The Pennsylvania State University

The Graduate School

Graduate Degree Program in the Huck Institutes of the Life Sciences

FUNCTIONAL GENOMICS OF THEOBROMA CACAO FATTY ACID BIOSYNTHESIS:

CONVERGENCE OF FATTY ACID DESATURATION, EMBRYO DEVELOPMENT, AND DEFENSE SIGNALING RESPONSES

A Dissertation in

Plant Biology

by

Yufan Zhang

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2014

The dissertation of Yufan Zhang was reviewed and approved* by the following:

Mark J. Guiltinan
Professor of Plant Molecular Biology
Dissertation Advisor
Chair of Committee

Sarah M. Assmann
Waller Professor of Biology

Siela N. Maximova
Professor of Horticulture

Ryan J. Elias
Associate Professor of Food Science

James L. Rosenberger
Professor of Statistics

Teh-hui Kao
Distinguished Professor of Biochemistry and Molecular Biology
Director of Plant Biology Program

*Signatures are on file in the Graduate School

ABSTRACT

Theobroma cacao L. (chocolate tree) is an important cash crop for 40-50 million farmers and their families in its tropical growing regions worldwide. Cocoa butter and cocoa powder extracted from cacao seeds provide the main raw ingredients for chocolate manufacturing, supporting a $80 billion global business. A unique fatty acid composition of cocoa butter makes its melting temperature close to the human body temperature, which is not only of particular importance for industrial uses, but also a valuable quality trait targeted by breeding programs. My Ph.D. dissertation focused mainly on the fatty acid biosynthesis pathway in cacao seeds. I identified a key desaturase gene *TcSAD1* from a large stearoyl-acyl carrier protein-desaturase gene family in cacao that plays a crucial role in converting stearic acid (18:0, saturated fatty acid) into oleic acid (18:1, unsaturated fatty acid). The expression of *TcSAD1* was highly correlated with the change of fatty acid composition during cacao seed development. The activity of TcSAD1 rescued all the Arabidopsis *ssi2* (a fatty acid desaturase) related mutant phenotypes, further supporting its *in vivo* functions. The discovery of the critical function of TcSAD1 offers a new strategy for screening for novel genotypes with desirable fatty acid compositions, and for use in breeding programs, to help pyramid genes for quality traits such as cocoa butter content. Moreover, because of the significance of fatty acid biosynthesis and lipid accumulation during cacao seed development, to further explore the regulatory mechanism, I functionally characterized of a master regulator, *TcLEC2* gene, which controls both zygotic and somatic embryo development of cacao. Transient overexpression of *TcLEC2* induced the expression of a variety of seed specific genes in cacao leaves. Furthermore, functions of TcLEC2 were explored during somatic embryogenesis, which is an *in vitro* propagation system for cacao. My results suggested that the activity of TcLEC2 determines the embryogenic capacity of the cacao tissue explants and correlated with embryogenic capacity of cultured cells. Transgenic embryos overexpressing *TcLEC2* produced a significantly higher number of embryos compared to non-transgenic embryos; however, most of these transgenic somatic

embryos exhibited abnormal phenotypes, and the development normally ceased at globular stage. This discovery may have future applications in increasing the efficiency of cacao mass propagation programs. Notably, in addition to major storage compounds in cacao seeds, fatty acids also function as signals involved in defense responses. I found that the endogenous level of 18:1 was modulated by exogenous glycerol application. Glycerol application on cacao leaves increased the level of glycerol-3-phosphate and lowered the level of 18:1 through an acylation reaction, which further triggered the defense responses. 100mM glycerol was sufficient to induce the accumulation of ROS, activate the expression of a variety of pathogen-related genes, and confer enhanced resistance against fungal pathogen *Phytophthora capsici*. My results demonstrated the potential of foliar glycerol application to become an environmentally safe means to induce the plant defense responses and fight important plant diseases in the field.

Together, my Ph.D. dissertation makes major contributions to three important research areas in cacao: (1) identification of the key gene regulating fatty acid composition in cocoa butter, (2) improvement of large-scale propagation system (somatic embryogenesis) of cacao, (3) enhancement of cacao foliar disease resistance. This thesis not only provides useful knowledge of the regulatory mechanisms of important quality traits at the molecular and genetic levels, but also demonstrates the potential of taking advantage of cacao genomic resources to accelerate cacao basic research and breeding programs.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am so thankful and proud to finally have this great opportunity to express my sincere appreciations and respects to those significant people in my life. I will never be who I am and where I am without your presence in my life. It always makes me nervous when I try to put something into words since I doubt I am capable of conveying my feelings thoroughly and precisely, but I would like to try.

Thank you my parents to raise and educate me. I hate to disappoint you in any kind, and that is probably what motivates me every day. I hope I lived up to your expectation. You've never thought your son be a Ph.D., have you? That is a big surprise to you I presume.

I couldn't appreciate more for the support, encouragement and guidance from my advisors Dr. Mark Guiltinan and Siela Maximova. I am so lucky to conduct my Ph.D. research in a lab I feel like home. You not only give me a "fish" that I can eat for a day, but also teach me to "fish" that I think I will eat for a lifetime. I am very grateful to Dr. Sarah Assmann, Dr. Ryan Elias, and Dr. James Rosenberger for serving on my committees and providing valuable comments to my thesis. Dr. Ryan Elias, especially thank you for offering me the opportunity to work in your food chemistry lab, which was a wonderful experience and expanded my knowledge. And of course Dr. Philip Smith, I want to thank you. Without your generous assists on the metabolomics analysis and technical advice, I could not even imagine any of my research goals could be ever achieved. My thanks also extend to all my lab mates in Dr. Guiltinan group I have been working with: Ann Young, Sharon Pishak, Merissa Kenzakoski, Adam Clemens, Luis Mejia, Sergio Florez, Andrew Fister, and Lena Sheaffer, for your numerous technical supports and inspirative discussions. I would also like to give my special thank to Dr. Zi Shi and Dr. Yuannian Jiao who provided tremendous help to me and welcomed me like a family. Wish you guys the very good luck in Georgia. I also want to thank my friends and teammates on the basketball team, Xiao Liu, Zhicheng

Tian, Kuo Han, Wei Luo, and etc. I always cherish the time we were running the court together.

Most importantly, is my wife Iris. Lucky me I met and married you. That is the best thing ever happened to me. Thank you for your unconditional support, encouragement and consideration.

To all my beloved you, I dedicate this thesis.

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction to *Theobroma cacao*

1.1.1 Botany

Theobroma cacao L. (chocolate tree) is an evergreen understory tropical tree in the family of *Malvaceae* (Sereno et al., 2006; Argout et al., 2011). It is endemic to the Amazon basin and is widely grown in West Africa, Central and South America, Malaysia, Indonesia and elsewhere in the humid tropics worldwide, where cacao production is associated with the livelihood of 40-50 million farmers and their families. Cacao seed provides the main ingredients for chocolate manufacturing and some other products, such as cocoa liquor, cocoa powder and cocoa butter. Cacao starts to produce small flowers directly on the trunk when they are about age three or four, and flowers are present throughout the year. Even though a few cacao cultivars are self-compatible, most cacao varieties are self-incompatible, and thus the pollination process needs to be facilitated by insect pollinators, such as midges of genus *Forcipomyia* and some species of aphids and thrips (Glendinning, 1972; Kaufmann, 1975). Once the flowers are successfully pollinated, it takes about 20 weeks to develop into a fully mature fruit, called a cocoa pod. Commonly, each cocoa pod contains 30~40 seeds, also known as cocoa beans. Cocoa pods can be harvested year-round for about 50 years.

Three major cacao cultivars are widely grown: Criollo, Forastero, and Trinitario. Among them, Criollo, the first cultivated cacao by humans, is a homozygous genotype producing white seeds with fine, fruity, and aromatic flavors (Frauendorfer and Schieberle, 2008; Argout et al., 2011). However, due to the poor agronomic productivity and disease susceptibility, only less than 10% of chocolate is made from Criollo (Rusconi and Conti, 2010). In contrast, approximately 80~90% of world chocolate production comes from Forastero, a genotype with outstanding yields and higher disease resistance, introduced from Amazon basin into Mesoamerica and Western Africa. Forastero produces purple seeds with rich chocolate flavors. Trinitario is a hybrid genotype between Criollo and Forastero with various and unique flavors (Brunetto et al., 2007; Sukha et al., 2008).

1.1.2 Major Compositions of Cacao Seeds

The main constituent of cacao seed is the lipid fraction, also known as cocoa butter, which accounts for approximate 50% weight of the cocoa bean (Lehrian and Keeney, 1980; Liendo et al., 1997). Cocoa butter is the storage lipid synthesized and deposited in the form of triacylglycerol during seed maturation. Notably, cocoa butter is composed of almost equal amount of palmitic acid (16:0), stearic acid (18:0) and oleic acid ($18:1^{n-9}$), the composition of which determines its unique melting temperature very close to human body temperature, thus providing the smoothness and mouth feel of chocolate, as well as the creamy texture of cosmetics on skin (Liendo et al., 1997). The solid residue after cocoa butter is extracted is called cocoa powder, which is rich in flavonoids and associated with the bitterness and astringent tastes of the chocolate. The main flavonoid compounds in fermented cacao beans are catechins ((-)-epicatechins, (+)-catechins, (+)-gallocatechins, and (-)-epigallocatechins), anthocyanins, and proanthocyanins. Notably, most of the catechins and epicatechins present in cocoa seeds are in a form of long polymers instead of dimers or trimers found in most flavonoid-rich foods (Rusconi and Conti, 2010). In addition, there is also about 2-10% of alkaloids, such as caffeine and theobromine, contained in the cocoa powder. Proteins, mainly composed of albumin (52%) and globulin (43%), constitute approximate 10-15% of the dry weight of cacao seeds.

1.1.3 Industrial Uses and Health Benefits of Cacao Seeds

After harvest, cocoa beans are naturally fermented and dried in plantations, and exported to the cacao product processors in Europe and North America. Cacao beans are roasted, Tand the shells are next removed before they are milled into chocolate liquor before the cocoa butter and cocoa powder are extracted. Cocoa butter and cocoa powder provide the main raw ingredients used for chocolate manufacturing. In 2011, the global market value of the chocolate industry surpassed $100 billion and the demand for cacao beans continues to increase (Anonymous, 2012). In North America and many North European countries, people consume lots of chocolate, which is regarded as one of the most concentrated sources of flavonoids from common diets. The beneficial effects of flavonoids on human health have been well recognized (Ross and Kasum, 2002; Yao et al., 2004). As the oxidative stress is implicated as a causal agent or contributing factor for many severe human

diseases, such as cancer, metabolic disorders, cardiovascular and inflammatory diseases (I. Andújar, 2012), the antioxidant capacity of flavonoids as a major component in cocoa beans and chocolate could increase the resistance to oxidative stress, and thus inhibit the lipid oxidation, prevent LDL-cholesterol oxidation, and lower the risk of the occurrences of those diseases, which could possibly explain why cocoa and chocolate have been described as potential medicine since the seventeenth century (Crozier et al., 2011).

1.1.4 Genomics and Molecular Genetics of Cacao

Theobroma cacao is a diploid tree with ten pairs of chromosomes (2n = 20) (Davie, 1935). Given the economic significance and the awareness of the importance of the sustainability of cocoa plantations, cacao whole genome sequencing projects were initiated and completed through several collaborative international teams. In 2010, the genome of a homozygous Belizean Criollo genotype (B97-61/B2) was first fully sequenced and assembled, the estimated size of which was 430-Mb. Key genes related with disease resistance, fatty acid composition, cocoa butter content, and flavonoid biosynthesis were surveyed genome-wide and functionally annotated (Argout et al., 2011). For example, the functional orthologous *NPR1* gene in cacao, *TcNPR1*, was successfully identified in the genome, and the *NPR* gene family in cacao was discovered and phylogenetically characterized. Moreover, 84 orthologous *T. cacao* genes potentially involved in fatty acid biosynthesis pathway were also identified in the genome. Later, a traditional cultivar Matina 1-6 in Amelonado genetic group was also sequenced, which represents a more widely grown variety of Foratero (Motamayor et al., 2013). Comparison of these two whole genome sequences revealed highly conserved synteny with 271 orthologous regions and the majority of genes identified are located in these conserved orthologous regions (Motamayor et al., 2013). The availability of two genome sequences, together with the cacao EST database and QTL maps greatly assists the biomarker discovery process and profoundly accelerates cacao-breeding programs, which opens a completely new era for cacao research (details in Section 1.4.3).

1.2 Embryogenesis

1.2.1 Comparison between Zygotic Embryogenesis and Somatic Embryogenesis

Embryogenesis defines the process during which a single cell differentiates into a mature embryo (Braybrook and Harada, 2008). In seed plants, zygotic embryogenesis (ZE) consists of two major developmental phases: morphogenesis phase and maturation phase. During the morphogenesis phase, through a series of cell division and differentiation, a fertilized zygote experiences the globular, heart-shape, torpedo stage and finally develops into an embryo with well-established body plan and clear specifications of shoot and root apical meristems. During the maturation phase, storage compounds, such as proteins, storage lipids, starch, and sugars, rapidly accumulate as the energy reservoir for the following germination. Zygotic embryos then normally acquire desiccation tolerance and become metabolically quiescent until the seed meets a favorable environmental condition suitable for germination and growth of the young seedling (Braybrook and Harada, 2008; Karami et al., 2009). However, distinctively, cacao, as well as some other tropical tree species, such as avocado, mango, and lychee, produces recalcitrant seeds, which are intolerant to desiccation and low temperature, and thus difficult for cryopreservation (Chandel et al., 1995).

Somatic embryogenesis (SE) is a developmental process originating from a somatic cell(s) without gametic fusion. In theory, every somatic cell obtains all the genetic materials required to regenerate into a whole plant, a capacity called totipotency (Karami et al., 2009). Since SE happens in an *in vitro* system, exogenously applied plant growth regulators, such as auxins and cytokinins, are required in the culture media to induce the process, which are normally provided by the maternal surrounding tissues during ZE (Santos-Mendoza et al., 2008; Gliwicka et al., 2013). Despite the distinctions, both ZE and SE share substantial similarities in morphological development stages and molecular regulatory mechanisms (Karami et al., 2009; Yang and Zhang, 2010). For example, many embryogenesis related genes play critical roles in both processes, such as *LEC* genes, which will be further discussed in the following sections. Therefore, SE is commonly used as a model system to study embryogenesis (Braybrook and Harada, 2008).

Generally, two different types of SE have been recognized: direct SE and indirect SE. Direct SE refers to a process in which the formation of somatic embryo experiences

minimal proliferation of dedifferentiated cells (Kurczynska et al., 2007), and it is believed that pre-embryogenic competent cells are pre-existing. Direct SE is induced without gene reprogramming to acquire the embryogenic status (Karami et al., 2009). However, during indirect SE, somatic cells first lose their cell identities and dedifferentiate into a cluster of undifferentiated cells, also known as callus. Then, the embryogenic pathway is activated in only certain callus cells, which differentiate into meristematic cells from which somatic embryos initiate and develop (Karami et al., 2009). The source of the explants used to initiate SE is critical to which pathway will predominate. For example, in Arabidopsis, immature zygotic embryos from globular to the early cotyledonary stage of development tend to go through indirect SE pathway to produce somatic embryos, while more advanced late cotyledonary zygotic embryos favor the direct SE pathway (Kurczynska et al., 2007).

1.2.2 Cacao Somatic Embryogenesis

Cacao trees are generally highly heterozygous and when propagated by seed, only a small fraction of individuals are high producing (Cheesman and Pound, 1932; Irizarry and Rivera, 1999; Irizarry and Goenaga, 2000). Thus, vegetative propagation systems provide a means to avoid the issue of trait variation, through cloning of the top elite individual genotypes. Several methods of vegetative propagation are commonly used with cacao (grafting and rooted cuttings techniques). In addition, *in vitro* somatic embryogenesis (SE) tissue culture offers an approach to accelerate the development and deployment of genetically improved genotypes because of its potentially very high multiplication rate and scalability.

Protocols for the regeneration of cacao plants via primary and secondary SE have been well documented (Pence et al., 1979; Lopez Baez et al., 1993; Li et al., 1998; Maximova et al., 2002). Floral explants, such as petals and staminodes in unopened cacao flowers, are the most commonly used and embryogenic tissues to induce primary SE. Briefly, after sterilization, floral tissues are placed on a culture media containing 2,4-D and thidiazuron (TDZ) for two weeks, followed by a second culture media containing 2,4-D and 6-benzylaminopurine (6BA) for another two weeks for callus initiation. Explants, normally entirely covered by compact callus at the time, are then transferred and maintained on hormone free embryo development (ED) media. Usually after another 2~3 weeks cultured on ED media, primary somatic globular embryos start to emerge from the surface of callus

cells with suspensors still attached to the explants. Subsequently, globular embryos further develop into heart shape, torpedo, and mature embryos with clear appearance of hypocotyl and cotyledons. Mature embryos are normally transferred to rooting media under light for cotyledon expansion, root elongation and vegetative shoot initiation. Healthy cacao plantlets with 3-4 green leaves and well-developed root system can be then potted into soil in the greenhouse, acclimated to *ex vitro* conditions and grown into plants. . Additionally, SE of cacao can be induced reiteratively using the cotyledons of somatic embryos with a 30-fold increase in SE efficiency (Maximova et al., 2002). Histology analysis revealed that primary and secondary SE might differ in the cellular origin of somatic embryos. Contrary to the multi-cellular origin of primary embryos, secondary embryos predominately emerge from the division of single cells, which possibly explained the higher SE efficiency of secondary SE compared to primary SE (Maximova et al., 2002). Remarkably, a comprehensive field test, including somatic embryo-derived plants, root cuttings, and conventional seed propagated plants for a total of 214 cacao seedlings, concluded that somatic embryo-derived plants exhibited normal phenotypes in the field environment with similar growth parameters to the conventional seed propagated plants (Maximova et al., 2008).

However, although SE is a favorable propagation system for cacao, varied embryogenesis efficiency and relatively low conversion rate are considered two major bottlenecks in the current system that needs to be further improved. For example, in one study (Li et al., 1998), even though somatic embryos have been successfully obtained from 19 genotypes representing three major cultivars, large variations in both frequencies of embryogenesis and average number of somatic embryos produced from each explant were observed, ranging from 1 to 100%, and 2 to 46/explant, respectively, which exist in both primary and secondary SE (Maximova et al., 2002). Moreover, developmental abnormalities of somatic embryos were frequently noted in the current system, such as embryos with fused hypocotyles or multiple cotyledons, which compromises the success rate of conversion. Therefore, a deeper understanding of mechanisms regulating the SE process in cacao could potentially lead to improvement of SE methods for commercial plant production. In this thesis, by leveraging the knowledge gained from the model plant

Arabidopsis, I aimed to identify key genes in cacao that determines the embryogenic potential of the explants and contributes to somatic embryo development, to facilitate the better characterization of the mechanisms of cacao SE and provide us the guidance to further optimize the current protocol.

1.2.3 Transcriptional Network Regulating Somatic Embryogenesis

In Arabidopsis, leafy cotyledon (*LEC*) transcription factors, including *AtLEC1* (Lotan et al., 1998), *AtLEC2* (Stone et al., 2001) and *AtFUS3* (Luerssen et al., 1998) have been characterized as master regulators of both zygotic and somatic embryo development (Braybrook and Harada, 2008). The *AtLEC2* gene encodes a B3 domain transcription factor, which binds specifically to the *RY* motifs in the 5' flanking regions of *AtLEC2*-induced genes (Braybrook et al., 2006). *AtLEC2* is exclusively expressed in developing zygotic embryos during both the early development and maturation phases. It is required for development and maintenance of suspensors and cotyledons and for the acquisition of desiccation tolerance and inhibition of premature germination (Stone et al., 2001). Loss-of-function Arabidopsis *lec2* mutants exhibit pleiotropic effects including abnormal suspensor anatomy, abnormal cotyledons with trichomes, precociously activated shoot apical meristems, highly pigmented cotyledon tips with prominent anthocyanin accumulation and reduced accumulation of seed storage compounds (Stone et al., 2001; Ledwon and Gaj, 2009; Feeney et al., 2013). AtLEC2 exerts its functions both by inducing a cascade of other transcription factors controlling various developmental and metabolic pathways, and directly targeting and regulating seed storage genes (Santos Mendoza et al., 2005; Baud et al., 2007). For example, *AtWRI1*, another key transcription factor crucial to embryo development, is a direct target of *AtLEC2* and is necessary to regulate normal fatty acid biosynthesis (Baud et al., 2007).

