 2006. KOBAS server: a web-based platform for automated annotation and pathway identification. Nucleic Acids Research.34:W720-W724.
Xiang J, Wu H, Zhang Y, Zhang Y, Wang Y, Li Z, Lin H, Chen H, Zhang J, Zhu D. 2017. Transcriptomic Analysis of Gibberellin- and Paclobutrazol-Treated Rice Seedlings under Submergence. International Journal of Molecular Sciences.18:2225.

Yang DL, Yao J, Mei CS, Tong XH, Zeng LJ, Li Q, Xiao LT, Sun TP, Li J, Deng XW, et al. 2012. Plant hormone jasmonate prioritizes defense over growth by interfering with gibberellin signaling cascade. Proc Natl Acad Sci U S A. May 8;109:E1192-1200. Epub 2012/04/25.

Yang SF, Hoffman NE. 1984. Ethylene Biosynthesis and its Regulation in Higher Plants. Annual Review of Plant Physiology and Plant Molecular Biology.35:155-189.

Zentella R, Zhang Z-L, Park M, Thomas SG, Endo A, Murase K, Fleet CM, Jikumaru Y, Nambara E, Kamiya Y, et al. 2007. Global Analysis of DELLA Direct Targets in Early Gibberellin Signaling in Arabidopsis. The Plant Cell.19:3037-3057.

Zhao Y. 2014. Auxin Biosynthesis. The Arabidopsis Book.12:e0173.

Zhou DX, Bisanz-Seyer C, Mache R. 1995. Molecular cloning of a small DNA binding protein with specificity for a tissue-specific negative element within the rps1 promoter. Nucleic acids research.23:1165-1169.

Chapter 4 General discussion

The present doctoral thesis was guided by two hypotheses: 1) difference between GID1 subfamilies (mainly GID1ac and GID1b) on sequence, structural and expression level and these differences gained by GID1 sub-families during the course of evolution, leading to diversification of the GID1 genes in eudicots; 2) genes which are up- and down-regulated by paclobutrazole (PBZ), a gibberellins (GAs) biosynthesis inhibitor, are the targets of endogenous GA. The overall objective of this project was to achieve a better understanding of evolutionary history of GID1s in land plants and especially aimed at uncovering the gibberellins-regulated transcriptome profile associated with *G. max* germination. The results of evaluating each hypothesis is summed up and explained below in terms of the major findings and their significance.

Comprehensive genome-wide analysis of the GID1 gene family in land plants

Intrigued by the previous finding related to GID1: 1) phylogenetic classification of GID1s into GID1ac and GID1b subfamilies (Voegele et al. 2011); 2) diverse expression pattern of GID1ac and GID1b in different tissues (Bellieny-Rabelo et al. 2016, Griffiths et al. 2006); 3) GA-regulated transcriptional down-regulation of GID1ac, but GID1b (Voegele et al. 2011); 4) higher affinity of GID1b for GA_3/GA_4 than GID1a and GID1c (Nakajima et al. 2006), a comprehensive genome-wide analysis of the GID1 gene family was conducted to study the expansion and diversification of GID1s in land plants. After careful screening, a total of 141 full-length GID1 sequences from 54 species were used for further analyses. Phylogenetic analyses divide the GID1 gene family into four clades. This phylogenetic separation is supported by previously proposed phylogenetic classification of this family in land plants (Voegele et al. 2011). This study identified that expansion and diversification of GID1s happened after the emergence of angiosperms. In addition, the study revealed contribution of whole-genome duplication to lineage-specific expansions of sub-families (GID1ac and

GID1b) in different eudicot lineages. More interestingly, by analyzing shared and divergent structural features of GID1ac and GID1b sub-families, this study uncovered important divergent residues in the GID1b GA-binding pocket that might provide increased GA affinity. Moreover, gene expression data from several species show that at least one GID1 gene is expressed in every sampled tissue, with a strong bias of GID1b expression towards underground tissues and dry legume seeds (which typically have low GA levels). Together, these data provide an in-depth look at evolution, conserved and divergent features and expression of the GID1 gene family in land plants.