LEC genes are also important during somatic embryogenesis. For example, *lec2* mutants produced SEs in Arabidopsis at a very low efficiency (Gaj et al., 2005), while ectopic expression of *AtLEC2* in Arabidopsis and tobacco vegetative tissue induced SE formations (Stone et al., 2001; Zuo et al., 2002; Guo et al., 2013). In addition, the capacity for SE was abolished in double (*lec1 lec2, lec1 fus3, lec2 fus3*) or triple (*fus3 lec1 lec2*) *LEC* mutants, which further confirms the critical and redundant roles of LEC proteins during SE

(Gaj et al., 2005). It is well known that exogenous application of hormones, such as synthetic auxin (2,4-D) and cytokinin, are required to induce SE (Fujimura and Komamine, 1979; Matsuta, 1992; Visser et al., 1992) and furthermore, a functional interaction between auxin and *AtLEC2* has been observed. In Arabidopsis, the expression of *AtLEC2* was significantly up-regulated in response to exogenously applied 2,4-D during the induction phase of SE (Ledwon and Gaj, 2009). Also, expression levels of *AtLEC2* were observed to be significantly higher in embryogenic callus compared to the non-embryogenic callus of the same age (Ledwon and Gaj, 2009). Interestingly, overexpression of *AtLEC2* in immature zygotic embryo transgenic explants was able to induce direct somatic embryogenesis, with little callus formation and in the absence of exogenous auxin (Ledwon and Gaj, 2009). Regarding this, Stone and Wojcikowska proposed that *AtLEC2* may activate genes involved in auxin biosynthesis, such as *YUC1*, *YUC2*, *YUC4* and *YUC10* (Stone et al., 2008; Wojcikowska et al., 2013). Taken together, *AtLEC2* is essential for maintaining embryogenic competency of plant somatic cells through complex interactions with transcriptional regulators and auxin (Karami et al., 2009).

1.3 Fatty Acid Biosynthesis in Plants

1.3.1 Functions of Fatty Acids and Lipids in Plants

Fatty acid biosynthesis is a fundamental metabolic pathway and occurs in every plant cell. However, after synthesis, only a fraction of fatty acids are present in the cells in the form of free fatty acids (Wang et al., 2000). Instead, they are modified and esterified to various backbones to form different kinds of lipids, such as glycerolipids, phospholipids, galactolipids, and sulfolipids. Lipids account for about 5-10% of the dry weight in plant vegetative tissues, and this value can reach as high as 50% in plant seeds, like cocoa beans (Crozier et al., 2011). Lipids play vital roles in many important biological processes, and blocking of fatty acid biosynthesis is usually lethal to the cell (Ohlrogge and Browse, 1995). For example, serving as the major component of biological membranes, phospholipid bilayers not only define the cell shape and region, but also compartmentalize different organelles within the cell, with specialized functions. The bio-membranes are also where

many important biochemical reactions take place, such as photosynthesis, respiration, and electron transfer reactions (Zhu et al., 2010; Horváth et al., 2012). The fatty acid composition of phospholipids greatly influences membrane fluidity, which is crucial to the cell normal activity.

In seeds, triacylglycerol is employed as the major form of energy reservoir, which is not only essential for seed germination, but also serves as an important food source for human beings (Theodoulou and Eastmond, 2012). In particular, fatty acids, like linolenic acid, can also be used as the precursor of jasmonic acid, which, as a second messenger, plays significant roles during signaling transduction pathway (Kachroo and Kachroo, 2009).

1.3.2 Fatty Acid Biosynthesis Pathway in Plants

The fatty acid biosynthesis pathway takes place in the plastid in plants (Figure 1-1). Acetyl-CoA, the initial carbon source, is provided by various biochemical reactions from both inside and outside the plastid. For example, in the leaf tissue, most acetyl-CoAs are produced from pyruvate decarboxylation during glycolysis in the plastid. Catalyzed by ACCase, acetyl-CoA is converted into malonyl-CoA, and further, the carboxy group is transferred to acyl carrier protein (ACP) in the priming step. Malonyl-ACP is then utilized as building blocks in the following repeated series of elongation reactions. During each elongation step, malonyl-ACP containing two carbon atoms is added into the pre-existing acyl chains by fatty acid synthase complex (FAS) until the saturated palmitic acid (16:0-ACP) and stearic acid (18:0-ACP) are formed. At this stage, both 16:0 and 18:0 can be desaturated by a family of soluble stromal Δ9-acyl-ACP desaturase in plastids to yield palmitoleic acid ($16:1^{n-9}$) and oleic acid ($18:1^{n-9}$). Subsequently, all the fatty acid acyl moieties (16:0, 18:0, $16:1^{n-9}$, $18:1^{n-9}$) are released from ACP by acyl-ACP-thioesterase, such as FATB and FATA, and acylated onto phosphorylated glycerol backbone in either ER or plastids. Further elongation and desaturation of fatty acid acyl moieties occur on the phosphorylated glycerol backbone to produce linoleic acid ($18:2^{n-9,12}$) and α-linolenic acid ($18:3^{n-9,12,15}$), and other long chain saturated and unsaturated fatty acids.

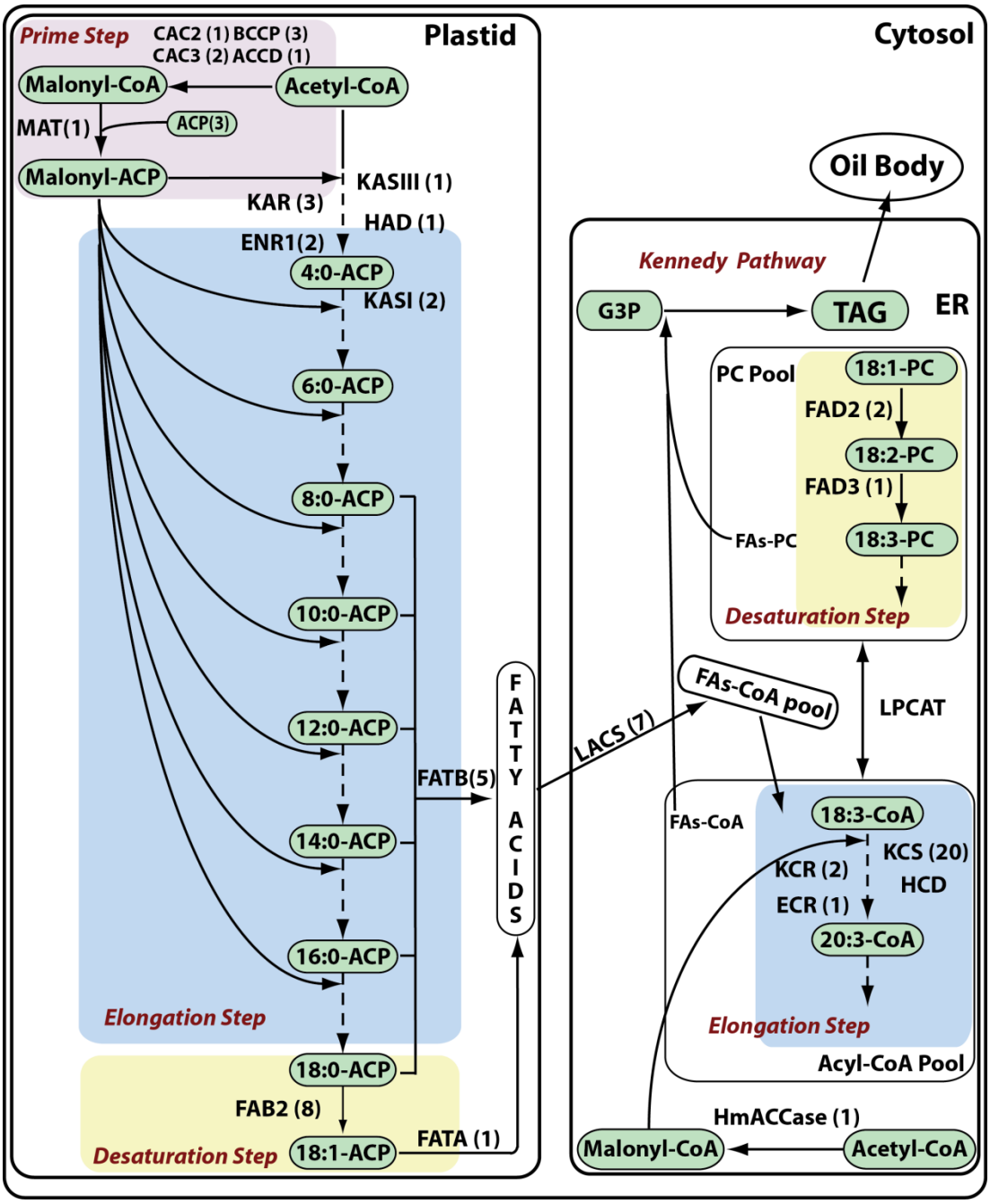


Figure 1-1 General fatty acid biosynthesis pathway in plants **(Argout et al., 2011).**

Different from the leafy phospholipids-rich fatty acid profile, the majority of the newly synthesized fatty acids are assembled in the form of triacylglycerols (TAGs) in developing seeds. TAG assembly, also termed the Kennedy Pathway (Figure 1-2), begins

with glycerol-3-phosphate (G3P) as the glycerol backbone, which is derived from dihydroxyacetone phosphate (DHAP) in the cytosol by sn-glycerol-3-phosphate dehydrogenease (G3PDH) [2]. Two consecutive acylations, from G3P through lysophosphatidic acid (LPA) to phosphatidic acid (PA), are catalyzed by acyl-CoA: sn-glycerol-3-phophate acyltransferase (GPAT) and acyl-CoA:lysophosphatidic acid acyltransferase (LPAAT), respectivively. Subsequently, phosphate groups are released by phosphatidic acid phosphatase (PAP) to produce sn-1,2-diacylglycerol (DAG). At the last step of acylation, acyl moieties from either acyl-CoA pools or acyl-PC pools are translocated on sn-3 position of glycerol backbone catalyzed by diacylglycerol transferase (DGAT) and phospholipid:diacylglycerol acyltransferase (PDAT), respectively. Newly assembled TAGs are extruded from ER into the cytosol in a form of a droplet in oil body.

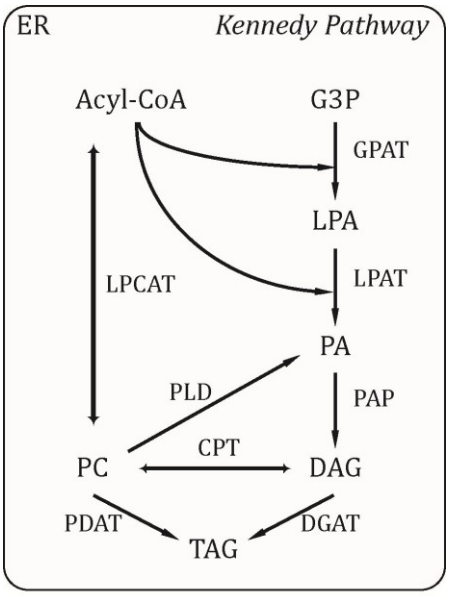


Figure 1-2 A simplified Kennedy Pathway in the ER, **adapted from (Chapman and Ohlrogge, 2012).**

1.3.3 Genomics of Cocoa Butter Biosynthesis

As mentioned above, cacao seeds contain about 50% total lipids (cocoa butter), and TAGs are the main component of cocoa butter (Crozier et al., 2011). In the first draft of the cacao genome, many key gene families involved in fatty acid biosynthesis pathway were

identified and functionally annotated in the genome, one of my main contributions to this work (Argout et al., 2011). Generally, the expansion of several gene families was observed in the fatty acid biosynthesis pathway compared to the Arabidopsis genome, such as *FATB* and *LACS* gene families, which might reflect the high requirement of oil production machinery in cacao seeds. In addition to the high lipid content, another important property of cocoa butter for industrial uses is its unique fatty acid composition, which consists of almost equal amounts of 16:0, 18:0 and $18:1^{n-9}$. Remarkably, during the 20 weeks of cacao seed development and maturation, dramatic alterations of fatty acid profiles occur within only 20 days when a polyunsaturated fatty acid $18:2^{n-6}$ and $18:3^{n-3}$-rich leafy membrane-like profile transitions to a final fatty acid profile rich in saturated fatty acid 16:0 and 18:0, and monounsaturated fatty acid $18:1^{n-9}$ (Patel et al., 1994). Moreover, as the dominant form of unsaturated fatty acid, $18:1^{n-9}$ gradually accumulates in developing cacao seeds, resulting in a significant impact on the physical properties of cocoa butter owing to its much lower melting temperature (16°C) compared to the other saturated fatty acids (16:0 – 62°C; 18:0 – 68°C) (Kachroo et al., 2007). Therefore, the ratio between $18:1^{n-9}$ and 16:0 + 18:0 has a significant impact on the final melting temperature of cocoa butter. Furthermore, positional analysis conducted by HPLC revealed that the composition of TAGs in cocoa butter has the monounsaturated 18:1 predominately occurring on the sn-2 position, yielding of 2-oleyl-glycerides of palmitic acid (16:0, P) and stearic acid (18:0, S) as POP, SOS, and POS (Takagi and Ando, 1995). Since sn-1 and sn-3 positions on glycerol are usually interchangeable and hard to distinguish, 18:1 at sn-2 position is of special interest because it provides critical crystallization properties of cocoa butter during chocolate manufacturing (Lipp and Anklam, 1998).

1.3.4 Stearoyl-acyl carrier protein-desaturase (SAD) Gene Family in Plants

In plants, stearoyl-acyl carrier protein (ACP)-desaturase (SAD, EC 1.14.99.6) is a nuclear-encoded, plastid-localized soluble desaturase that introduces the first Δ9 double bond into the saturated fatty acid resulting in the conversion of 18:0-ACP into $18:1^{n-9}$-ACP (Fox et al., 1993). As the dominant form of monounsaturated fatty acid exported from the plastid, $18:1^{n-9}$ can be further desaturated into polyunsaturated fatty acid derivatives, such as $18:2^{n-6}$ and $18:3^{n-3}$, both of which serve as major components of cell membrane systems in

the form of phospholipids (Ohlrogge and Browse, 1995). In this respect, given the fact that most plants lack other desaturases that utilize 18:0 as the substrate, the activity of SAD is of particular interest because of its significant effects on the ratio of saturated fatty acids to unsaturated fatty acids (Lindqvist et al., 1996), and the effects on the fluidity and rigidity of membrane system and the relationship of this to the adaption of plants to various environmental conditions. For example, sunflower, flax, and castor grown in lower temperature conditions contained higher proportions of polyunsaturated fatty acid compared to those in high temperature growing regions (Green, 1986; Garces et al., 1992).

Given the functional importance of fatty acid saturation in plant development and industrial application, SAD genes from many plant species have been identified and characterized and high correlations between the activities of SADs and levels of 18:0 and $18:1^{n-9}$ have been widely observed (Knutzon et al., 1992; Nishida et al., 1992; Slocombe et al., 1992; Gibson, 1993; Tong et al., 2006; Schluter et al., 2011; Shilman et al., 2011). For instance, a mutation of the Arabidopsis AtSSI2 gene, resulting in a single amino acid substitution resulted in an elevated level of 18:0 in Arabidopsis leaves (Lightner et al., 1994). Likewise, antisense-mediated reduction of SAD enzymatic activity in Brassica led to a dramatically increased level of 18:0 in the mature seed oil, from 2% up to 40% (Knutzon et al., 1992). Likewise, in soybean, a high 18:0 phenotype was strongly correlated with the activity of one of the SAD isozymes SACPD-C, an observation useful for screening of soybean varieties for unique seed oil profiles (Zhang et al., 2008).

In Arabidopsis, the family of SAD exhibits a high degree of amino acid sequence similarity (>70%) in spite of their minor functional differences (Kachroo et al., 2007). The three-dimensional crystal structure of homodimeric SAD protein from castor seed revealed that each SAD monomer comprises 11 conserved α-helices, which are crucial for ligand binding affinity, substrate chain-length selectivity, and double bond insertion position (Fox et al., 1993; Lindqvist et al., 1996; Cahoon et al., 1997). More impressively, replacements of several key amino acid residues in those conserved functional domains were sufficient to significantly alter the preferential mode of SAD. For example, replacement of five amino acids (T181A/F200A/N205S/T206L/A207G) of a castor Δ^9 SAD resulted in a new specificity as a Δ^6-16:0-ACP desaturase (Cahoon et al., 1997). Similarly, mutation of three

amino acids (T117R/G188L/D280K) in castor Δ^9 SAD enables it convert stearoyl-ACP to the allylic alcohol (*E*)-10-18:1-9-OH instead of 18:1^{n-9} (Whittle et al., 2008). Therefore, it is feasible to precisely manipulate activity of SAD taking advantage of the in-depth knowledge of its functional mechanism and advanced genetic approaches.

1.4 Fatty Acid Derived Defense Signals and Defense Response

1.4.1 Fatty Acid Derived Signals in Plant Defense

Profoundly, in addition to the significant roles of 18:1 as the major component of storage lipids and as the precursor for membrane fatty acid biosynthesis, levels of 18:1 and activities of SADs are also involved in regulation of the defense response in many plant species via the fatty acid-derived signaling pathway (Kachroo and Kachroo, 2009). In Arabidopsis, a reduction level of 18:1 in the *ssi2* mutant induced the SA mediated defense pathway, resulting in constitutive expression of *PR* genes, activation of the hypersensitive response and enhancement of broad-spectrum resistance to bacterial and oomycete pathogens (Shah et al., 2001; Kachroo et al., 2003; Kachroo et al., 2004). Similar defense responses were also observed in soybean and rice (Kachroo et al., 2008; Jiang et al., 2009), showing that silenced *SAD* activities were sufficient to induce the same type of defense-signaling pathway and enhance resistance to multiple pathogens. Together, it appears that 18:1 derived defense signaling pathway is conserved among many plant species.

1.4.2 Effects of Glycerol and Glycerol-3-phosphate on Defense Pathway

Glycerol is an environmentally friendly, non-toxic, edible and biodegradable sugar alcohol. As a byproduct of the booming biodiesel industry, large amounts of low-cost glycerol are expected to be produced in the coming years (Yang et al., 2012) and thus the utility of glycerol in agricultural applications has been previously explored as a potential added-value byproduct of biofuels production (Tisserat and Stuff, 2011). The application of glycerol in foliar sprays was found to stimulate short-term plant growth in diverse plant species under greenhouse conditions; however the mechanism of this enhancement was not reported.

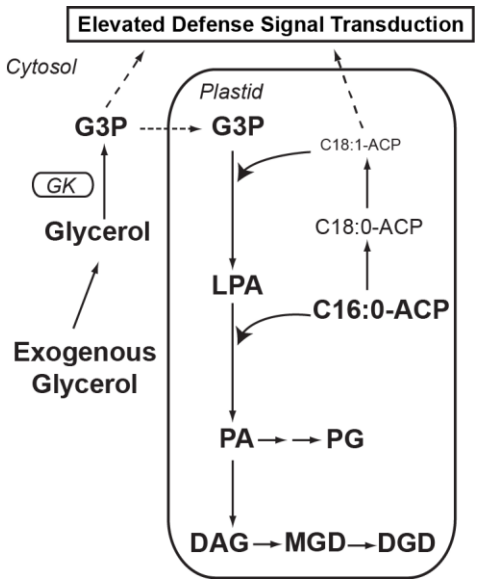


Figure 1-3 Abbreviated schemes of glycerol metabolism and glycerolipid biosynthesis pathway **(adapted from (Chanda et al., 2008))**

One physiological effect of exogenous application of glycerol is the elevation of the endogenous level of G3P via a glycerol kinase mediated reaction (Lu et al., 2001; Eastmond, 2004; Kachroo et al., 2004; Chandra-Shekara et al., 2007; Kachroo et al., 2008). As a precursor and backbone for the glycerolipid biosynthesis pathway, the plastidial level of G3P determines the carbon flux into the pathway, and thus regulates the level of oleic acid (18:1) through acylation (Figure 1-3). In this aspect, the exogenous application of glycerol is able to lower the level of 18:1 and thus mimic the *ssi2* mutant related phenotypes, such as reactive oxygen species (Campos et al., 2007) generation, *PR* gene expression, and the hypersensitive response (Chandra-Shekara et al., 2007). Additionally, the level of G3P was demonstrated to be closely associated with activation of the plant immune system as well. For example, the basal resistance of Arabidopsis to fungus pathogen *Colletotrichum higginsianum* (Chanda et al., 2008) was shown to be induced by G3P. Furthermore, G3P was recognized as a novel mobile signal that is able to translocate from locally infected tissues to distal tissues to induce systemic acquired resistance (SAR) against a broad spectrum of pathogens upon secondary infection (Chanda et al., 2011). Therefore, taken together, glycerol might participate in the defense responses through a fatty acid-derived

signaling pathway and might be also involved in induction of SAR through the elevation of G3P.

1.4.3 Cacao Diseases and Disease Control in Cacao

The main constraint on cacao cultivation is a range of destructive pathogenic diseases, which result in about 30-40% annual world production loss. Among them, three most severe diseases, black pod, frosty pod, and witches' broom, are caused by *Phytophthora spp*, *Moniliophthora roreri*, and *Moniliophthora perniciosa*, respectively, which have been described as the "trilogy of crippling fungal disease" (Fulton, 1989; Evans, 2007). Black pod disease spreads in all cocoa growing regions with varied degrees of aggressiveness. For example, *Phytophthora megakarya* is the most destructive pathogen in West Africa, which may cause 60-100% yield loss; *Phytophthora palmivora* is less severe but has more global spread, resulting in about 25% yield loss; *Phytophthora capsici* only exist in America (Flament et al., 2001; Risterucci et al., 2003). *Moniliophthora roreri* and *perniciosa* are both hemibiotrophic pathogens (Mondego et al., 2008). Cacao is the only host of *M. roreri* to our current knowledge while *M. perniciosa* is known to infect several crop species (Lanaud et al., 2009; Meinhardt et al., 2014). Witches' broom is the most devastating disease in America, which infects all the meristematic tissues of cacao, including shoots, flowers, and pods (de Oliveira et al., 2012). Recently, genomes of *M. roreri* and *perniciosa* have been fully sequenced (Meinhardt et al., 2014), and sequencing projects of *P. megakarya* and *palmivora* are in progress. Identification of responsible secreted peptides during infection process allows us to better elucidate the infection mechanism and seek for a solution for cacao disease control.

Foliar sprays (insecticides and fungicides) have been commonly applied as conventional and effective tools for disease management of many crop species (Dann et al., 1998; Reuveni et al., 1998; Buck et al., 2008). In cacao, copper based fungicides and insecticides like organochlorides are sprayed to cacao foliage against *Moniliophthora*, *Phytophthora spp*, and mirids (Sosan and Akingbohungbe, 2009; Okoya et al., 2013). However, there is a concern about environmental consequences of heavy metal fungicides and potential risks to human health due to mishandling and carryover effects of pesticide residues (Okoya et al., 2013). In this aspect, endophytes, which refer to non-pathogenic

microorganisms that live asymptomatically within intra- and intercellular spaces of plant tissues, exhibit a great potential as biosafety biocontrol agents to enhance host defense against pathogens and herbivores (Mejía et al., 2014). Symbiontic endophytes are well known for their beneficial effects on promoting plant growth, such as N_2 fixing, synthesis of enzymes to provide nutrient availability, and enhancing tolerance to biotic and abiotic stress (Hanada et al., 2010). Interestingly, recent studies also suggested that after penetrating and colonization in the tissues, endophytes are capable of inducing systemic resistance in plants against pathogenic pathogens. For instance, *Colletotrichum tropicale* is a dominant endophytic fungus isolated from healthy cacao leaves. In both greenhouse study and field trials, cacao endophyte-free plants inoculated with *C. tropicale* exhibited higher resistance against *Phytophthora palmivora* and milder symptoms of black pod diseases (Mejía et al., 2014). It was further revealed that many defense related genes were induced by the inoculation of *C. tropicale*, and among them, a highly *C. tropicale* inducing gene *Tc00g04254* (one putative trypsin inhibitor) was able to confer higher resistance against *Phytophthora capsici* when overexpressed in cacao leaves in the absence of *C. tropicale* (Mejía et al., 2014). Together, employment of endophytes as biocontrol agents appears to be a promising approach to enhance disease resistance of cacao plants.

In addition to the above passive prevention strategies to control diseases, tremendous efforts have been made to construct QTL maps to identify the regions on the genome associated with the tolerance to the pathogens. For instance, 256 F1 progenies from the cross between “Pound7”and “UF273” were included in one study, aiming to map for the resistance QTL against frosty pod and black pod diseases (Brown et al., 2007). Consequently, five QTLs for frosty pod on chromosomes 2, 7, and 8, and three QTLs for black pod on chromosomes 4, 8, and 10 were found with favorable alleles from UF273 and Pound7, respectively. Currently, at least 76 QTL maps were constructed for various cacao diseases (Lanaud et al., 2009). However, it is problematic to merge all these maps into a single genetic map since most of these mapping trails utilized different cacao varieties from different generations (F1, F2, back cross, etc) and the resistance traits were evaluated by different parameters, and different genetic markers were applied. Profoundly, through a comparative mapping strategy, a consensus QTL map was constructed and several “hot

spots" associated with resistance were observed (Lanaud et al., 2009). Together, taking advantage of this QTL map, multiple genetic resistance sources could be accumulated in the same variety through breeding programs.

In the genome era of cacao research, together with several other genomic resources, such as EST database and microarray analysis on pathogen infection, more and more attention has been paid to identify key defense related genes. To further facilitate the evaluation of functions of genes of interests, stable transgenic plants and a rapid transient gene expression and leaf pathogen assay were developed in our lab. For example, the function of a class I chitinase gene in cacao, *TcChi1*, was previously characterized in the defense response against *Colletotrichum gloeosporioides* (Maximova et al., 2006). The stable transgenic cacao plant overexpressing *TcChi1* exhibited higher resistance against the fungal pathogen. Consistent result was also obtained when *TcChi1* was transiently overexpressed in cacao stage C leaves and conferred higher resistant against fungal pathogens compared to control plants. Likewise, *NPR* gene family in cacao was previously functional characterized by Shi (Shi et al., 2010), suggesting that a *NPR1*-dependent defense signaling pathway is conserved between Arabidopsis and cacao. Transient overexpression of *TcNPR1* allow us rapidly confirm its key roles involved in the defense responses against *Phytophthora capsici* in cacao leaves (unpublished data). Together, these studies demonstrated a great potential of combining genomics resources and transient gene expression approaches to rapidly identify and functionally characterize key genes on the defense pathway, assisting us to better elucidate the defense mechanism in cacao with the long term goal to accelerate breeding programs.

1.5 Conclusion

The research presented in my thesis aims to further our understanding of the somatic embryogenesis system and the key quality traits of cacao at genetic levels by identification and functional characterization of the key regulatory genes in cacao. Specifically, transcriptional network regulating somatic embryogenesis and fatty acid central metabolism in developing cacao seeds are the main focuses in this thesis. I

discovered the critical roles of a B3 domain transcription factor, TcLEC2, during both zygotic and somatic embryogenesis, which functions in concert with many other embryo specific genes regulating the embryogenic potential of somatic cells and the development of the embryos. I also functionally dissected the SAD gene family in cacao and provided evidence indicating that the activity of one SAD isoform, TcSAD1, which was proved to be functional equivalent to the orthologous gene in Arabidopsis, was highly correlated with the alteration of fatty acid profiles in developing cacao seeds. Additionally, our data also supported the potential roles of fatty acid derived signaling pathway in cacao defense responses. The endogenous level of oleic acid was lowered by the exogenous glycerol application, which could trigger the defense responses, activate the expression of defense related genes, and thus confer disease resistance against fungal pathogens.

Together, our results not only offer several key genes, such as *TcLEC2* and *TcSAD1*, as potential biomarkers useful for breeding programs, but also provide valuable knowledge in somatic embryogenesis, fatty acid biosynthesis, and defense responses of cacao. The details on each research project are presented in the following chapters.

1.6 References

Anonymous (2012) Cocoa Market Update. World Cocoa Foundation http://worldcocoafoundation.org/wp-content/uploads/Cocoa-Market-Update-as-of-3.20.2012.pdf

Argout X, Salse J, Aury JM, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN, Abrouk M, Murat F, Fouet O, Poulain J, Ruiz M, Roguet Y, Rodier-Goud M, Barbosa-Neto JF, Sabot F, Kudrna D, Ammiraju JS, Schuster SC, Carlson JE, Sallet E, Schiex T, Dievart A, Kramer M, Gelley L, Shi Z, Berard A, Viot C, Boccara M, Risterucci AM, Guignon V, Sabau X, Axtell MJ, Ma Z, Zhang Y, Brown S, Bourge M, Golser W, Song X, Clement D, Rivallan R, Tahi M, Akaza JM, Pitollat B, Gramacho K, D'Hont A, Brunel D, Infante D, Kebe I, Costet P, Wing R, McCombie WR, Guiderdoni E, Quetier F, Panaud O, Wincker P, Bocs S, Lanaud C (2011) The genome of Theobroma cacao. Nature Genetics **43:** 101-108

Baud S, Mendoza MS, To A, Harscoet E, Lepiniec L, Dubreucq B (2007) WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in Arabidopsis. Plant Journal **50:** 825-838

Braybrook SA, Harada JJ (2008) LECs go crazy in embryo development. Trends Plant Sci **13:** 624-630

Braybrook SA, Stone SL, Park S, Bui AQ, Le BH, Fischer RL, Goldberg RB, Harada JJ (2006) Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. Proc Natl Acad Sci U S A **103:** 3468-3473

Brown JS, Phillips-Mora W, Power EJ, Krol C, Cervantes-Martinez C, Motamayor JC, Schnell RJ (2007) Mapping QTLs for Resistance to Frosty Pod and Black Pod Diseases and Horticultural Traits in Theobroma cacao L.. Crop Sci. **47:** 1851-1858

Brunetto MadR, Gutiérrez L, Delgado Y, Gallignani M, Zambrano A, Gómez Á, Ramos G, Romero C (2007) Determination of theobromine, theophylline and caffeine in cocoa samples by a high-performance liquid chromatographic method with on-line sample cleanup in a switching-column system. Food Chemistry **100:** 459-467

Buck GB, Korndörfer GH, Nolla A, Coelho L (2008) Potassium Silicate as Foliar Spray and Rice Blast Control. Journal of Plant Nutrition **31:** 231-237

Cahoon EB, Lindqvist Y, Schneider G, Shanklin J (1997) Redesign of soluble fatty acid desaturases from plants for altered substrate specificity and double bond position. Proc Natl Acad Sci U S A **94:** 4872-4877

Campos MA, Rosa DD, Teixeira JÉC, Targon MLPN, Souza AA, Paiva LV, Stach-Machado DR, Machado MA (2007) PR gene families of citrus: their organ specific-biotic and abiotic inducible expression profiles based on ESTs approach. Genetics and Molecular Biology **30:** 917-930

Chanda B, Venugopal SC, Kulshrestha S, Navarre DA, Downie B, Vaillancourt L, Kachroo A, Kachroo P (2008) Glycerol-3-Phosphate Levels Are Associated with Basal Resistance to the Hemibiotrophic Fungus Colletotrichum higginsianum in Arabidopsis. Plant Physiology **147:** 2017-2029

Chanda B, Xia Y, Mandal MK, Yu K, Sekine KT, Gao QM, Selote D, Hu Y, Stromberg A, Navarre D, Kachroo A, Kachroo P (2011) Glycerol-3-phosphate is a critical mobile inducer of systemic immunity in plants. Nature Genetics **43:** 421-427

Chandel KPS, Chaudhury R, Radhamani J, Malik SK (1995) Desiccation and Freezing Sensitivity in Recalcitrant Seeds of Tea, Cocoa and Jackfruit. Annals of Botany **76:** 443-450

Chandra-Shekara AC, Venugopal SC, Barman SR, Kachroo A, Kachroo P (2007) Plastidial fatty acid levels regulate resistance gene-dependent defense signaling in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America **104:** 7277-7282

Chapman KD, Ohlrogge JB (2012) Compartmentation of Triacylglycerol Accumulation in Plants. Journal of Biological Chemistry **287:** 2288-2294

Cheesman EE, Pound FJ (1932) Uniformity trials on cacao. Tropical Agriculture **Vol. IX:** 277-287

Crozier SJ, Preston AG, Hurst JW, Payne MJ, Mann J, Hainly L, Miller DL (2011) Cacao seeds are a "Super Fruit": A comparative analysis of various fruit powders and products. Chem Cent J **5:** 5

Dann E, Diers B, Byrum J, Hammerschmidt R (1998) Effect of treating soybean with 2,6-dichloroisonicotinic acid (INA) and benzothiadiazole (BTH) on seed yields and the level of disease caused by Sclerotinia sclerotiorum in field and greenhouse studies. European Journal of Plant Pathology **104:** 271-278

Davie JH (1935) Chromosome studies in the malvaceae and certain related families. II. Genetica **17:** 487-498

de Oliveira BV, Teixeira GS, Reis O, Barau JG, Teixeira PJ, do Rio MC, Domingues RR, Meinhardt LW, Paes Leme AF, Rincones J, Pereira GA (2012) A potential role for an extracellular methanol oxidase secreted by Moniliophthora perniciosa in Witches' broom disease in cacao. Fungal Genet Biol **49:** 922-932

Eastmond PJ (2004) Glycerol-insensitive Arabidopsis mutants: gli1 seedlings lack glycerol kinase, accumulate glycerol and are more resistant to abiotic stress. Plant Journal **37:** 617-625

Evans HC (2007) Cacao diseases - The trilogy revisited. Phytopathology **97:** 1640-1643

Feeney M, Frigerio L, Cui Y, Menassa R (2013) Following vegetative to embryonic cellular changes in leaves of Arabidopsis overexpressing LEAFY COTYLEDON2. Plant Physiology **162:** 1881-1896

Flament MH, Kebe I, Clément D, Pieretti I, Risterucci AM, N'Goran JAK, Cilas C, Despréaux D, Lanaud C (2001) Genetic mapping of resistance factors to Phytophthora palmivora in cocoa. Genome **44:** 79-85

Fox BG, Shanklin J, Somerville C, Munck E (1993) Stearoyl-acyl carrier protein delta 9 desaturase from Ricinus communis is a diiron-oxo protein. Proc Natl Acad Sci U S A **90:** 2486-2490

Frauendorfer F, Schieberle P (2008) Changes in Key Aroma Compounds of Criollo Cocoa Beans During Roasting. Journal of Agricultural and Food Chemistry **56:** 10244-10251

Fujimura T, Komamine A (1979) Involvement of endogenous auxin in somatic embryogenesis in a carrot cell-suspension culture. Zeitschrift Fur Pflanzenphysiologie **95:** 13-19

Fulton R (1989) The cacao disease trilogy: black pod, Monilia pod rot, and witches'-broom. Plant Disease**:** 601

Gaj MD, Zhang S, Harada JJ, Lemaux PG (2005) Leafy cotyledon genes are essential for induction of somatic embryogenesis of Arabidopsis. Planta **222:** 977-988

Garces R, Sarmiento C, Mancha M (1992) Temperature regulation of oleate desaturase in sunflower (Helianthus annuus L.) seeds. Planta **186:** 461-465

Gibson KJ (1993) Palmitoleate formation by soybean stearoyl-acyl carrier protein desaturase. Biochim Biophys Acta **1169:** 231-235

Glendinning DR (1972) NATURAL POLLINATION OF COCOA. New Phytologist **71:** 719-729

Gliwicka M, Nowak K, Balazadeh S, Mueller-Roeber B, Gaj MD (2013) Extensive modulation of the transcription factor transcriptome during somatic embryogenesis in Arabidopsis thaliana. PLoS One **8:** e69261

Green AG (1986) Genetic control of polyunsaturated fatty acid biosynthesis in flax (Linum usitatissimum) seed oil. Theor Appl Genet **72:** 654-661

Guo FD, Liu CL, Xia H, Bi YP, Zhao CZ, Zhao SZ, Hou L, Li FG, Wang XJ (2013) Induced Expression of AtLEC1 and AtLEC2 Differentially Promotes Somatic Embryogenesis in Transgenic Tobacco Plants. PLoS One **8**

Hanada RE, Pomella AW, Costa HS, Bezerra JL, Loguercio LL, Pereira JO (2010) Endophytic fungal diversity in Theobroma cacao (cacao) and T. grandiflorum (cupuacu) trees and their potential for growth promotion and biocontrol of black-pod disease. Fungal Biol **114:** 901-910

Horváth I, Glatz A, Nakamoto H, Mishkind ML, Munnik T, Saidi Y, Goloubinoff P, Harwood JL, Vigh L (2012) Heat shock response in photosynthetic organisms: Membrane and lipid connections. Progress in Lipid Research **51:** 208-220

I. Andújar MCR, R. M. Giner, and J. L. Ríos (2012) Cocoa Polyphenols and Their Potential Benefits for Human Health. Oxidative Medicine and Cellular Longevity **2012:** 23

Irizarry H, Goenaga R (2000) Clonal selection in cacao based on early yield performance of grafted trees. J. Agric. Univ. P.R. **8:** 153-163

Irizarry H, Rivera E (1999) Early yield of five cacao families at three locations in Puerto Rico. J. Agric. Univ. P.R. **82:** 167-176

Jiang CJ, Shimono M, Maeda S, Inoue H, Mori M, Hasegawa M, Sugano S, Takatsuji H (2009) Suppression of the rice fatty-acid desaturase gene OsSSI2 enhances resistance to blast and leaf blight diseases in rice. Mol Plant Microbe Interact **22:** 820-829

Kachroo A, Fu DQ, Havens W, Navarre D, Kachroo P, Ghabrial SA (2008) An oleic acid-mediated pathway induces constitutive defense signaling and enhanced resistance to multiple pathogens in soybean. Mol Plant Microbe Interact **21:** 564-575

Kachroo A, Fu DQ, Havens W, Navarre D, Kachroo P, Ghabrial SA (2008) An oleic acid-mediated pathway induces constitutive defense signaling and enhanced resistance to multiple pathogens in soybean. Molecular Plant-Microbe Interactions **21:** 564-575

Kachroo A, Kachroo P (2009) Fatty Acid-derived signals in plant defense. Annu Rev Phytopathol **47:** 153-176

Kachroo A, Lapchyk L, Fukushige H, Hildebrand D, Klessig D, Kachroo P (2003) Plastidial fatty acid signaling modulates salicylic acid- and jasmonic acid-mediated defense pathways in the Arabidopsis ssi2 mutant. Plant Cell **15:** 2952-2965

Kachroo A, Shanklin J, Whittle E, Lapchyk L, Hildebrand D, Kachroo P (2007) The Arabidopsis stearoyl-acyl carrier protein-desaturase family and the contribution of leaf isoforms to oleic acid synthesis. Plant Mol Biol **63:** 257-271

Kachroo A, Venugopal SC, Lapchyk L, Falcone D, Hildebrand D, Kachroo P (2004) Oleic acid levels regulated by glycerolipid metabolism modulate defense gene expression in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America **101:** 5152-5157

Karami O, Aghavaisi B, Mahmoudi Pour A (2009) Molecular aspects of somatic-to-embryogenic transition in plants. Journal of chemical biology **2:** 177-190

Kaufmann T (1975) Ecology and Behavior of Cocoa Pollinating Ceratopogonidae in Ghana, W. Africa. Environmental Entomology **4:** 347-351

Knutzon DS, Thompson GA, Radke SE, Johnson WB, Knauf VC, Kridl JC (1992) Modification of Brassica seed oil by antisense expression of a stearoyl-acyl carrier protein desaturase gene. Proc Natl Acad Sci U S A **89:** 2624-2628

Kurczynska EU, Gaj MD, Ujczak A, Mazur E (2007) Histological analysis of direct somatic embryogenesis in Arabidopsis thaliana (L.) Heynh. Planta **226:** 619-628

Lanaud C, Fouet O, Clément D, Boccara M, Risterucci AM, Surujdeo-Maharaj S, Legavre T, Argout X (2009) A meta–QTL analysis of disease resistance traits of Theobroma cacao L. Molecular Breeding **24:** 361-374

Ledwon A, Gaj MD (2009) LEAFY COTYLEDON2 gene expression and auxin treatment in relation to embryogenic capacity of Arabidopsis somatic cells. Plant Cell Reports **28:** 1677-1688

Lehrian DW, Keeney PG (1980) Changes in lipid components of seeds during growth and ripening of cacao fruit. Journal of the American Oil Chemists' Society **57:** 61-65

Li ZJ, Traore A, Maximova S, Guiltinan MJ (1998) Somatic embryogenesis and plant regeneration from floral explants of cacao (Theobroma cacao L.) using thidiazuron. In Vitro Cellular & Developmental Biology-Plant **34:** 293-299

Liendo R, Padilla FC, Quintana A (1997) Characterization of cocoa butter extracted from Criollo cultivars of Theobroma cacao L. Food Research International **30:** 727-731

Lightner J, Wu JR, Browse J (1994) A Mutant of Arabidopsis with Increased Levels of Stearic-Acid. Plant Physiology **106:** 1443-1451

Lindqvist Y, Huang W, Schneider G, Shanklin J (1996) Crystal structure of delta9 stearoyl-acyl carrier protein desaturase from castor seed and its relationship to other di-iron proteins. EMBO J **15:** 4081-4092

Lipp M, Anklam E (1998) Review of cocoa butter and alternative fats for use in chocolate—Part A. Compositional data. Food Chemistry **62:** 73-97

Lopez Baez O, Bollon H, Eskes A, Petiard V (1993) Somatic embryogenesis and plant-regeneration from flower parts of cocoa Theobroma-cacao L. Comptes Rendus De L Academie Des Sciences Serie Iii-Sciences De La Vie-Life Sciences **316:** 579-584

Lotan T, Ohto M, Yee KM, West MA, Lo R, Kwong RW, Yamagishi K, Fischer RL, Goldberg RB, Harada JJ (1998) Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. Cell **93:** 1195-1205

Lu M, Tang XY, Zhou JM (2001) Arabidopsis NHO1 is required for general resistance against Pseudomonas bacteria. Plant Cell **13:** 437-447

Luerssen H, Kirik V, Herrmann P, Misera S (1998) FUSCA3 encodes a protein with a conserved VP1/AB13-like B3 domain which is of functional importance for the regulation of seed maturation in Arabidopsis thaliana. Plant Journal **15:** 755-764

Matsuta N (1992) Effect of auxin on somatic embryogenesis from leaf callus in Grape (Vitis-Spp). Japanese Journal of Breeding **42:** 879-883

Maximova S, Marelli J-P, Young A, Pishak S, Verica J, Guiltinan M (2006) Over-expression of a cacao class I chitinase gene in Theobroma cacao L. enhances resistance against the pathogen, Colletotrichum gloeosporioides. Planta **224:** 740-749

Maximova S, Young A, Pishak S, Guiltinan M (2008) Field performance of Theobroma cacao L. plants propagated via somatic embryogenesis. In Vitro Cellular & Developmental Biology - Plant **44:** 487-493

Maximova SN, Alemanno L, Young A, Ferriere N, Traore A, Guiltinan MJ (2002) Efficiency, genotypic variability, and cellular origin of primary and secondary somatic embryogenesis of Theobroma cacao L. In Vitro Cellular & Developmental Biology-Plant **38:** 252-259