GA-regulated transcriptional profile of soybean during germination

Gibberellins (GA) are key positive regulators of seed germination (Groot et al. 1987, Karssen et al. 1989, Urbanova and Leubner-Metzger 2016). GA enhances seed germination by promoting cell elongation and weakening of the surrounding tissues (Kucera et al. 2005, Ogawa et al. 2003). Although the GA effects on seed germination have been studied in a number of species (Groot et al. 1987, Ikuma and Thimann 1960, Karssen et al. 1989, Yomo and Iinuma 1966), little is known about the transcriptional reprogramming modulated by GA during this phase in species other than *Arabidopsis thaliana* (Cao et al. 2005, Ogawa et al. 2003, Zentella et al. 2007). Therefore, to identify GA-responsive genes, we performed a transcriptome study in a time-course experiment during soybean seed germination using paclobutrazol (PBZ), a GA biosynthesis inhibitor. A total of 770 genes were identified as differentially expressed upon PBZ treatment. We found a number of differentially expressed cell wall metabolism genes, supporting their roles in cell expansion during germination. Several genes involved in the biosynthesis and signaling of other phytohormones were also modulated, indicating an intensive hormonal crosstalk at the embryonic axis. Interestingly, we also found 26 photosynthesis genes that are up-regulated by PBZ at 24 hours of imbibition (HAI) and down-regulated at 36 HAI, which led us to suggest that this is part of a strategy to implement an autotrophic growth program in the absence of GA-driven mobilization of reserves. Finally, we identified 30

transcription factors (mostly from the MYB, bHLH and bZIP families) that are down-regulated by PBZ and are likely downstream GA targets that will drive transcriptional changes during germination.

REFERENCES

Bellieny-Rabelo D, Alves Gamosa de Oliveira E, da Silva Ribeiro E, Pessoa Costa E, Elenir Amâncio Oliveira A, Motta Venancio T. 2016. Transcriptome analysis uncovers key regulatory and metabolic aspects of soybean embryonic axes during germination. Scientific Reports.6:36009.

Cao D, Hussain A, Cheng H, Peng J. 2005. Loss of function of four DELLA genes leads to light- and gibberellin-independent seed germination in Arabidopsis. Planta.223:105-113.

Griffiths J, Murase K, Rieu I, Zentella R, Zhang Z-L, Powers SJ, Gong F, Phillips AL, Hedden P, Sun T-p, et al. 2006. Genetic characterization and functional analysis of the GID1 gibberellin receptors in Arabidopsis. The Plant cell.18:3399-3414.

Groot SPC, Bruinsma J, Karssen CM. 1987. The role of endogenous gibberellin in seed and fruit development of tomato: Studies with a gibberellin-deficient mutant. Physiologia Plantarum.71:184-190.

Ikuma H, Thimann KV. 1960. Action of Gibberellic Acid on Lettuce Seed Germination. Plant physiology.35:557-566.

Karssen CM, Zaorski S, Kepczynski J, Groot SPC. 1989. Key Role for Endogenous Gibberellins in the Control of Seed Germination. Ann Bot.63:71-80.

Kucera B, Cohn MA, Leubner-Metzger G. 2005. Plant hormone interactions during seed dormancy release and germination. Seed Science Research.15:281-307.

Nakajima M, Shimada A, Takashi Y, Kim YC, Park SH, Ueguchi-Tanaka M, Suzuki H, Katoh E, Iuchi S, Kobayashi M, et al. 2006. Identification and characterization of Arabidopsis gibberellin receptors. Plant Journal.46:880-889.

Ogawa M, Hanada A, Yamauchi Y, Kuwahara A, Kamiya Y, Yamaguchi S. 2003. Gibberellin biosynthesis and response during Arabidopsis seed germination. The Plant cell.15:1591-1604.