Meinhardt L, Costa GG, Thomazella D, Teixeira PJ, Carazzolle M, Schuster S, Carlson J, Guiltinan M, Mieczkowski P, Farmer A, Ramaraj T, Crozier J, Davis R, Shao J, Melnick R, Pereira G, Bailey B (2014) Genome and secretome analysis of the hemibiotrophic fungal pathogen, Moniliophthora roreri, which causes frosty pod rot disease of cacao: mechanisms of the biotrophic and necrotrophic phases. BMC Genomics **15:** 164

Mejía LC, Herre EA, Sparks JP, Winter K, García MN, Van Bael SA, Stitt J, Shi Z, Zhang Y, Guiltinan MJ, Maximova SN (2014) Pervasive effects of a dominant foliar

endophytic fungus on host genetic and phenotypic expression in a tropical tree. Frontiers in Microbiology **5**

Mondego JM, Carazzolle MF, Costa GG, Formighieri EF, Parizzi LP, Rincones J, Cotomacci C, Carraro DM, Cunha AF, Carrer H, Vidal RO, Estrela RC, Garcia O, Thomazella DPT, de Oliveira BV, Pires ABL, Rio MCS, Araujo MRR, de Moraes MH, Castro LAB, Gramacho KP, Goncalves MS, Neto JPM, Neto AG, Barbosa LV, Guiltinan MJ, Bailey BA, Meinhardt LW, Cascardo JCM, Pereira GAG (2008) A genome survey of Moniliophthora perniciosa gives new insights into Witches' Broom disease of cacao. Bmc Genomics **9:** 548

Motamayor J, Mockaitis K, Schmutz J, Haiminen N, Iii D, Cornejo O, Findley S, Zheng P, Utro F, Royaert S, Saski C, Jenkins J, Podicheti R, Zhao M, Scheffler B, Stack J, Feltus F, Mustiga G, Amores F, Phillips W, Marelli J, May G, Shapiro H, Ma J, Bustamante C, Schnell R, Main D, Gilbert D, Parida L, Kuhn D (2013) The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. Genome Biology **14:** r53

Nishida I, Beppu T, Matsuo T, Murata N (1992) Nucleotide sequence of a cDNA clone encoding a precursor to stearoyl-(acyl-carrier-protein) desaturase from spinach, Spinacia oleracea. Plant Mol Biol **19:** 711-713

Ohlrogge J, Browse J (1995) Lipid Biosynthesis. Plant Cell **7:** 957-970

Okoya AA, Ogunfowokan AO, Asubiojo OI, Torto N (2013) Organochlorine Pesticide Residues in Sediments and Waters from Cocoa Producing Areas of Ondo State, Southwestern Nigeria. ISRN Soil Science **2013:** 12

Patel VK, Shanklin J, Furtek DB (1994) Changes in fatty-acid composition and stearoyl-acyl carrier protein desaturase expression in developing Theobroma-cacao L embryos. Planta **193:** 83-88

Pence VC, Hasegawa PM, Janick J (1979) Asexual embryogenesis in Theobroma-cacao L. Journal of the American Society for Horticultural Science **104:** 145-148

Reuveni R, Dor G, Reuveni M (1998) Local and systemic control of powdery mildew (Leveillula taurica) on pepper plants by foliar spray of mono-potassium phosphate. Crop Protection **17:** 703-709

Risterucci AM, Paulin D, Ducamp M, N'Goran JAK, Lanaud C (2003) Identification of QTLs related to cocoa resistance to three species of Phytophthora. Theoretical and Applied Genetics **108:** 168-174

Ross JA, Kasum CM (2002) DIETARY FLAVONOIDS: Bioavailability, Metabolic Effects, and Safety. Annual Review of Nutrition **22:** 19-34

Rusconi M, Conti A (2010) Theobroma cacao L., the Food of the Gods: a scientific approach beyond myths and claims. Pharmacol Res **61:** 5-13

Rusconi M, Conti A (2010) Theobroma cacao L., the Food of the Gods: A scientific approach beyond myths and claims. Pharmacological Research **61:** 5-13

Santos-Mendoza M, Dubreucq B, Baud S, Parcy F, Caboche M, Lepiniec L (2008) Deciphering gene regulatory networks that control seed development and maturation in Arabidopsis. The Plant Journal **54:** 608-620

Santos Mendoza M, Dubreucq B, Miquel M, Caboche M, Lepiniec L (2005) LEAFY COTYLEDON 2 activation is sufficient to trigger the accumulation of oil and seed specific mRNAs in Arabidopsis leaves. FEBS Letters **579:** 4666-4670

Schluter PM, Xu S, Gagliardini V, Whittle E, Shanklin J, Grossniklaus U, Schiestl FP (2011) Stearoyl-acyl carrier protein desaturases are associated with floral isolation in sexually deceptive orchids. Proc Natl Acad Sci U S A **108:** 5696-5701

Sereno M, Albuquerque PB, Vencovsky R, Figueira A (2006) Genetic Diversity and Natural Population Structure of Cacao (Theobroma cacao L.) from the Brazilian Amazon Evaluated by Microsatellite Markers. Conservation Genetics **7:** 13-24

Shah J, Kachroo P, Nandi A, Klessig DF (2001) A recessive mutation in the Arabidopsis SSI2 gene confers SA- and NPR1-independent expression of PR genes and resistance against bacterial and oomycete pathogens. Plant Journal **25:** 563-574

Shi Z, Maximova S, Liu Y, Verica J, Guiltinan M (2010) Functional analysis of the Theobroma cacao NPR1 gene in arabidopsis. BMC Plant Biology **10:** 248

Shilman F, Brand Y, Brand A, Hedvat I, Hovav R (2011) Identification and Molecular Characterization of Homeologous Delta 9-Stearoyl Acyl Carrier Protein Desaturase 3 Genes from the Allotetraploid Peanut (Arachis hypogaea). Plant Molecular Biology Reporter **29:** 232-241

Slocombe SP, Cummins I, Jarvis RP, Murphy DJ (1992) Nucleotide sequence and temporal regulation of a seed-specific Brassica napus cDNA encoding a stearoyl-acyl carrier protein (ACP) desaturase. Plant Mol Biol **20:** 151-155

Sosan MB, Akingbohungbe AE (2009) Occupational Insecticide Exposure and Perception of Safety Measures among Cacao Farmers in Southwestern Nigeria. Archives of Environmental & Occupational Health **64:** 185-193

Stone SL, Braybrook SA, Paula SL, Kwong LW, Meuser J, Pelletier J, Hsieh TF, Fischer RL, Goldberg RB, Harada JJ (2008) Arabidopsis LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis. Proc Natl Acad Sci U S A **105:** 3151-3156

Stone SL, Kwong LW, Yee KM, Pelletier J, Lepiniec L, Fischer RL, Goldberg RB, Harada JJ (2001) LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. Proc Natl Acad Sci U S A **98:** 11806-11811

Sukha D, Butler D, Umaharan P, Boult E (2008) The use of an optimised organoleptic assessment protocol to describe and quantify different flavour attributes of cocoa liquors made from Ghana and Trinitario beans. European Food Research and Technology **226:** 405-413

Takagi T, Ando Y (1995) Stereospecific analysis of monounsaturated triacylglycerols in cocoa butter. Journal of the American Oil Chemists' Society **72:** 1203-1206

Theodoulou FL, Eastmond PJ (2012) Seed storage oil catabolism: a story of give and take. Current Opinion in Plant Biology **15:** 322-328

Tisserat B, Stuff A (2011) Stimulation of Short-term Plant Growth by Glycerol Applied as Foliar Sprays and Drenches under Greenhouse Conditions. Hortscience **46:** 1650-1654

Tong L, Shu-Ming P, Wu-Yuan D, Dan-Wei M, Ying X, Meng X, Fang C (2006) Characterization of a new stearoyl-acyl carrier protein desaturase gene from Jatropha curcas. Biotechnology Letters **28:** 657-662

Visser C, Qureshi JA, Gill R, Saxena PK (1992) Morphoregulatory role of thidiazuron - substitution of auxin and cytokinin requirement for the induction of somatic embryogenesis in geranium hypocotyl cultures. Plant Physiology **99:** 1704-1707

Wang C, Zien CA, Afitlhile M, Welti R, Hildebrand DF, Wang X (2000) Involvement of phospholipase D in wound-induced accumulation of jasmonic acid in Arabidopsis. The Plant Cell Online **12:** 2237-2246

Whittle EJ, Tremblay AE, Buist PH, Shanklin J (2008) Revealing the catalytic potential of an acyl-ACP desaturase: tandem selective oxidation of saturated fatty acids. Proc Natl Acad Sci U S A **105:** 14738-14743

Wojcikowska B, Jaskola K, Gasiorek P, Meus M, Nowak K, Gaj MD (2013) LEAFY COTYLEDON2 (LEC2) promotes embryogenic induction in somatic tissues of Arabidopsis, via YUCCA-mediated auxin biosynthesis. Planta **238:** 425-440

Yang F, Hanna MA, Sun R (2012) Value-added uses for crude glycerol--a byproduct of biodiesel production. Biotechnol Biofuels **5:** 13

Yang X, Zhang X (2010) Regulation of Somatic Embryogenesis in Higher Plants. Critical Reviews in Plant Sciences **29:** 36-57

Yao LH, Jiang YM, Shi J, TomÁS-BarberÁN FA, Datta N, Singanusong R, Chen SS (2004) Flavonoids in Food and Their Health Benefits. Plant Foods for Human Nutrition **59:** 113-122

Zhang P, Burton JW, Upchurch RG, Whittle E, Shanklin J, Dewey RE (2008) Mutations in a Delta(9)-Stearoyl-ACP-Desaturase Gene Are Associated with Enhanced Stearic Acid Levels in Soybean Seeds. Crop Science **48:** 2305-2313

Zhu X-G, Long SP, Ort DR (2010) Improving Photosynthetic Efficiency for Greater Yield. Annual Review of Plant Biology **61:** 235-261

Zuo J, Niu QW, Ikeda Y, Chua NH (2002) Marker-free transformation: increasing transformation frequency by the use of regeneration-promoting genes. Curr Opin Biotechnol **13:** 173-180

CHAPTER 2: THE THEOBROMA CACAO B3 DOMAIN TRANSCRIPTION FACTOR TCLEC2 PLAYS A DUAL ROLE IN CONTROL OF EMBRYO DEVELOPMENT AND MATURATION

Yufan Zhang[a, b], Adam Clemens[a], Siela Maximova[a, b], and Mark Guiltinan[a,b,1]

[a]Huck Institute of Life Sciences, The Pennsylvania State University, University Park, PA 16802

[b]The Department of Plant Science, The Pennsylvania State University, University Park, PA 16802

[1]Corresponding author

2.1 Abstract

The *Arabidopsis thaliana LEC2* gene encodes a B3 domain transcription factor, which plays critical roles during both zygotic and somatic embryogenesis. LEC2 exerts significant impacts on determining embryogenic potential and various metabolic processes through a complicated genetic regulatory network. An ortholog of the Arabidopsis Leafy Cotyledon 2 gene *(AtLEC2)* was characterized in *Theobroma* cacao (*TcLEC2)*. *TcLEC2* encodes a B3 domain transcription factor preferentially expressed during early and late zygotic embryo development. The expression of *TcLEC2* was higher in dedifferentiated cells competent for somatic embryogenesis (embryogenic calli), compared to non-embryogenic calli. Transient overexpression of *TcLEC2* in immature zygotic embryos resulted in changes in gene expression profiles and fatty acid composition. Ectopic expression of *TcLEC2* in cacao leaves changed the expression levels of several seed related genes. The overexpression of *TcLEC2* in cacao explants greatly increased the frequency of regeneration of stably transformed somatic embryos. *TcLEC2* overexpressing cotyledon explants exhibited a very high level of embryogenic competency and when cultured on hormone free medium, exhibited an iterative embryogenic chain-reaction. Our study revealed essential roles of TcLEC2 during both zygotic and somatic embryo development. Collectively, our evidence

supports the conclusion that *TcLEC2* is a functional ortholog of *AtLEC2* and that it is involved in similar genetic regulatory networks during cacao somatic embryogenesis. To our knowledge, this is the first detailed report of the functional analysis of a *LEC2* ortholog in a species other then Arabidopsis. *TcLEC2* could potentially be used as a biomarker for the improvement of the SE process and screen for elite varieties in cacao germplasm.

2.2 Introduction

The current knowledge of somatic embryogenesis of cacao and the key transcription factors controlling this process in Arabidopsis was earlier reviewed in **Section 1.2**. In the present study, we identified a putative ortholog of *AtLEC2* in cacao, *TcLEC2*. We characterized the expression patterns of *TcLEC2* during both zygotic and somatic embryogenesis and explored the relationships between the activity of TcLEC2 in modulating the embryogenic potential of callus and in regulation of the fatty acid biosynthesis pathway.

2.3 Results

Gene isolation and sequence comparison

The Arabidopsis *AtLEC2* gene (At1G28300) is part of a large family of B3 domain containing proteins involved in a wide variety of functions. In the Arabidopsis genome, 87 genes were previously annotated as B3 domain containing genes that were further classified into five different families: auxin response factor (ARF), abscisic acid-insensitive3 (ABI3) of which *AtLEC2* is a member, high level expression of sugar inducible (HSI), related to ABI3/VP1 (RAV) and reproductive meristem (REM) (Romanel et al., 2009).

In order to identify a putative ortholog of *AtLEC2* in cacao, the full-length amino acid sequence of Arabidopsis *AtLEC2* was blasted against the predicted proteome of the Belizean Criollo genotype (B97-61/B2) (http://cocoagendb.cirad.fr/ (Argout et al., 2011)) using blastp algorithm with E-value cut-off of $1e^{-5}$ (Altschul et al., 1990), which resulted in identification of 13 possible candidate genes (Supplemental Figure 2-1). As a second

approach to identify cacao *LEC2* gene(s), the predicted protein sequences of each of the 13 candidate genes were used to search the predicted proteome from a second sequenced cacao genome of cv. Matina 1-6 v1.1 (http://www.cacaogenomedb.org (Motamayor et al., 2013)) by Blastp and a set of nearly identical cognate genes were identified for each (Supplemental Figure 2-1). No additional related genes were identified in this variety of cacao. Of the 13 candidate genes, the gene *Tc06g015590* resulted in the best alignment with *AtLEC2*, resulting in a blastp expect value of 3E-75.

To identify the most likely *AtLEC2*-orthologous gene, a phylogenetic analysis was performed with the 13 candidate cacao genes and several representative genes from each of the five B3 domain families in Arabidopsis (Figure 2-1A). The 13 cacao genes clustered with three of the B3 domain containing gene families (HIS, ABI3, RAV). Three cacao genes clustered within the ABI3 subfamily, with one cacao gene pairing with each of the three Arabidopsis members of this group (*Tc04g004970* with *AtFUS3*, *Tc01g024700* with *AtABI3* and *Tc06g015590* with *AtLEC2*), again suggesting that *Tc06g015590* is the most likely ortholog of *AtLEC2* in the cacao genome. This gene exists as a single copy and we tentatively designated it as *TcLEC2*.

The annotation of *TcLEC2* (*Tc06g015590*) in the cacao genome database predicted two translational start sites 72 bp apart. PCR primers were designed based on the most 5' potential translation start site and a predicted full-length coding sequence of *TcLEC2* was amplified from cDNA extracted from *SCA6* mature zygotic cotyledons. A 1368 bp fragment was sequenced and after alignment with the *TcLEC2* genomic sequence, a gene model was constructed, consisting of six exons and five introns, nearly identical to the *AtLEC2* gene structure (Figure 2-1B). The lengths of the first and last exons differ slightly and the remaining four are identical. The *TcLEC2* encodes an open reading frame of 455 amino acid residues with the B3 domain predicted in the central region of the polypeptide. The full-length TcLEC2 protein shares 42% identity with AtLEC2 (Supplemental Figure 2-2); however, they are 81% identical within the B3 domain (Figure 2-1C).

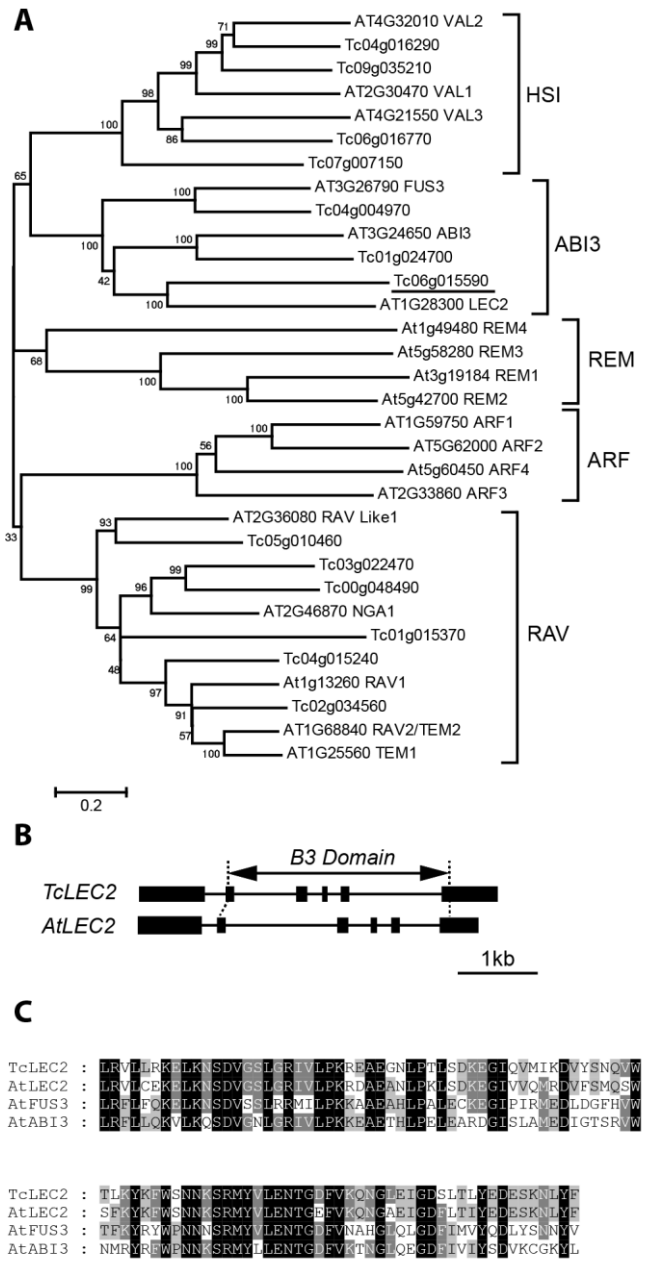


Figure 2-1 Phylogenetic analysis and gene structure of B3 domain containing genes in cacao. A. Unrooted neighbor-joining consensus tree of full-length amino acid sequences of selected Arabidopsis and *Theobroma cacao* B3 domain containing genes. The scale bar represents 0.2 estimated substitutions per residue and values next to nodes indicate

bootstrap values from 1000 replicates. Five families of B3 domain containing genes were identified. The gene most closely related to *AtLEC2* (underlined) was designated as *TcLEC2*. B. Comparison of *TcLEC2* and *AtLEC2* gene structures. Boxes represent exons and lines indicate introns. Location of the conserved B3 domain is indicated. C. Amino acid alignment of B3 domains from *TcLEC2*, *AtLEC2*, *AtFUS3*, and *AtABI3*. Residues in black boxes are identical in all four proteins; residues in dark grey boxes are identical in three of four proteins; residues in light grey boxes are identical in two of four proteins.

TcLEC2 is expressed primarily in endosperm and early mature embryo cotyledons

To investigate the function of TcLEC2 in cacao, its expression was measured by qRT-PCR in various tissues including: leaves at developmental stages A, C, and E (defined in (Mejia et al., 2012)), unopened flowers, open flowers, roots, endosperm and zygotic seeds at 14, 16, 18, 20 weeks after pollination (WAP). A cacao beta-tubulin gene (*TcTUB1*, *Tc06g000360*) that was previously shown to exhibit stable expression levels during cacao seed development (unpublished data) was used for normalization. *TcLEC2* was exclusively expressed in cacao endosperm and cotyledon (Figure 2-2), and significant levels of transcript were not detected in other tissues, consistent with the *AtLEC2* expression pattern in Arabidopsis (Stone et al., 2001; Sreenivasulu and Wobus, 2013). Moreover, the expression of *TcLEC2* was significantly higher in cacao cotyledons at 14 and 18 WAP compared to 16 and 20 WAP, stages previously defined as the onsets of cacao embryo morphogenesis and the seed maturation phase, respectively (Patel et al., 1994). A similar biphasic expression pattern was reported for *LEC2* in Arabidopsis (Sreenivasulu and Wobus, 2013), suggesting a potential role of TcLEC2 in early developmental induction and in maturation phases of zygotic embryogenesis. Notably, the transcript of *TcLEC2* was accumulated to high levels in endosperm (90 days after pollination) when the embryo had just begun development (Figure 2-2). The endosperm functions to provide nutritive support to the developing embryo, and for crosstalk between maternal tissue and the embryo, being a critical determinant of successful embryo development (Costa et al., 2004). Therefore, the abundance of *TcLEC2* transcript in the endosperm of developing cacao ovules suggests that

TcLEC2 expression in endosperm could be involved in controlling embryo initiation in cacao.

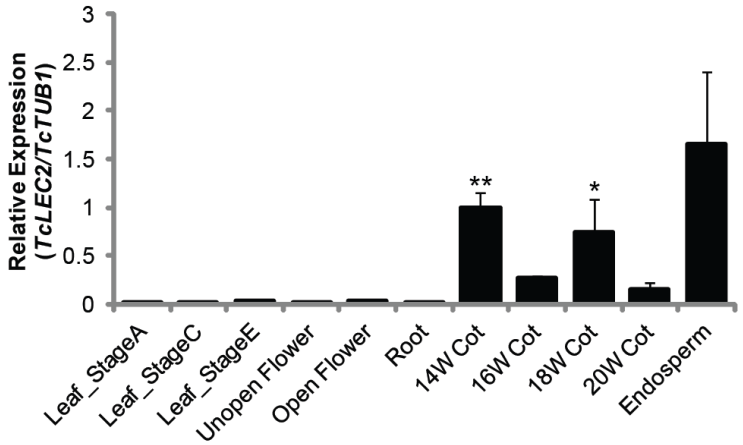


Figure 2-2 TcLEC2 expression pattern in different cacao tissues. Tissues include: leaves and flowers at different developmental stages, roots and zygotic cotyledons from seeds collected at 14, 16, 18 and 20 weeks after pollination (14W Cot, 16W Cot, 18W Cot and to 20W Cot respectively). The expression levels were analyzed by qRT-PCR and *TcLEC2* gene normalized relative to that of *TcTUB1* gene. Bars represent mean values (n=3; mean ± SE). Significance was established by t-test (** represents p-value < 0.01by t-test; * represents p-value < 0.05 by t-test).

Ectopic expression of *TcLEC2* was sufficient to activate seed specific gene expression in cacao leaves

To test the function of cacao TcLEC2 in regulation of gene expression and to identify its putative downstream targets, a rapid transient transformation assay using cacao leaf tissue was utilized (Shi et al., 2013) (see Materials and Methods, Supplemental Figure 2-3). *TcLEC2* was ectopically overexpressed under the *E12-Ω* modified CaMV35S promoter (E12Ω::TcLEC2, pGZ12.0108, GenBank Accession: KF963132, Supplemental Figure 2-4) in fully expanded young stage C cacao leaves using *Agrobacterium* vacuum infiltration.

Agrobacterium containing empty based vector pGH00.0126 (control vector, GenBank Accession: KF018690, EGFP only) was also infiltrated in parallel as a control. As expected, *TcLEC2* was highly expressed only in leaves transformed with E12Ω::TcLEC2 vector but was not detectable in control leaves (Figure 2-3). To identify the potential targets of TcLEC2, a set of cacao putative orthologs of genes involved in seed development in Arabidopsis was also assayed via qRT-PCR (Table 1). The predicted ortholog of *AGAMOUS-Like* 15, a MADS box type transcription factor involved in the induction of somatic embryogenesis from shoot apical meristems (Harding et al., 2003), was highly induced (>129 fold) by *TcLEC2* ectopic overexpression (Figure 2-3), which was consistent with the observation that LEC2 and AGL15 were able to activate each other in Arabidopsis (Zheng et al., 2009). The predicted ortholog *of ABA INSENSITIVE 3* (*ABI3*), which encodes a B3 domain transcription factor active during seed development and previously identified as a downstream target of AtLEC2 in Arabidopsis (Nambara et al., 1992; Monke et al., 2012), was also induced (>9 fold) by TcLEC2 (Figure 2-3). However, another B3 domain transcription factor *FUSCA 3* (*FUS3*) (Luerssen et al., 1998; Yamamoto-Toyoda et al., 2007) was not responsive to *TcLEC2* overexpression in leaf tissues under our experimental conditions (Table 2-1). The predicted ortholog *of WRINKLED 1* (*WRI1*), an AP2/EREB family transcription factor that is the direct downstream target of AtLEC2 and specifies AtLEC2 function toward fatty acid biosynthesis pathway in Arabidopsis (Baud et al., 2007; Maeo et al., 2009), was induced more than ten-fold by TcLEC2 (Figure 2-3). Moreover, two genes encoding for *OLEOSIN* proteins, involved in the structure of oil bodies, were also activated in cacao leaves by *TcLEC2* ectopic overexpression (Figure 2-3). Collectively, these results indicated that TcLEC2 was sufficient to induce the ectopic transcription of several important seed specific genes in cacao leaves, supporting its function as a key regulator of embryo and seed development.

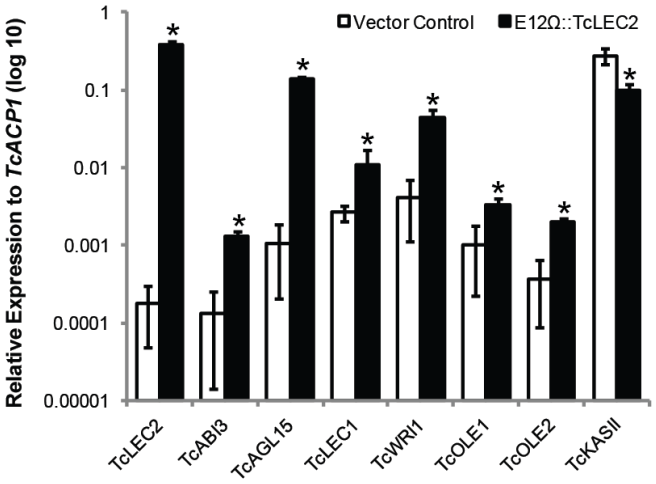


Figure 2-3 Genes induced by ectopic overexpression of *TcLEC2* in attached cacao leaf transient assay. Expression levels of *TcLEC2* and TcLEC2 induced genes (*TcAGL15*, *TcABI3*, *TcWRI1*, *TcOLE1*, and *TcOLE2*) in *TcLEC2* ectopic expressing attached cacao leaves compared to vector control by qRT-PCR. The expression levels of genes were normalized relative to that of *TcACP1*. (n=3, mean ± SE) * represents for p-value < 0.05 by t-test.

Table 2-1 Changes in gene expression levels in response to *TcLEC2* ectopic expression in leaf tissues of genes involved in various dimensions of seed development calculated from qRT-PCR measurements. Expression levels were normalized to TcACP1 (column 3 and 4). Fold changes were calculated as a ratio of expression levels induced by TcLEC2 relative to expression levels in tissues transformed with the vector control lacking*TcLEC2* and are the mean of three biological replicates (n=3, mean ± SE). Significance was established by t-test with p-value cut-off of 0.05.

Gene	Gene_ID	Vector Control	E12Ω::TcLEC2	Fold Change	Gene Function
TcAGL15	*Tc01g040120*	0.0002 ± 0.0001	0.3821 ± 0.0422	129.2	*MADS* box transcription factor, regulate GA biosynthesis
TcWRI1	*Tc10g012790*	0.0001 ± 0.0001	0.0013 ± 0.002	10.53	*AP2/ERWEBP* transcription factor, regulate fatty acid biosynthesis and

					embryo development
TcABI3	*Tc01g024700*	0.001 ± 0.0008	0.1355 ± 0.0103	9.57	*B2,B3* domain transcription factor, regulation ABA induced gene expression
TcOLE2	*Tc09g004410*	0.0027 ± 0.0006	0.011 ± 0.0061	5.49	Oleosin, oily body structure protein
TcLEC1	*Tc07g001180*	0.0041 ± 0.003	0.433 ± 0.0133	4.14	*HAP3* subunit CCAAT-binding transcription factor
TcOLE1	*Tc04g001560*	0.001 ± 0.0008	0.0033 ± 0.0006	3.27	Oleosin, oily body structure protein
TcKASII	*Tc09g006480*	0.2742 ± 0.0607	0.0989 ± 0.0222	-2.77	3-ketoacyl-ACP-synthase II
TcFUS3	*Tc04g004970*	nd	nd	Not Induced	*B3* domain transcription factor, direct bind to RY motif
TcLEC1_Like	*Tc06g020950*	0.0011 ± 0.0006	0.0016 ± 0.0003	Not Induced	*HAP3* subunit CCAAT-binding transcription factor
TcVicilin	*Tc04g024090*	0.0003 ± 0.0002	0.0002 ± 0.0001	Not Induced	most abundant seed storage protein in cacao
TcBBM	*Tc05g019690*	nd	nd	Not Induced	*AP2* transcription factor in developing embryos and seeds
TcPKL	*Tc09g001610*	0.3973 ± 0.0578	0.3928 ± 0.0733	Not Induced	*CHD* chromatin remodeling factor
TcWUS	*Tc01g001780*	nd	nd	Not Induced	Master regulator of stem cell fate determination in shoot apical meristem
TcYUC2	*Tc09g009820*	0.0016 ± 0.0002	0.001 ± 0.0001	Not Induced	flavin monooxygenase in auxin biosynthesis
TcYUC4	*Tc09g013260*	0.0093 ± 0.0069	0.0085 ± 0.0021	Not Induced	flavin monooxygenase in auxin biosynthesis

TcLEC2 expression is associated with embryogenic competency of callus cells

Based on the above results, we reasoned that TcLEC2 might also be a key regulator of somatic embryogenesis. To explore this, *TcLEC2* was measured in tissues grown with or without the SE inducing hormone 2,4-D (Figure 2-4). Staminodes from the highly embryogenic cacao genotype *PSU-SCA6* were used to produce primary somatic embryos (Figure 2-4A, panel i) following our previously published protocol (Maximova et al., 2002). Cotyledon explants (Figure 2-4A, panel i, red box) were excised and placed on secondary embryogenesis induction media with (SCG) or without (SCG-2,4D) auxin 2,4-D (required for SE induction). After two weeks on these media, the tissues were transferred biweekly to hormone free embryo development media (ED). The explants cultured on SCG media started to produce calli two-weeks after culture initiation (ACI) (panel ii) and secondary somatic embryos were visible after four additional weeks (panel iv and vi). However, on SCG-2,4D, explants expanded and gradually turned green during the first six weeks, then stopped developing and turned brown. Neither calli nor embryos were produced from the explants on SCG-2,4D medium (panel iii, v and vii).

TcLEC2 expression levels were measured in tissues cultured on both SCG and SCG-2,4D media throughout the culture period (Figure 2-4B). *TcLEC2* expression was detectable in primary somatic embryo cotyledons at time 0, then decreased significantly one day after explants were placed on either SCG or SCG-2,4D media (Figure 2-4B). *TcLEC2* expression remained low in both treatments for the following two weeks, indicating that *TcLEC2* was not rapidly responsive to exogenous auxin treatment during the induction period. However, between day 32 and 36 ACI, *TcLEC2* expression levels were slightly increased and variable in both treatments. Notably, at 46 days ACI the development of embryos was first observed on SCG media arising from calli (embryogenic calli). RNA was extracted from the embryogenic calli (without visible embryos) and a large increase in *TcLEC2* gene expression was observed by qRT-PCR. On SCG-2,4D media, embryos were not observed and *TcLEC2* expression was not detectable.

A common occurrence in tissue culture is dedifferentiation of different types of calli that vary in their totipotency to regenerate somatic embryos (Delporte et al., 2013; She et al., 2013). With cacao tissue cultures, we and our collaborators have frequently observed two types of calli, those that produce abundant embryos (embryogenic calli) and those that produce few if any embryos (non-embryogenic calli) (unpublished observations). To

investigate the relationship between TcLEC2 activity and embryogenic potential of the calli, *TcLEC2* gene expression was compared in embryogenic and non-embryogenic calli growing from explants cultured on SCG media. The observed average levels of *TcLEC2* expression were 20-fold higher in the embryogenic calli compared to the non-embryogenic calli of the same age (Figure 2-4C), suggesting a tight association between *TcLEC2* expression and embryogenic competency. Given the role of AtLEC2 in controlling embryo development in Arabidopsis, we hypothesized that TcLEC2 may play a similar role in the control of cacao somatic embryo development.

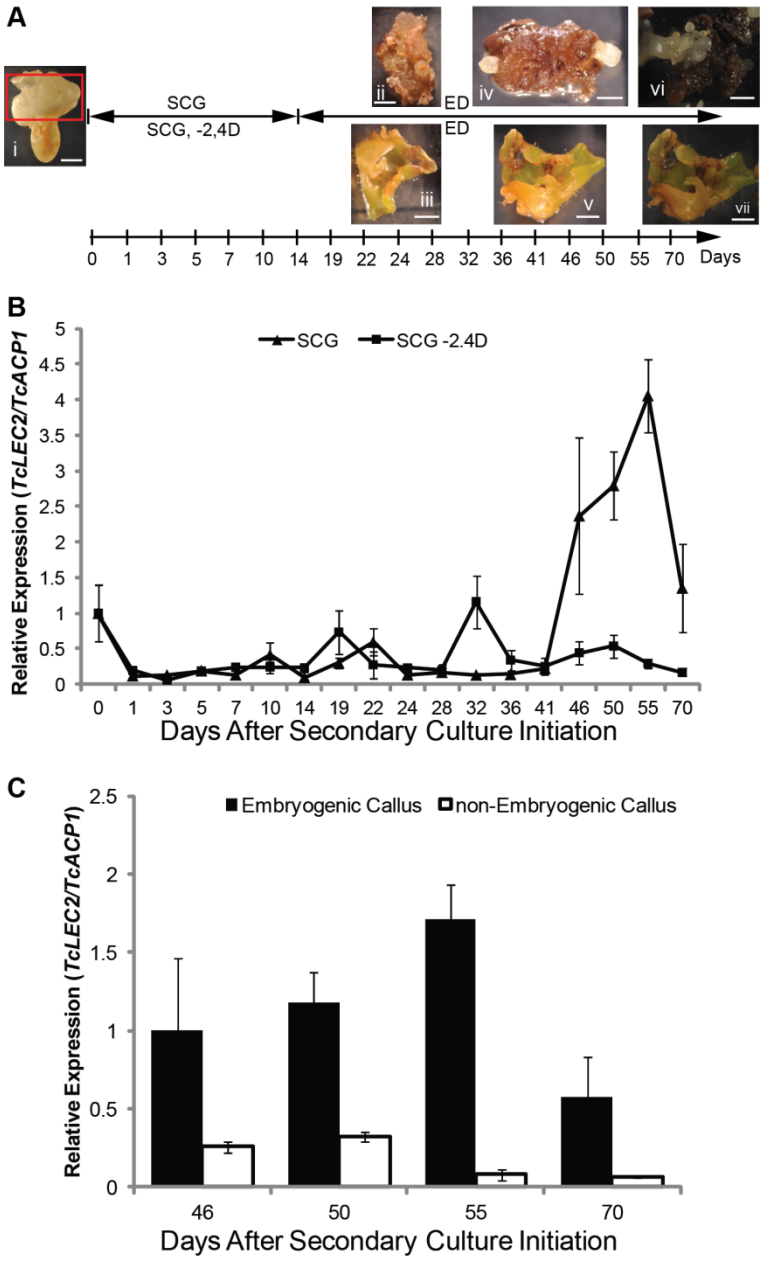


Figure 2-4 *TcLEC2* expression correlates with embryogenic potential. A. Illustration of the cacao secondary somatic embryogenesis stages and time frame, indicating the points used for sample collections. Representative images of several key stages of embryo development: (i) cotyledon stage *PSU-SCA6* embryo used as explants to initiate secondary somatic embryogenesis cultures; (i) & (iii) cotyledon explants on hormone-free medium at 28 days ACI, from cultures initiated on SCG medium containing 2,4D and modified SCG without 2,4D, respectively; (iv) & (v) cotyledon explants on ED at 46 days ACI (same treatments as above); (vi) & (vii) cotyledon explants on ED at 70 days ACI (same treatments as above); (Bars = 2 mm). B. Time course expression pattern of *TcLEC2* during cacao secondary somatic embryogenesis from cultures initiated on SCG medium containing 2,4D and modified SCG without 2,4D. Expression of *TcLEC2* was normalized relative to that of *TcACP1* (n=3 or 4, mean ± SE). C. Expression levels of *TcLEC2* at different time points in embryogenic and non-embryogenic calli. Expression of *TcLEC2* was normalized relative to that of *TcACP1*. Bars represent mean ± SE (n=3 or 4). Significance was established by t-test (* represents for p-value < 0.05).

Overexpression of *TcLEC2* significantly increased efficiency of somatic embryogenesis and regeneration of transgenic embryos

The current methods for *Agrobacterium*-mediated transformation of cacao genotype results in reproducible but very low rates of transgenic embryo recovery (Maximova et al., 2003). We speculate that this is a result of very low co-incidence of stable T-DNA integration into the cacao genome and the same cells entering the embryogenic pathway. We hypothesized that overexpression of *TcLEC2* might enhance the rate of somatic embryogenesis and thus improve the recovery of transgenic SEs through increased co-incidence with T-DNA integration events.

To test this, we performed *Agrobacterium*-mediated transformation experiments on cotyledon explants excised from primary embryos for co-cultivation with *Agrobacterium* containing the control vector (pGH00.0126) or the E12Ω::TcLEC2 vector (pGZ12.0108). Two weeks after co-cultivation, the initial transient expression levels of GFP in tissues transformed with the control vector were always higher than E12Ω::TcLEC2 (Supplemental

Figure 2-5). This may be due to the larger size of the E12Ω::TcLEC2 containing plasmid relative to the control vector, the inclusion of a repeated promoter element, or the addition of a third highly expressed transgene. We have observed this phenomenon with other unrelated plasmids containing transgenes (unpublished data).

During the subsequent weeks of culture on embryogenesis media, large numbers of non-transgenic embryos (GFP negative) were observed in all three independent transformation trials regardless of the presence of the *TcLEC2* transgene (Supplemental Figure 2-6). There was no consistently significant difference observed between the transformations of control vector and E12Ω::TcLEC2 in terms of the cumulative non-transgenic embryo production (Supplemental Figure 2-6). To identify stably transformed embryos, GFP fluorescence was observed by stereomicroscopy as a visualization marker. With the control vector lacked the *TcLEC2* transgene, no GFP expressing embryos were observed on over 176 cotyledon explants cultured in three separate experiments. Surprisingly, the transformation with E12Ω::TcLEC2, containing the*TcLEC2* transgene, resulted in the recovery of over 300 stable transgenic embryos distributed over the entire surface of the cotyledon explant (Figure 2-5 A-B). This result was dramatically higher than the stable transformation results we have observed over many years, with several different transgenes, where the prior record for a single transformation (about 200 cotyledon explants) was 8 GFP positive embryos (Maximova et al., 2003; Maximova et al., 2006; Mejia et al., 2012). Thus, although TcLEC2 did not impact the initial levels of transient transformation (Supplemental Figure 2-5) or embryogenesis frequency of non-transgenic embryos (Supplemental Figure 2-6), it greatly increased the frequency of transgenic embryo production, which confirmed our hypothesis.

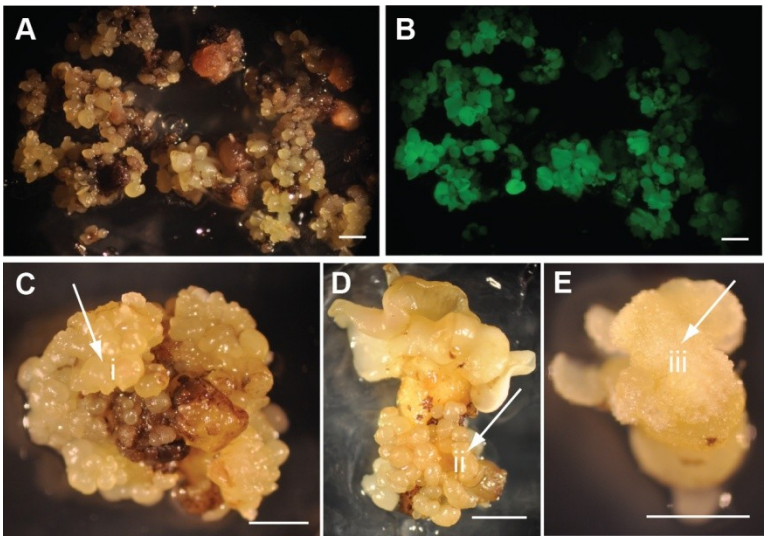


Figure 2-5 Effect of stable overexpression of *TcLEC2* in cacao secondary somatic embryos. A&B. Secondary embryogenic explants transformed with *Agrobacterium*, regenerating stable transgenic E12Ω::TcLEC2 embryos were photographed under white and with GFP imaging optics, respectively. C, D&E. Transgenic somatic embryos expressing E12Ω::TcLEC2. (i) embryo-like structure formed on top of cotyledon (ii) embryo-like structure formed along embryo axis (iii) callus-like structure formed on top of cotyledon.

Although a large number of transgenic *TcLEC2* embryos were obtained, most of them exhibited prominent developmental and morphological abnormalities (Figure 2-5 C, D, and E), and most ceased development at the globular or heart stage and the initiations of cotyledons were significantly compromised. The few embryos that did develop to cotyledon stage formed callus on top of the cotyledons (Figure 2-5E) and new embryos were occasionally initiated along the embryo axis (Figure 2-5C and D). The attempts to recover plants from any of these embryos were unsuccessful.