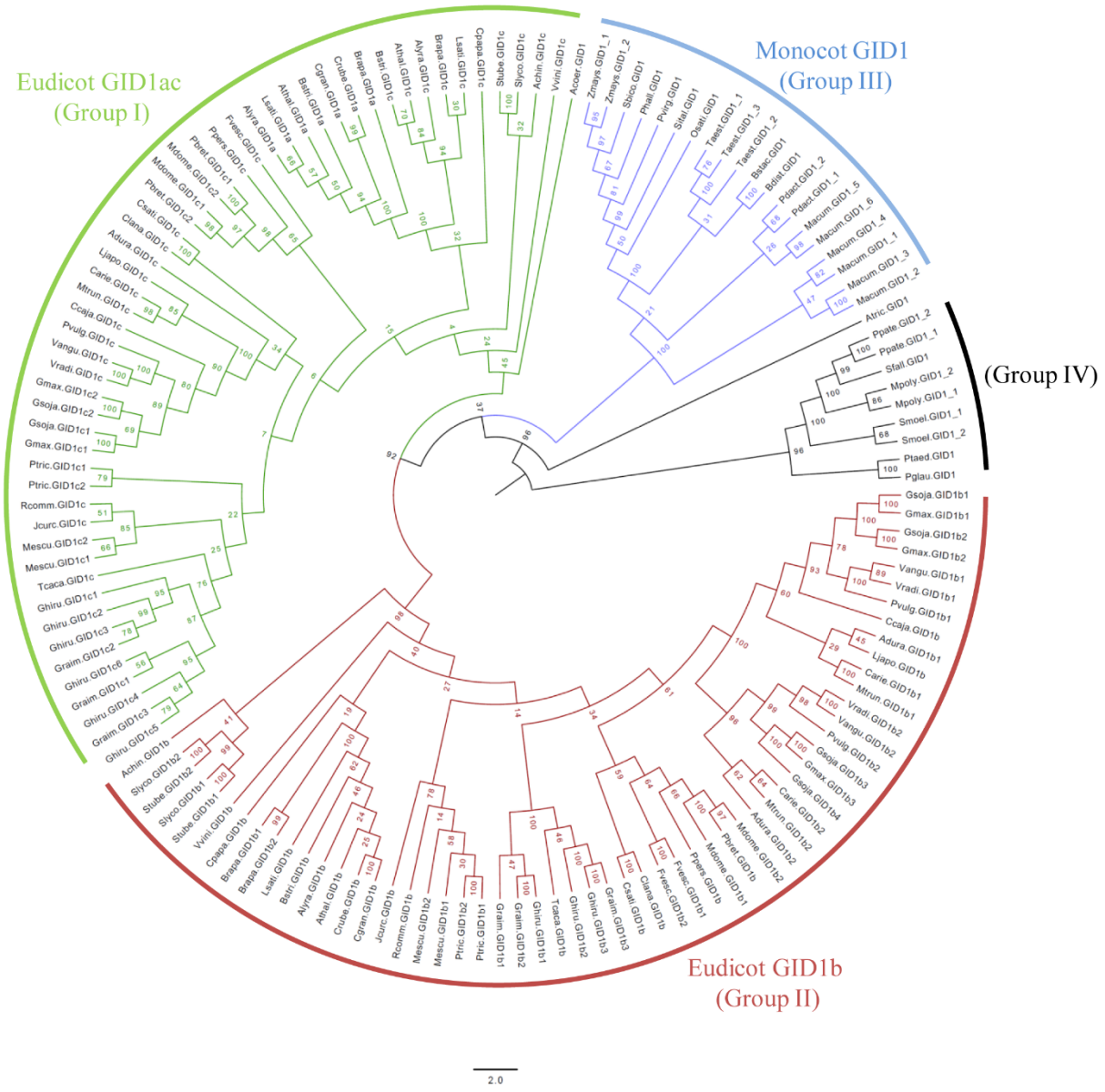
Urbanova T, Leubner-Metzger G. 2016. Gibberellins and seed germination. In: Annual Plant Reviews, Volume 49. p. 253-284.

Voegele A, Linkies A, Muller K, Leubner-Metzger G. 2011. Members of the gibberellin receptor gene family GID1 (GIBBERELLIN INSENSITIVE DWARF1) play distinct roles during Lepidium sativum and Arabidopsis thaliana seed germination. Journal of Experimental Botany.62:5131-5147.