To test the effect of stable overexpression of *TcLEC2* transgene on iterative somatic embryogenesis, cotyledons from fully developed mature transgenic E12Ω::TcLEC2 embryos were excised and cultured for tertiary embryo production as previously described (Maximova et al., 2002). Cotyledon explants from non-transformed *PSU-SCA6* SEs were cultured as controls. Remarkably, cotyledon explants from transgenic E12Ω::TcLEC2 lines

started to produce tertiary embryos as early as four weeks ACI (Figure 2-6B), compared to six weeks for *PSU-SCA6* lines (Figure 2-6A). Additionally, while the majority of tertiary embryo production from *PSU-SCA6* lines was completed by 14 weeks ACI, after which very few SEs were produced (Figure 2-6C), explants from transgenic E12Ω::TcLEC2 lines continued to produce large numbers of embryos until twenty weeks ACI, when the experiment was terminated (Figure 2-6D). In total, within the twenty week period, transgenic E12Ω::TcLEC2 lines produced about 2.5 times more tertiary embryos per explant (p-value < 0.001) compared to *PSU-SCA6* lines (Figure 2-6E.)

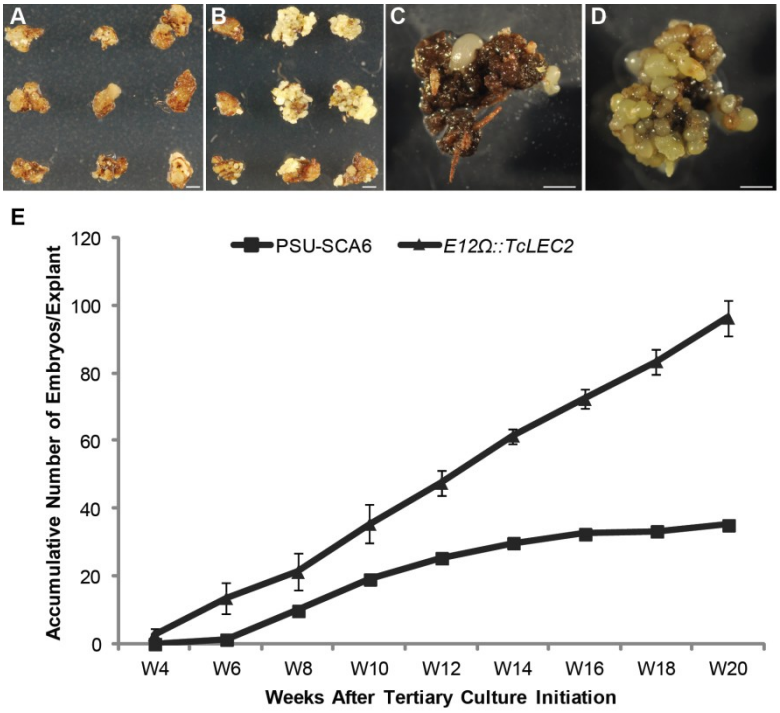


Figure 2-6 Overexpression of *TcLEC2* increases tertiary somatic embryogenesis efficiency. A. Tertiary *PSU-SCA6* culture on hormone free medium at 4 weeks ACI. B. Tertiary stable transgenic E12Ω::TcLEC2 culture on hormone free medium at 4 weeks ACI. C. Tertiary *PSU-SCA6* culture on hormone free medium 20 weeks after culture initiation. D. Tertiary stable transgenic E12Ω::TcLEC2 culture on hormone free medium at 20 weeks ACI. E. Average number of tertiary embryos produced per explant from *PSU-SCA6* and stable transgenic E12Ω::TcLEC2 explants (n=4, mean ± SE) (Bars = 2mm).

Overexpression of *TcLEC2* altered the expression of genes involved in fatty acid biosynthesis

In addition to its role in initiation of embryogenesis, it has been well documented in Arabidopsis that AtLEC2 also regulates de novo fatty acid biosynthesis during embryo development. Evidence includes, but is not limited to, (a) transgenic 35S::AtLEC2 ovules exhibited a mature seed-like fatty acid profile (Stone et al., 2008); (b) ectopic overexpression of *AtLEC2* in leaves resulted in accumulation of seed specific lipids and very long chain fatty acids (Santos Mendoza et al., 2005); (c) *AtLEC2* directly regulates expression of *AtWRI1,* which is known to play a role in regulation of fatty acid metabolism in developing embryos (Baud et al., 2007). Since fatty acids, in the form of triacylglycerols (TAGs), are major storage components of mature cacao seeds, we examined the role of TcLEC2 in control of fatty acid biosynthesis in cacao immature zygotic embryos (IZEs).

E12Ω::TcLEC2 was transiently overexpressed in IZEs (12 weeks old) in parallel with the control vector (pGZ00.0126). High transient expression was confirmed by fluorescence microscopy to detect EGFP on approximately 90% of explants surfaces (Supplemental Figure 2-7). Overexpression levels of *TcLEC2* in transformed IZEs were further confirmed by qRT-PCR and compared to the basal levels of *TcLEC2* in control vector transformed IZEs (Figure 2-7). Consistent with the observations in attached leaf transient assay (Figure 2-3), the overexpression of *TcLEC2* resulted in elevated transcript of *TcAGL15* and *TcLEC1* in the IZE tissues (Figure 2-7A). Unlike in transiently transformed leaf tissue, induced expression of *TcWRI1* was not detected in the transformed IZEs under our qRT-PCR condition (40 cycles) (Supplemental Figure 2-9).

To obtain further insights into TcLEC2 regulatory functions during embryo development, we identified the most likely orthologs of genes for key enzymes controlling the fatty acid biosynthesis and production of TAGs in the cocoa genome by homology to Arabidopsis gene sequences (Figure 2-7C and Supplemental Figure 2-8) and compared their expression levels in IZEs tissues overexpressing E12Ω::TcLEC2 and control vector (Figure 2-7A). mRNA levels of *TcKASII* (*Tc09g006480*), a condensing enzyme β-ketoacyl-[acyl-carrier-protein] synthase II responsible for the elongation of C16 to C18 (Wu and Xue,

2010), was two-fold lower (p-value < 0.05) in the *TcLEC2* transformed tissue compared to the controls (Figure 2-7A). In addition, the predicted ortholog of *TcFatA* (*Tc01g022130*) and two isoforms of *TcFatB* (*Tc01g022130* and *Tc03g015170*), two types of acyl-[acyl-carrier-protein] thioesterases that specifically export C18:1 (FatA) and other saturated fatty acid moieties (FatB) from plastid into cytosol (Moreno-Perez et al., 2012), were significantly up-regulated by more than 1.5 fold (p-value < 0.05). Interestingly, the predicted diacylglycerol acyltransferase 2 (*TcDGAT2*, *Tc01g000140*), a key enzyme that catalyzed the last step of TAG assembly through an acyl-CoA dependent pathway (Lung and Weselake, 2006), was significantly up-regulated by 1.5 fold (p-value<0.05). No significant differences in the expression levels of two isoforms of fatty acid desaturase 2 (*FAB2, Tc04g017510 and Tc08g012550*) were observed (Supplemental Figure 2-9).

To determine if these changes in gene expression resulted in altered metabolite profiles, fatty acid composition was measured by gas chromatograph/mass spectrometry (GC/MS) in IZEs tissues transformed with both E12Ω::TcLEC2 and control vector. Overexpression of *TcLEC2* resulted in a significant increase of the level of cis-vaccenic acid C18:1^{n-7} (p-value < 0.001), an isoform of oleic acid (OA), and significantly decreased the level (p-value < 0.001) of linoleic acid (LA, C18:2^{n-6}) compared to tissues transformed with vector control (Figure 2-7B).

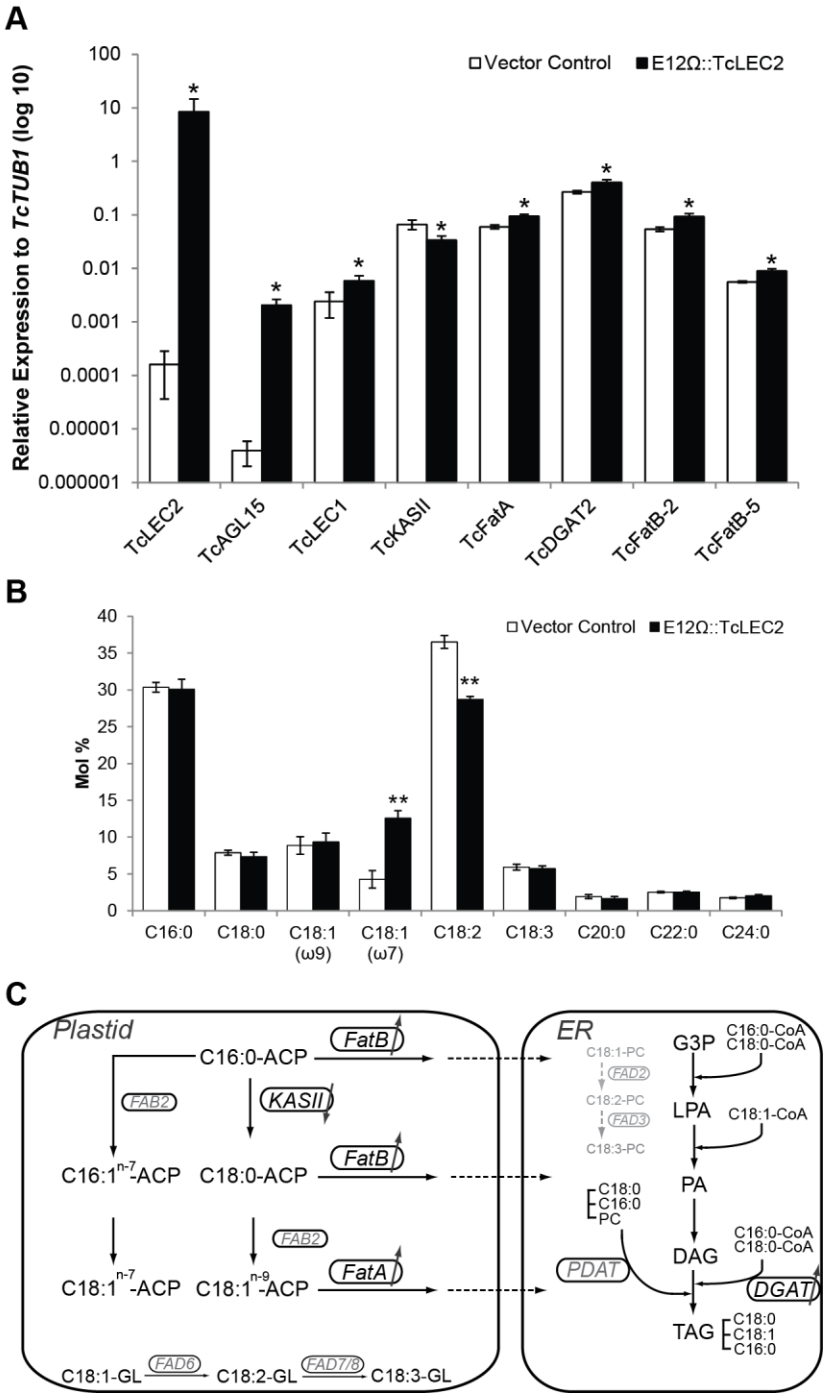


Figure 2-7 Transient overexpression of E12Ω::TcLEC2 in IZE altered fatty acid compositions and gene expression. A. Changes of the expression levels of responsive enzymes on the fatty acid biosynthesis pathway; expression levels of genes were normalized relative to that of *TcTUB1*; (n=3, mean ± SE) * represents for p-value < 0.05 by t-test. B. Molar percentages of fatty acid compositions in cacao immature zygotic embryos

transiently overexpressing vector control and E12Ω::TcLEC2, respectively; (n=3, mean ± SE). ** represents for p-value < 0.001 by t-test. C. Diagram of proposed model to explain the relationship between gene expression levels and altered fatty acid compositions. Enzymes are marked in circle. Enzymes that were regulated by the activity of TcLEC2 are in black, otherwise, in grey.

Abbreviation:

ER, endoplasmic reticulum; ACP, acyl carrier protein; CoA, Coenzyme A; FAB2, fatty acid desaturase; Fat, fatty acyl-ACP thioesterase; KAS, 3-ketoacyl-ACP synthase; FAD2, oleoyl desaturase; FAD3, linoleoyl desaturase; FAD6, oleoyl desaturase on membrane glycerolipid; FAD7/8, linoleoyl desaturase on membrane glycerolipids; PC, phosphatidylcholine; G3P, glycerol-3-phosphate; LPA, lysophosphatidate; PA, phosphatidate; DAG, diacylglycerol; TAG, triacylglycerol; PDAT, phospholipid:diacylglycerol acyltransferase; DGAT, 1,2-sn-diacylglcyerol transferase.

2.4 Discussion

TcLEC2 is involved in cacao somatic embryogenesis

Somatic embryogenesis has long been considered a superior propagation system for many crops (Barry-Etienne et al., 2002; Samaj et al., 2003; Kita et al., 2007) because of its inherent high multiplication rate and potential for year round, uniform disease free plant production. Although theoretically, every somatic plant cell has the capacity to dedifferentiate and redifferentiate into a whole plant (totipotency), the competencies of plant cells to enter the somatic embryogenesis developmental pathway varies dramatically between different tissues, developmental stages, and species. Accumulated evidence has revealed that the activity of AtLEC2 is highly associated with embryogenic competency and involves interactions with several other regulatory factors. Our results are consistent with a role of cacao TcLEC2 in the regulation of somatic embryogenesis similar to AtLEC2 in Arabidopsis. Supporting evidence includes; (1) ectopic overexpression of *TcLEC2* in cacao stage C leaves was able to induce the expression of seed transcription factor genes, such as

TcAGL15, *TcABI3* and *TcLEC1;* (2) the induced expression level of *TcLEC2* was associated with embryogenic capacity in explants; (3) constitutive overexpression of *TcLEC2* in secondary somatic embryo tissue leads to earlier and increased regeneration of tertiary embryos compared to *PSU-SCA6* controls. Collectively, our evidence supports the conclusion that TcLEC2 is a functional ortholog of AtLEC2 and that it is involved in similar genetic regulatory networks during cacao somatic embryogenesis.

Transient overexpression of *TcLEC2* in cotyledon explants by itself was not sufficient to increase embryogenesis efficiency of non-transgenic somatic embryos (Supplemental Figure 2-6). This suggests that there are other factor(s) that are required for cell dedifferentiation and redifferentiation, which are not present during the period of time examined in our embryogenesis culture system. However, the constitutive overexpression of *TcLEC2* in stably transformed cells resulted in greatly enhanced somatic embryogenesis as early as four weeks compared to six to seven without *TcLEC2* overexpression (Figure 2-5B, 6B and 6E), implying that the enhanced activity of TcLEC2 is sufficient to promote the efficiency of somatic embryogenesis in cacao.

The very high degree of genotype variation in embryogenic capacity for SE in cacao limits its' practical application for large scale propagation (Maximova et al., 2002). Therefore, *TcLEC2* could be a useful molecular marker for screening cacao genotypes for high embryogenic capacities. Additionally, the levels of *TcLEC2* expression in callus and other tissues *in vitro* could be used for evaluating the effect of different media and other variables for further optimization of the SE protocols. Potentially, we could explore the possibility to promote somatic embryogenesis in cacao leaves or other tissues by ectopically expressing *TcLEC2*.

TcLEC2 regulates fatty acid biosynthesis during cacao seed maturation

Fatty acid composition and lipid profiles of cacao seeds are important quality traits for chocolate industry. Therefore, there is great interest in identification of the genetic networks regulating its biosynthesis. LEC2, and its partners LEC1, ABI3 and FUS3 are known to be critical regulators of fatty acid and lipid biosynthesis in Arabidopsis and other species, and thus impact many aspects of seed development. Moreover, of particular

relevance to applications of this knowledge, the level of WRI1, a downstream target of LEC1, LEC2 and FUS3, was highly correlated with seed oil content in different *B. napus* genotypes (Tan et al., 2011). Our observations that *TcLEC2* overexpression resulted in increased expression of *TcLEC1* and *TcWRI1* (Figure 2-3) in attached cacao leaves led us to speculate that this might result in changes in fatty acid composition and TAG assembly. Indeed, transient overexpression of *TcLEC2* in zygotic embryos resulted in increased C18:1^{n-7} and decreased C18:2^{n-6} levels (Figure 2-7B), similar to changes occurring during cacao seed maturation when profiles change from mainly polyunsaturated fatty acids (C18:2^{n-6} and C18:3^{n-3}) to almost exclusively saturated (C16:0 and C18:0) and monounsaturated fatty acid (C18:1^{n-9}) (Patel et al., 1994). However, given the fact that the expression of *TcWRI1* was not induced by overexpression of *TcLEC2* in immature zygotic embryos, it suggests that WRI1 is not required to mediated impacts of TcLEC2 on fatty acid biosynthesis, and that the regulatory network between TcLEC2 and other transcription factors on fatty acid biosynthesis is not the same in cacao as they are in Arabidopsis and *B. napus*. The overexpression of *TcLEC2* also resulted in changes in gene expression for some of the major structural genes for fatty acid biosynthesis, and this could provide an explanation for the fatty acid composition shifts we observed. C18:1^{n-7} is synthesized from C16:0 via the production of C16:1^{n-9} by FAB2 and further elongation to C18:1^{n-7} (Barthet, 2008). The decreased expression level of *TcKASII* may increase the substrate availability of C16:0, which could serve as a substrate for TcFAB2 for the production of C16:1^{n-9} and further leading to C18:1^{n-7} accumulation (Figure 2-7C). The increased levels of *TcFatA* and two isoforms of *TcFatB* (all significantly up-regulated by more than 1.5 fold) could contribute to increased production and accumulation of saturated fatty acid (C16:0 and C18:0) and monounsaturated fatty acid (C18:1^{n-9}) during cacao seed maturation (Figure 2-7C).

Interestingly, the expression of *TcDGAT2* was also significantly increased by overexpression of *TcLEC2*, but the expression level of *TcDGAT1.1* was not affected (Supplemental Figure 2-9). The activities of DGAT genes were highly correlated with the oil content and compositions in oilseeds (Zheng et al., 2008) and three known types of DGAT genes (DGAT1, DGAT2 and DGAT3) are different in terms of substrate specificities and

subcellular localizations (Shockey et al., 2006). According to an unpublished study, the expression of *TcDGAT2* in yeast has led to accumulation of more C18:0 in TAG fraction compared to the expression of *TcDGAT1* (Zhang, 2012). Considering the fact that the majority of TAGs in cacao mature seeds consist of unsaturated fatty acid (C18:1) exclusively on sn-2 and saturated fatty acids (C16:0 and C18:0) on sn-1 and 3 (Figure 2-7C), it is plausible to speculate, that the activity of TcDGAT2 is more significant to catalyze the final acylation on sn-3 of TAG assembly compared to TcDGAT1. This argument was further supported by our result indicating that the expression level of *TcDGAT2* was approximately five times higher than *TcDGAT1* in cacao immature seeds (Figure 2-7A and Supplemental Figure 2-9). Collectively, our data indicates that TcLEC2 could be involved in regulation of lipid biosynthesis during cacao seed maturation through control of *TcDGAT2* gene expression. However, whether TcLEC2 is able to directly trans-activate *TcDGAT2* or its action is mediated through other transcription factors, remains unknown. Further research on the regulatory mechanism controlling fatty acid biosynthesis and TAG assembly in cacao will contribute to identification of the key enzymes in the pathway and aid the screening process for elite cacao varieties to meet industrial demands.

2.5 Conclusion

The isolation and functional characterization of *LEC2* ortholog from cacao genome reveal crucial roles of TcLEC2 in regulating both zygotic and somatic embryogenesis. The exclusive expression pattern in seed and the identification of its regulatory targets, such as *AGL15* and *WRI1*, strongly indicate the functional similarities between AtLEC2 and TcLEC2. However, the impacts of TcLEC2 on fatty acid biosynthesis in cacao also suggest that TcLEC2 is able to directly or indirectly interact with many key enzymes on the pathway, which has not been well characterized yet in Arabidopsis. Furthermore, the correlation between the activity of TcLEC2 and embryogenic potential during cacao somatic embryogenesis provides us a great opportunity to better understand and improve our current inefficient and variable propagation system of cacao.

2.6 Materials and Methods

Phylogenetic Analysis and Sequence Alignment

B3 domain containing genes in *Theobroma cacao* were identified by blastp using AtLEC2 (At1g28300) as queries (E-value cut off $1e^{-5}$). Multiple protein sequence alignment was performed by MUSCLE (Edgar, 2004). The phylogenetic tree was constructed by MEGA4.1 using neighbor-joining algorithm with Poisson correction model and the option of pairwise deletion (Tamura et al., 2007). Bootstrap values represent 1000 replicates. Full-length Arabidopsis *AtLEC2, AtABI3,* and *AtFUS3,* protein sequences were used to search the Cocoa Genome Database (http://cocoagendb.cirad.fr/) by tblastn (Altschul et al., 1990) to obtain the full-length *TcLEC2, TcABI3* and *TcFUS3* nucleotide sequences, respectively. The functional B3 domains were predicted using InterPro program (http://www.ebi.ac.uk/interpro/) on EMBL-EBI website. B3 domain containing proteins from five subfamilies in *Arabidopsis* were identified and selected according to (Romanel et al., 2009).

RNA Extraction, *TcLEC2* Cloning and Expression Vector Construction

Plant tissues collected from SCA6 genotype of cacao were first ground in liquid nitrogen. Total RNA was extracted using Plant RNA Purification Reagent (Life Technologies, Cat. 12322-012, following manufactures protocol). The concentration of RNA was measured using a Nanodrop 2000c (Thermo Scientific). RNA was further treated with RQ1 RNase-free DNase (Promega, Cat. M6101) to remove potential genomic DNA contamination (following the manufacturer's protocol). 250ng of treated RNA was reverse-transcribed by M-MuLV Reverse Transcriptase (New England Biolabs) with oligo-$(dT)_{15}$ primers. The full length *TcLEC2* was amplified from *SCA6* mature seed cotyledon cDNA with the primer pair

(*TcLEC2-5'*-SpeI: GCACTAGTATGGAAAACTCTTACACACC and

TcLEC2-3'-HpaI: GCGTTAACTCAAAGTGAAAAATTGTAGTGATTGAC) and cloned into pGH00.0126 (Maximova et al., 2003) driven by the E12-Ω promoter resulting in plasmid pGZ12.0108 (Supplementary Figure 2-7). The recombinant binary plasmid was introduced into *A. tumefaciens* strain *AGL1* (Lazo et al., 1991) by electroporation.

TcLEC2 Expression Analysis by qRT-PCR

RNA samples were extracted and reverse-transcribed into cDNA as described above. The primers to detect *TcLEC2* transcripts were designed based on the coding sequence of *TcLEC2 (Tc06g015590 (Argout et al., 2011))* (*TcLEC2-Realtime-5'*: TGACCAGCTCTGGTGCTGACAATA; *TcLEC2-Realtime-3'*: TGATGTTGGGTCCCTTGGGAGAAT). qRT-PCR was performed in a 10 µl mixture containing 4 µl diluted-cDNA (1:50), 5µl SYBR Green PCR Master Mix (Takara), 0.2 µl Rox, and 0.4 µl each 5 µM primers. Each reaction was performed in duplicates in Roche Applied Biosystem StepOne Plus Realtime PCR System under the following program: 15 min at 94°C, 40 cycle of 15 s at 94°C, 20s at 60°C, and 40 s at 72°C. The specificity of the primer pair was examined by PCR visualized on a 2% agarose Gel and dissociation curve. An acyl carrier protein (*Tc01g039970*, *TcACP1 (Argout et al., 2011) TcACP1-5'*: GGAAAGCAAGGGTGTCTCGTTGAA and *TcACP1-3'*: GCGAGTTGAAATCTGCTGTTGTTTGG), and a tubulin gene in cacao (*Tc06g000360, TcTUB1* (Argout et al., 2011) *TcTUB1-5'*: GGAGGAGTCTCTATAAGCTTGCAGTTGG and *TcTUB1-3'*: ACATAAGCATAGCCAGCTAGAGCCAG) were used as the reference genes.

Cacao attached leaf and immature zygotic embryo transient gene expression assay

A. tumefaciens strain *AGL1* carrying either control vector (pGH00.0126, GenBank Accession: KF018690, EGFP only) or E12Ω::TcLEC2 (pGZ12.0108, GenBank Accession: KF963132, Supplemental Figure 2-4) were inoculated in 100 ml 523 medium with 50 µg/ml kanamycin and grown with shaking (200 rpm, 25°C) overnight to optical density (O.D.) of 1.0 at 420 nm. *AGL1* was pelleted at 1500xg for 17 min at room temperature and resuspended in induction media (Maximova et al., 1998) to to O.D. of 1.0 at 420 nm. *AGL1* was induced for 3 h at 100 rpm at 25°C and Silwet added to a final concentration of 0.02%. For the attached leaf transient transformation assay we used fully expanded, young leaves (developmental stage C as defined in (Shi et al., 2010)) from genotype *SCA6* grown in a greenhouse. The petioles of the leaves were wrapped with parafilm and set in the groove of a modified vacuum desiccator to create a seal and to avoid damage to leaves. The leaves were soaked in AGL1 induction media in the desiccator and were vacuum infiltrated at -22 psi for 2 min using a vacuum pump (GAST Model No. 0523-V4F-G582DX). Vacuum infiltration was performed three times to increase transformation efficiency. The transformed cacao leaves remained on the plant for three days after infiltration then

collected and evaluated by fluorescence microscopy. The regions with high GFP expression (>80% coverage) were selected and subjected to further analysis. For immature zygotic embryo transient transformation assay, developing fruit (open pollinated *Sca6*, four months after pollination) from the USDA germplasm collection in Puerto Rico. Zygotic embryo cotyledons were collected and suspended in the ED media (Maximova et al., 2002) before transformation. Zygotic cotyledons were soaked in AGL1 induction media and transformation was performed as described above for leaf transient expression assays. The transformed tissues were analyzed five days after infiltration.

Cacao stable transformation of primary somatic embryos

Primary somatic embryogenesis was performed as previously described (Li et al., 1998). Glossy cotyledons from healthy and mature primary embryos were cut into 4 mm X 4 mm square pieces, and infected using *A. tumefaciens* strain AGL1 carrying the T-DNA binary vectors as previously described (Maximova et al., 2003) with minor modifications: (1) the AGL1 was pelleted and resuspended in induction media (Maximova et al., 1998) to reach the O.D. of 1.0 at 420 nm instead of 0.5; (2) after transformation, the infected cotyledons were co-cultivated with *A. tumefaciens* strain AGL1 on the filter paper for 72h at 25°C in the dark instead of 48h. The transformed explants were cultivated and transgenic secondary somatic embryos were identified by screening for GFP fluorescence as previously described (Maximova et al., 2003).

Fatty acid profiling by GC/MS

Fresh plant tissues were ground in liquid nitrogen and fatty acid methyl esters (FAME) were prepared using approximately 30 mg of tissue extracted in 1 ml buffer containing MeOH/fuming HCl/Dichloromethane (10:1:1, v/v) while incubated without shaking at 80°C for 2h. Fatty acid methyl esters were re-extracted in 1 ml buffer H2O/Hexane/Dichloromethane (5:4:1, v/v) with vortexing for 1min. The hexane (upper phase) was separated by centrifugation at 1500xg for 5 min, transferred to Agilent glass GC vials and evaporated to dryness under a vacuum. The FAMEs were then dissolved in 500 µl hexane for GC/MS analysis. Pentadecanoic acid (C15:0) (Sigma, Cat. P6125) was used as the internal standard added prior to the extraction and methyl nonadecanoate (C19:0-methyl

ester) (Sigma, Cat. N5377) was used as the spike control, added into the sample prior to the GC injection. Fatty acid derivatives were analyzed on an Agilent 6890 Gas Chromatograph equipped with FAME Mix Omegawax 250 Capillary GC column (Sigma, Cat. 24136). A Waters GCT Classic mass spectrometry was directly connected to the GC operation. EI of 70 eV was applied. Peak height areas were used to quantify the abundance of each fatty acid species, and the mass spectra were interpreted by comparing with the NIST/EPA/NIH Mass Spectra Library (Stein, 2008).

Accession Numbers

Sequence data from this article can be found in either The Arabidopsis Information Resource (TAIR) or CocoaGenDB (http://cocoagendb.cirad.fr/gbrowse/cgi-bin/gbrowse/theobroma/) under the following accession numbers in the table:

Abbreviations

SCA6: *Scavana 6*; LEC: leafy cotyledon; ABI3: ABA INSENSITIVE 3; FUS3: FUSCA3; SE: somatic embryogenesis; ZE: zygotic embryogenesis; WRI1: WRINKLED 1; AGL15: AGAMOUS-like 15; OLE: OLEOSIN; YUC: YUCCA; ARF: auxin response factor; HSI: high level expression of sugar inducible; RAV: related to ABI3/VP1; REM: reproductive meristem; TUB1: tubulin; ACP1: acyl carrier protein; 2,4D: 2,4-Dichlorophenoxyacetic acid; SCG: secondary embryogenesis induction media; ACI: after culture initiation; IZE: immature zygotic embryo; TAG: triacylglycerol; GC/MS: gas chromatograph/mass spectrometry; OA: oleic acid; LA: linoleic acid; ER, endoplasmic reticulum; CoA, Coenzyme A; FAB2, fatty acid desaturase; Fat, fatty acyl-ACP thioesterase; KAS, 3-ketoacyl-ACP synthase; FAD2, oleoyl desaturase; FAD3, linoleoyl desaturase; FAD6, oleoyl desaturase on membrane glycerolipid; FAD7/8, linoleoyl desaturase on membrane glycerolipids; PC, phosphatidylcholine; G3P, glycerol-3-phosphate; LPA, lysophosphatidate; PA, phosphatidate; DAG, diacylglycerol; TAG, triacylglycerol; PDAT, phospholipid:diacylglycerol acyltransferase; DGAT, 1,2-sn-diacylglcyerol transferase

Acknowledgement

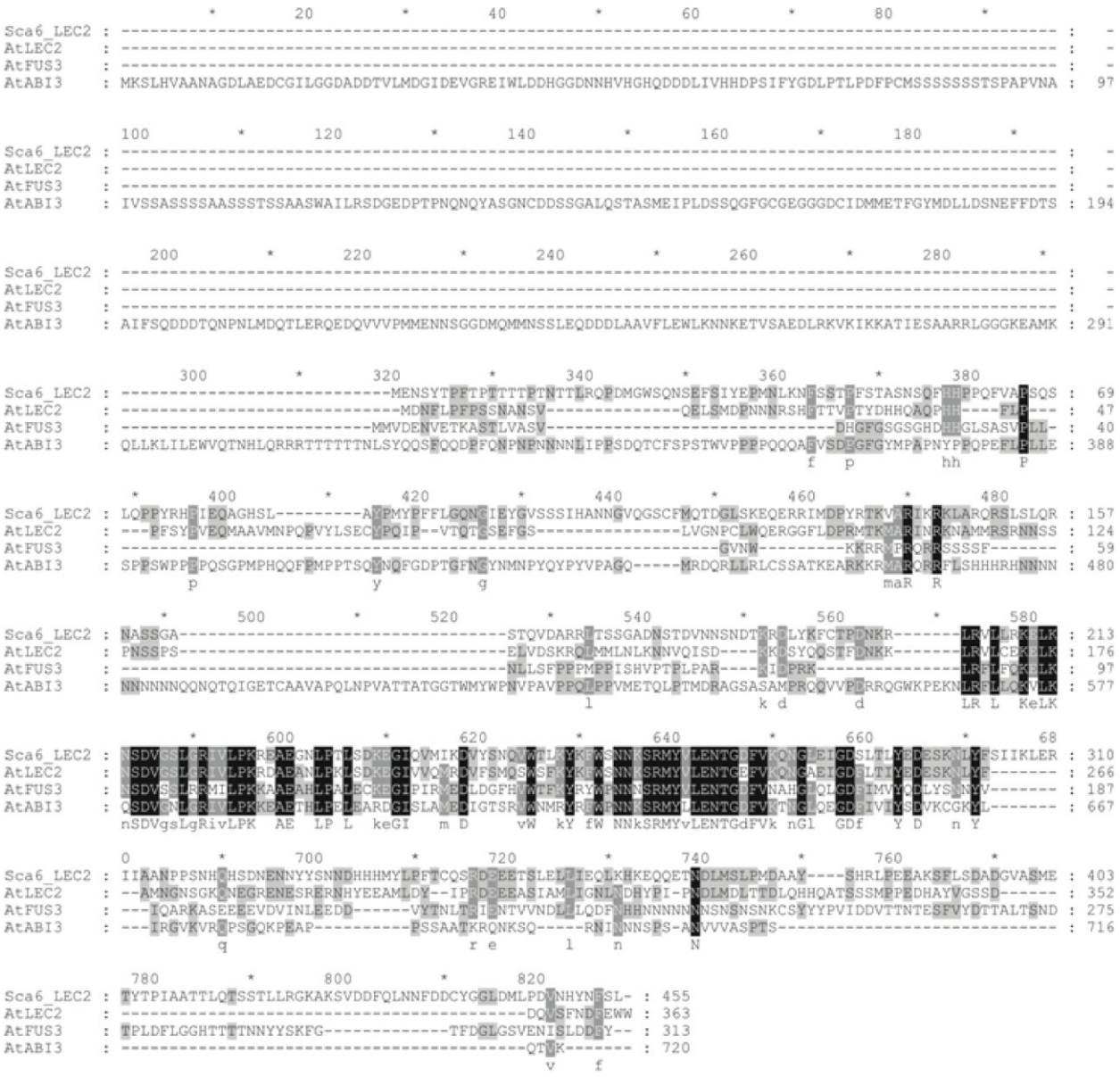
We would like to thank Dr. Phillip Smith from Proteomics and Mass Spectrometry Core Facility for the great help with GC analysis, Lena Landherr and Sharon Pishak for the technical assistance in maintenance of our cacao tissue culture pipeline. We are also grateful to Andrew Fisher for valuable comments to improve the manuscript. We thank Brian Irish of the USDA ARS located in Mayaguez, Puerto Rico, for provision of cacao pods. This work was supported in part by The Pennsylvania State University, College of Agricultural Sciences, The Huck Institutes of Life Sciences, the American Research Institute Penn State Endowed Program in the Molecular Biology of Cacao and a grant from the National Science Foundation BREAD program in cooperation with the Bill and Melinda Gates Foundation (NSF0965353) which supported the development of the transient gene expression assay for cacao.

2.7 Supplemental Data and Figures

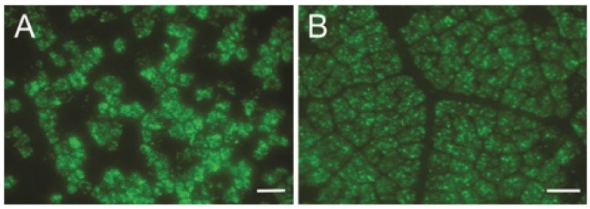
Supplemental Figure 2-1. Correspondent gene comparison from Criollo and Forastero genome database.

Criollo Genes	Forastero Genes	E-values
Tc06_g015590	Thecc1EG029838	0
Tc04_g004970	Thecc1EG017321	4.00E-159
Tc01_g024700	Thecc1EG003056	0
Tc09_g035210	Thecc1EG042290	0
Tc04_g016290	Thecc1EG019701	0
Tc06_g016770	Thecc1EG029976	0
Tc07_g007150	Thecc1EG031311	0
Tc00_g048490	Thecc1EG005014	0
Tc05_g010460	Thecc1EG022729	2.00E-105
Tc02_g034560	Thecc1EG011862	0
Tc03_g022470	Thecc1EG015590	0
Tc01_g015370	Thecc1EG040397	0
Tc04_g015240	Thecc1EG019508	0

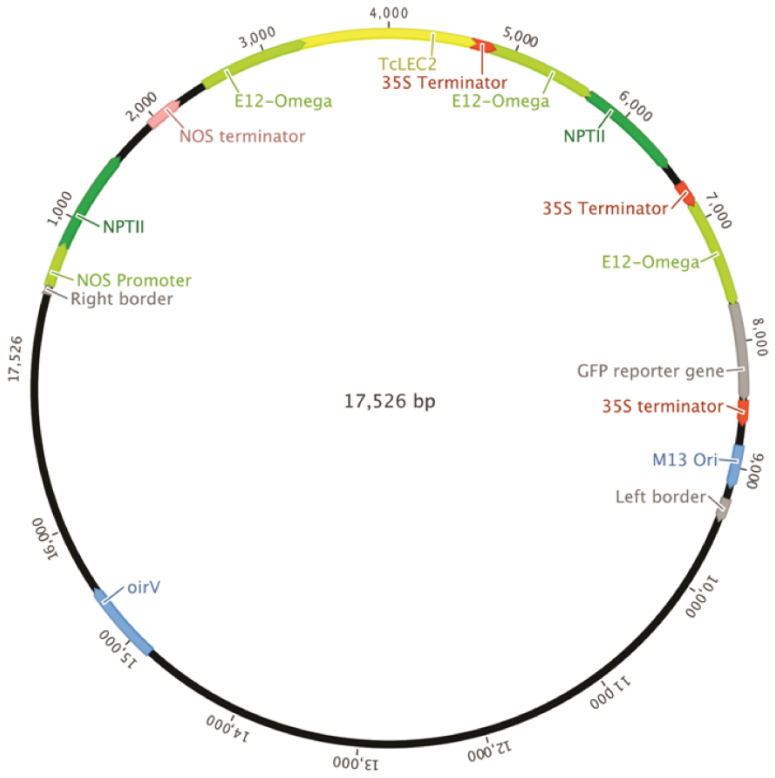
Supplemental Figure 2-1 Full-length amino acid alignment of TcLEC2, AtLEC2, AtFUS3, and AtABI3. Residues in black boxes are identical in all four proteins; residues in dark grey boxes are identical in three of four proteins; residues in light grey boxes are identical in two of four proteins.



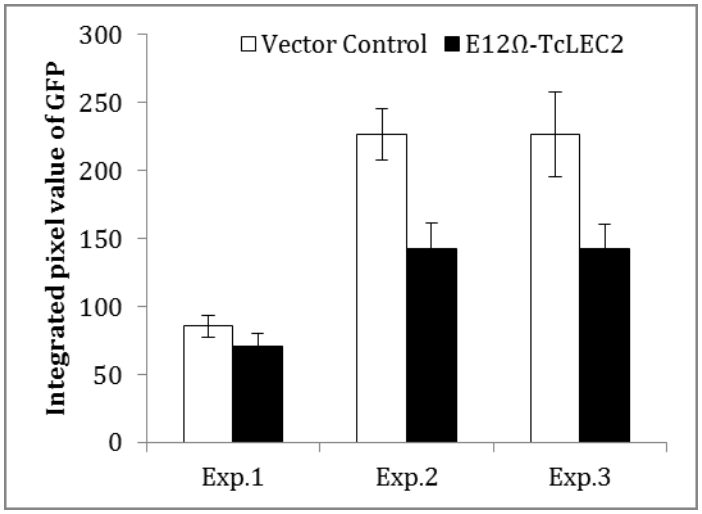
Supplemental Figure 2-2 Ectopic overexpression of control vector (pGH00.0126) and E12Ω::TcLEC2 in cacao attached leaf transient assay. Fluorescent micrographs of GFP expression (visualization marker) in leaves were captured three days after transformation (Bars = 0.4mm). A. GFP fluorescence image of cacao stage C leaves transformed with control vector. B. GFP fluorescence image of cacao stage C leaves transformed with E12Ω::TcLEC2.



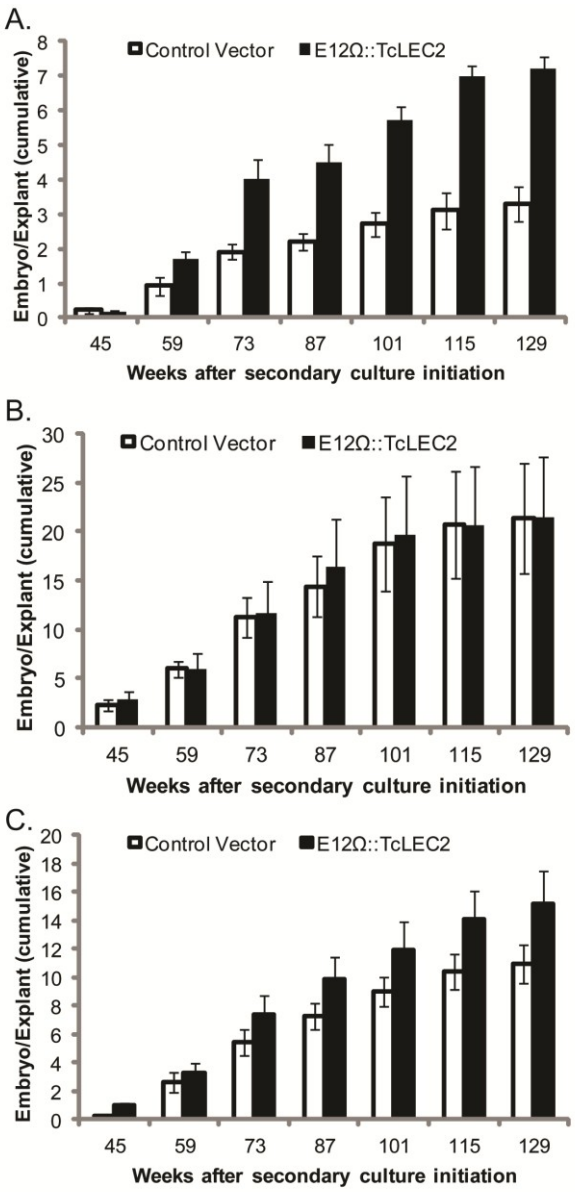
Supplemental Figure 2-3 Vector map of E12Ω-TcLEC2. Location of the TcLEC2 and GFP transgenes are indicated as are the NPTII selectable marker genes, and the location of all plant promoter and terminator elements. The control vector plasmid (pGH00.0126, GenBank: KF018690.1) is identical but lacks the E12Ω-TcLEC2-35S Terminator transgene segment.