Yomo H, Iinuma H. 1966. Production of gibberellin-like substance in the embryo of barley during germination. Planta.71:113-118.

Zentella R, Zhang Z-L, Park M, Thomas SG, Endo A, Murase K, Fleet CM, Jikumaru Y, Nambara E, Kamiya Y, et al. 2007. Global Analysis of DELLA Direct Targets in Early Gibberellin Signaling in Arabidopsis. The Plant Cell.19:3037-3057.

APPENDIX A



APPENDIX A.1: Phylogenetic reconstruction of the 141 GID1 proteins identified in 54 plant species.

Multiple sequence alignment of GID1 proteins was carried with PROMALS3D and phylogenetic reconstruction was performed with RAxML. GID1 proteins were classified in four groups, represented in different colors. Node labels indicate bootstrap support values.

APPENDIX B

APPENDIX B1: Pairwise correlation between biological replicates.

Pairwise Samples	Pearson correlation
12CR1 v/s 12CR2	0.98
12CR1 v/s 12CR3	0.98
12CR2 v/s 12CR3	0.99
24CR1 v/s 24CR2	0.98
24CR1 v/s 24CR3	0.98
24CR2 v/s 24CR3	0.99
36CR1 v/s 36CR2	0.98
36CR1 v/s 36CR3	0.98
36CR2 v/s 36CR3	0.97
12PR1 v/s 12PR2	0.95
12PR1 v/s 12PR3	0.98
12PR2 v/s 12PR3	0.97
24PR1 v/s 24PR2	0.98
24PR1 v/s 24PR3	0.98
24PR2 v/s 24PR3	0.99
36PR1 v/s 36PR2	0.98
36PR1 v/s 36PR3	0.99

36PR2 v/s 36PR3	0.98

APPENDIX B2: TRANSCRIPTIONAL LEVELS (IN RPKM) OF EXPRESSED SOYBEAN PROTEIN-CODING GENES.

This list of expressed genes presented in Chapter 3 may be found in S3 file in supplemental sheet 480814_file04.xlsx2 at below link:

https://www.biorxiv.org/content/biorxiv/early/2018/11/29/480814/DC2/embed/media-2.xlsx?download=true

APPENDIX B3: THE DIFFERENTIALLY EXPRESSED GENES BETWEEN CONTROL VS PBZ AT 12 HAI, 24 HAI AND 36 HAI.

This list of differentially expressed genes presented in Chapter 3 may be found in a supplemental file name 480814_file05.xlsx at below link:

https://www.biorxiv.org/content/biorxiv/early/2018/11/29/480814/DC3/embed/media-3.xlsx?download=true

APPENDIX B4: GENES THOSE ARE UP-REGULATED BY PBZ AT 24 HAI AND DOWN-REGULATED AT 36 HAI.

This list of genes up- and down-regulated by PBZ at 24 HAI and 36 HAI, respectively, presented in chapter 3 may be found in a S5 file in supplemental sheet 480814_file04.xlsx2 at below link:

https://www.biorxiv.org/content/biorxiv/early/2018/11/29/480814/DC2/embed/media-2.xlsx?download=true

APPENDIX B5: Transcriptional level (in RPKM) of hormone biosynthesis, signaling and transporter genes (many of them are also shown in Figures 3 and 4).

This list of genes presented in chapter 3 may be found in supplementary file name 480814_file06.xlsx at below link:

https://www.biorxiv.org/content/biorxiv/early/2018/11/29/480814/DC4/embed/media-4.xlsx?download=true

APPENDIX B6: Blastp analysis of genes modulated by GA in A*rabidopsis thaliana* and Soybean. Refer to the main paper for details on the datasets.

This list of GA regulated genes in ***A. thaliana*** **and soybean**, presented in chapter 3 may be found in a S5 file in supplemental sheet 480814_file04.xlsx2 at below link:

https://www.biorxiv.org/content/biorxiv/early/2018/11/29/480814/DC2/embed/media-2.xlsx?download=true