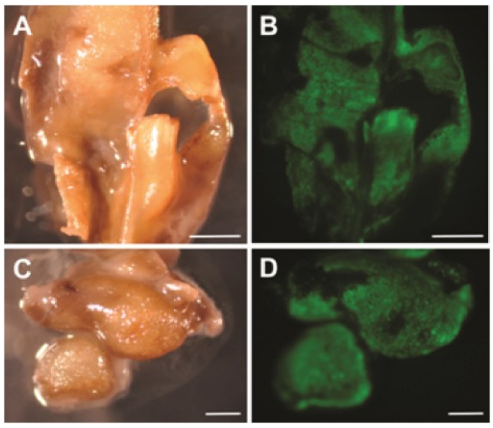
Supplemental Figure 2-5 Relative transient GFP expression levels of TcLEC2 transformation in SE compared to PSUSCA6.



Supplemental Figure 2-6 Comparison of average of total number of non-transgenic embryo produced per cotyledonary explant. Sixteen pieces of cotyeldonary explants were placed on each media plate. Three or four plates (taken as biological replicates) were used for transient transformation of control vector or E12Ω::TcLEC2 in each transformation trial (n=3 or 4, mean ± SE). A. Transformation trial 1 (n=3). B. Transformation trial 2 (n=4). C. Transformation trial 3 (n=4).



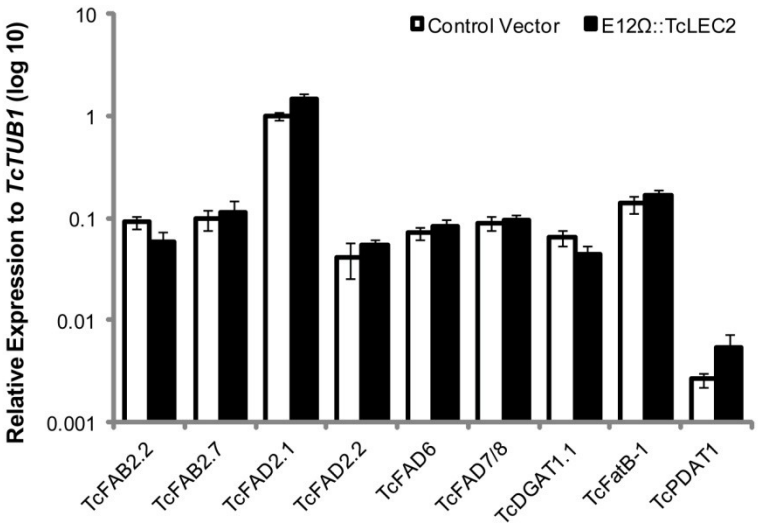
Supplemental Figure 2-7 Overexpression of control vector and E12Ω::TcLEC2 in cacao zygotic embryo transient assay. Fluorescent micrographs of GFP expression (visualization marker) in leaves were captured five days after transformation (Bars = 2mm). A&B. IZE transformed with control vector with white light and GFP fluorescence imaging. C&D. IZE transformed with E12Ω::TcLEC2 with white light and GFP fluorescence imaging.



Supplemental Figure 2-8 List of fatty acid biosynthesis related genes in cacao. The expression of these genes were compared in cacao IZE transiently overexpressing control vector and E12Ω::TcLEC2.

Gene	Tc_ID
TcKASII	*Tc09g006480*
TcFAB2.2	*Tc04g017510*
TcFAB2.7	*Tc08g012550*
TcFAD2.1	*Tc05g018800*
TcFAD2.2	*Tc01g015280*
TcFAD6	*Tc09g002840*
TcFAD7/8	*Tc05g002310*
TcFatA	*Tc01g022130*
TcFatB_1	*Tc09g010360*
TcFatB_2	*Tc01g022130*
TcFatB_5	*Tc03g015170*
TcDGAT1.1	*Tc09g007600*
TcDGAT2	*Tc01g000140*
TcPDAT1	*Tc09g029110*

Supplemental Figure 2-9 Expression levels of genes that are not significantly affected by transient overexpression of *TcLEC2* in cacao IZE compared to control vector (n=3, mean ± SE, significant levels were determined by t-test). The gene encoding TcWRI1 was also measured but no expression was detected.


Control Vector
E12Ω::TcLEC2
Relative Expression to TcTUB1 (log 10)
10
1
0.1
0.01
0.001
TcFAB2.2
TcFAB2.7
TcFAD2.1
TcFAD2.2
TcFAD6
TcFAD7/8
TcDGAT1.1
TcFatB-1
TcPDAT1

2.8 References

1. Sreenivasulu N, Wobus U: **Seed-development programs: a systems biology-based comparison between dicots and monocots**. *Annu Rev Plant Biol* 2013, **64**:189-217.
2. Romanel EA, Schrago CG, Counago RM, Russo CA, Alves-Ferreira M: **Evolution of the B3 DNA binding superfamily: new insights into REM family gene diversification**. *PLoS One* 2009, **4**(6):e5791.
3. Argout X, Salse J, Aury JM, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN *et al*: **The genome of Theobroma cacao**. *Nat Genet* 2011, **43**(2):101-108.
4. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool**. *J Mol Biol* 1990, **215**(3):403-410.
5. Motamayor JC, Mockaitis K, Schmutz J, Haiminen N, Livingstone D, 3rd, Cornejo O, Findley SD, Zheng P, Utro F, Royaert S *et al*: **The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color**. *Genome Biol* 2013, **14**(6):R53.
6. Mejia LC, Guiltinan MJ, Shi Z, Landherr L, Maximova SN: **Expression of Designed Antimicrobial Peptides in Theobroma cacao L. Trees Reduces Leaf Necrosis Caused by Phytophthora spp.** *Acs Sym Ser* 2012, **1095**:379-395.
7. Stone SL, Kwong LW, Yee KM, Pelletier J, Lepiniec L, Fischer RL, Goldberg RB, Harada JJ: **LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development**. *Proc Natl Acad Sci U S A* 2001, **98**(20):11806-11811.
8. Patel VK, Shanklin J, Furtek DB: **Changes in fatty-acid composition and stearoyl-acyl carrier protein desaturase expression in developing Theobroma-cacao L embryos**. *Planta* 1994, **193**(1):83-88.

9. Costa LM, Gutierrez-Marcos JF, Dickinson HG: **More than a yolk: the short life and complex times of the plant endosperm**. *Trends in Plant Science* 2004, **9**(10):507-514.
10. Shi Z, Zhang Y, Maximova S, Guiltinan MJ: **TcNPR3 from Theobroma cacao functions as a repressor of the pathogen defense response**. *BMC Plant Biology* 2013, **13**:204.
11. Harding EW, Tang WN, Nichols KW, Fernandez DE, Perry SE: **Expression and maintenance of embryogenic potential is enhanced through constitutive expression of AGAMOUS-Like 15**. *Plant Physiol* 2003, **133**(2):653-663.
12. Zheng YM, Ren N, Wang H, Stromberg AJ, Perry SE: **Global Identification of targets of the Arabidopsis MADS domain protein AGAMOUS-Like15**. *Plant Cell* 2009, **21**(9):2563-2577.
13. Nambara E, Naito S, Mccourt P: **A mutant of arabidopsis which is defective in seed development and storage protein accumulation is a new ABI3 allele**. *Plant Journal* 1992, **2**(4):435-441.
14. Monke G, Seifert M, Keilwagen J, Mohr M, Grosse I, Hahnel U, Junker A, Weisshaar B, Conrad U, Baumlein H *et al*: **Toward the identification and regulation of the Arabidopsis thaliana ABI3 regulon**. *Nucleic Acids Res* 2012, **40**(17):8240-8254.
15. Luerssen K, Kirik V, Herrmann P, Misera S: **FUSCA3 encodes a protein with a conserved VP1/ABI3-like B3 domain which is of functional importance for the regulation of seed maturation in Arabidopsis thaliana**. *Plant Journal* 1998, **15**(6):755-764.
16. Yamamoto-Toyoda A, Kagaya Y, Tanaka S, Tsutsumida K, Kagaya M, Kojima M, Sakakibara H, Hattori T: **The mechanism of the regulation of seed maturation by the FUS3 transcription factor revealed by transcriptome analysis**. *Plant Cell Physiol* 2007, **48**:S190-S190.
17. Maeo K, Tokuda T, Ayame A, Mitsui N, Kawai T, Tsukagoshi H, Ishiguro S, Nakamura K: **An AP2-type transcription factor, WRINKLED1, of Arabidopsis thaliana binds to the AW-box sequence conserved among proximal upstream regions of genes involved in fatty acid synthesis**. *The Plant Journal* 2009, **60**(3):476-487.

18. Baud S, Mendoza MS, To A, Harscoet E, Lepiniec L, Dubreucq B: **WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in Arabidopsis**. *Plant Journal* 2007, **50**(5):825-838.

19. Maximova SN, Alemanno L, Young A, Ferriere N, Traore A, Guiltinan MJ: **Efficiency, genotypic variability, and cellular origin of primary and secondary somatic embryogenesis of Theobroma cacao L.** *In Vitro Cell Dev-Pl* 2002, **38**(3):252-259.

20. She MY, Yin GX, Li JR, Li X, Du LP, Ma WJ, Ye XG: **Efficient Regeneration Potential is Closely Related to Auxin Exposure Time and Catalase Metabolism During the Somatic Embryogenesis of Immature Embryos in Triticum aestivum L**. *Mol Biotechnol* 2013, **54**(2):451-460.

21. Delporte F, Muhovski Y, Pretova A, Watillon B: **Analysis of expression profiles of selected genes associated with the regenerative property and the receptivity to gene transfer during somatic embryogenesis in Triticum aestivum L.** *Mol Biol Rep* 2013, **40**(10):5883-5906.

22. Maximova S, Miller C, Antunez de Mayolo G, Pishak S, Young A, Guiltinan MJ: **Stable transformation of Theobroma cacao L. and influence of matrix attachment regions on GFP expression**. *Plant Cell Rep* 2003, **21**(9):872-883.

23. Mejia L, Guiltinan M, Shi Z, Landherr L, Maximova S: **Expression of designed antimicrobial peptides in *Theobroma cacao* L. trees reduces leaf necrosis caused by *Phytophthora* spp**. In: *Acs Sym Ser.* vol. 1095: American Chemical Society; 2012: 379-395.

24. Maximova SN, Marelli JP, Young A, Pishak S, Verica JA, Guiltinan MJ: **Over-expression of a cacao class I chitinase gene in Theobroma cacao L. enhances resistance against the pathogen, Colletotrichum gloeosporioides**. *Planta* 2006, **224**:740-749.

25. Stone SL, Braybrook SA, Paula SL, Kwong LW, Meuser J, Pelletier J, Hsieh TF, Fischer RL, Goldberg RB, Harada JJ: **Arabidopsis LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis**. *Proc Natl Acad Sci U S A* 2008, **105**(8):3151-3156.

26. Santos Mendoza M, Dubreucq B, Miquel M, Caboche M, Lepiniec L: **LEAFY COTYLEDON 2 activation is sufficient to trigger the accumulation of oil and seed specific mRNAs in Arabidopsis leaves**. *FEBS Letters* 2005, **579**(21):4666-4670.

27. Wu G-Z, Xue H-W: **Arabidopsis β-ketoacyl-[acyl carrier protein] synthase I is crucial for fatty acid synthesis and plays a role in chloroplast division and embryo development**. *The Plant Cell Online* 2010, **22**(11):3726-3744.

28. Moreno-Perez AJ, Venegas-Caleron M, Vaistij FE, Salas JJ, Larson TR, Garces R, Graham IA, Martinez-Force E: **Reduced expression of FatA thioesterases in Arabidopsis affects the oil content and fatty acid composition of the seeds**. *Planta* 2012, **235**(3):629-639.

29. Lung SC, Weselake RJ: **Diacylglycerol acyltransferase: a key mediator of plant triacylglycerol synthesis**. *Lipids* 2006, **41**(12):1073-1088.

30. Barry-Etienne D, Bertrand B, Vasquez N, Etienne H: **Comparison of somatic embryogenesis-derived coffee (Coffea arabica L.) plantlets regenerated in vitro or ex vitro: morphological, mineral and water characteristics**. *Ann Bot* 2002, **90**(1):77-85.

31. Kita Y, Nishizawa K, Takahashi M, Kitayama M, Ishimoto M: **Genetic improvement of the somatic embryogenesis and regeneration in soybean and transformation of the improved breeding lines**. *Plant Cell Rep* 2007, **26**(4):439-447.

32. Samaj J, Baluska F, Pretova A, Volkmann D: **Auxin deprivation induces a developmental switch in maize somatic embryogenesis involving redistribution of microtubules and actin filaments from endoplasmic to cortical cytoskeletal arrays**. *Plant Cell Rep* 2003, **21**(10):940-945.

33. Tan H, Yang X, Zhang F, Zheng X, Qu C, Mu J, Fu F, Li J, Guan R, Zhang H *et al*: **Enhanced seed oil production in canola by conditional expression of Brassica napus LEAFY COTYLEDON1 and LEC1-LIKE in developing seeds**. *Plant Physiol* 2011, **156**(3):1577-1588.

34. Barthet VJ: **(n-7) and (n-9) cis-Monounsaturated fatty acid contents of 12 Brassica species**. *Phytochemistry* 2008, **69**(2):411-417.

35. Zheng P, Allen WB, Roesler K, Williams ME, Zhang S, Li J, Glassman K, Ranch J, Nubel D, Solawetz W *et al*: **A phenylalanine in DGAT is a key determinant of oil content and composition in maize**. *Nat Genet* 2008, **40**(3):367-372.

36. Shockey JM, Gidda SK, Chapital DC, Kuan JC, Dhanoa PK, Bland JM, Rothstein SJ, Mullen RT, Dyer JM: **Tung tree DGAT1 and DGAT2 have nonredundant functions in triacylglycerol biosynthesis and are localized to different subdomains of the endoplasmic reticulum**. *Plant Cell* 2006, **18**(9):2294-2313.

37. Zhang Y: **Recombinant expression of plant diacylglycerol acyltransferases from tissues that accumulate saturated fatty acids**. Edmonton, Alberta: University of Alberta; 2012.

38. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput**. *Nucleic Acids Res* 2004, **32**(5):1792-1797.

39. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0**. *Mol Biol Evol* 2007, **24**(8):1596-1599.

40. Lazo GR, Stein PA, Ludwig RA: **A DNA transformation-competent Arabidopsis genomic library in Agrobacterium**. *Biotechnology (N Y)* 1991, **9**(10):963-967.

41. Maximova SN, Dandekar AM, Guiltinan MJ: **Investigation of Agrobacterium-mediated transformation of apple using green fluorescent protein: high transient expression and low stable transformation suggest that factors other than T-DNA transfer are rate-limiting**. *Plant Mol Biol* 1998, **37**(3):549-559.

42. Shi Z, Maximova SN, Liu Y, Verica J, Guiltinan MJ: **Functional analysis of the Theobroma cacao NPR1 gene in Arabidopsis**. *BMC Plant Biol* 2010, **10**:248.

43. Li ZJ, Traore A, Maximova S, Guiltinan MJ: **Somatic embryogenesis and plant regeneration from floral explants of cacao (Theobroma cacao L.) using thidiazuron**. *In Vitro Cell Dev-Pl* 1998, **34**(4):293-299.

44. Stein SE: **NIST/EPA/NIH mass spectral library and NIST mass spectral search program**. 2008.

CHAPTER 3: CHARACTERIZATION OF A STEAROYL-ACYL CARRIER PROTEIN DESATURASE GENE FAMILY FROM CHOCOLATE TREE, *THEOBROMA CACAO* L.

Yufan Zhang[a, b], Siela Maximova[a, b], and Mark Guiltinan[a,b,1]

[a]Huck Institute of Life Sciences, The Pennsylvania State University, University Park, PA 16802

[b]The Department of Plant Science, The Pennsylvania State University, University Park, PA 16802

[1]Corresponding author

3.1 Abstract

In plants, the conversion of stearoyl-ACP to oleoyol-ACP is catalyzed by a plastid-localized soluble stearoyl-acyl carrier protein (ACP) desaturase (SAD). The activity of SAD significantly impacts the ratio between saturated and unsaturated fatty acids, and fatty acid composition. The cacao genome contains eight putative SAD isoforms with high amino acid sequence similarities and functional domain conservation with SAD genes from other species. Sequence variation in known functional domains between different SAD family members suggested that these eight SAD isoforms might have distinct functions in plant development, a hypothesis supported by their diverse expression patterns in various cacao tissues. Notably, *TcSAD1* is universally expressed across all tissues, and its expression pattern in seeds is highly correlated with the dramatic change in fatty acid composition during seed maturation. Interestingly, *TcSAD3* and *TcSAD4* appear to be exclusively and highly expressed in flowers, functions of which remain unknown. To test the function of *TcSAD1* in vivo, transgenic complementation of the Arabidopsis *ssi2* mutant was performed,

which demonstrated that TcSAD1 successfully rescued all AtSSI2 related phenotypes, further supporting the functional orthology between these two genes. The identification of the major *SAD* gene responsible for cocoa butter biosynthesis provides new strategies for screening for novel genotypes with desirable fatty acid compositions, and for use in breeding programs, to help pyramid genes for quality and other traits such as disease resistance.

3.2 Introduction

Current knowledge of the function of stearoyl-acyl carrier protein-desaturase (SAD) in various plant species was earlier reviewed in **Section 1.3**. The cacao genome was previously shown to contain 8 putative *SAD* genes (Argout et al., 2011). In the present study, we examined the gene family in detail, and explored the expression patterns of each *SAD* gene in various cacao tissues. A single gene primarily involved in the synthesis of 18:1 pools in developing cacao seeds was identified and functionally characterized via complementation of the Arabidopsis *ssi2* mutant. This information can be used to develop biomarkers for screening and breeding of new cacao varieties with novel fatty acid compositions of cocoa butter.

3.3 Material and Methods

Gene identification, phylogenetic analysis and protein prediction

SAD isoforms in *Theobroma cacao* were identified by blastp (Altschul et al., 1990) using full-length amino acid sequence of Arabidopsis AtSSI2 (At2g43710) as the query (E-value cut off $1e^{-5}$). Multiple protein sequence alignment was performed by MUSCLE (Edgar, 2004). The phylogenetic tree was constructed by MEGA4.1 using neighbor-joining algorithm with Poisson correction model and pairwise deletion (Tamura et al., 2007). Bootstrap values represent 1000 replicates. The phylogenetic tree was rooted using the amino acid sequence of PpSAD (XP_001772953) from *Physcomitrella patens*. Molecular mass and isoelectric point of SAD isozymes were predicted on ExPASy server (http://web.expasy.org/compute_pi/) (Gasteiger et al., 2005). Transient signal peptides

were predicted using ChloroP 1.1 server (Emanuelsson et al., 1999) and TargetP 1.1 server (Emanuelsson et al., 2000).

Fatty acid profiling by GC-MS

Plant tissue from glycerol treated and control leaves (four biological replicates) were ground in liquid nitrogen and fatty acid methyl esters (FAME) were prepared from each sample using approximately 30 mg of tissue per sample. Briefly, 1 ml of a MeOH/fuming HCl/Dichloromethane (10:1:1 v/v) solution was added to each tissue sample, which was incubated without shaking at 80˚C for 2h. Fatty acid methyl esters were re-extracted in 1 ml buffer H_2O/Hexane/Dichloromethane (5/4/1, v/v) with vortexing for 1 min. The hexane (upper phase) was separated by centrifugation at 1500 g for 5 min, transferred to glass GC vials (Agilent) and evaporated to dryness under vacuum. The FAMEs were then dissolved in 500 µl hexane for GC-MS analysis. Pentadecanoic acid (C15:0) (Sigma, Cat. P6125) was used as the internal standard added prior to the extraction and methyl nonadecanoate (C19:0-methyl ester) (Sigma, Cat. N5377) was used as the spike control, added into the sample prior to the GC injection. Samples were analyzed on an Agilent 6890N gas chromatograph coupled to a Waters GCT time of flight mass spectrometer. Mass spectra were acquired in electron ionization mode (70 eV) from 45 to 500 Da at a rate of one scan/second. The samples were separated on an Omegawax® 250 Capillary GC column (30 m x 0.25mm 0.25 uM phase thickness, Sigma, Cat. 24136) using helium at a constant flow of 1.0 ml/min. The initial oven temperature was 100˚C held for 1 min then increased at 15˚C/min to a temperature of 150˚C, and then increased at 4˚C/min to a final temperature of 280˚C. Samples (1 ul) were injected onto the column using a split/splitless injector maintained at 240˚C with a split ratio of 50/1.

RNA extraction and RT-qPCR analysis of gene expression

Plant tissues were first ground in liquid nitrogen. Total RNA was extracted using Plant RNA Purification Reagent (Life Technologies, Cat. 12322-012, following the manufacturer's instructions). The concentration of RNA was measured using a Nanodrop 2000c (Thermo Scientific). 500 ng of RNA was further treated with RQ1 RNase-free DNase (Promega, Cat. M6101) at 37°C for 30 min to remove potential genomic DNA contamination (following the

manufacturer's protocol). The treated RNA was reverse-transcribed by M-MuLV Reverse Transcriptase (New England Biolabs) with oligo-dT_{15} primers to obtain cDNA. RT-qPCR was performed in a total reaction volume of 10 µl containing 4 µl diluted-cDNA (1:50), 5 µl SYBR Green PCR Master Mix (Takara), 0.2 µl ROX (Takara, an internal reference dye to correct for fluorescent fluctuation), and 0.4 µl each 5 µM primers. Each reaction was performed in duplicates using Roche Applied Biosystem Step One Plus Realtime PCR System under the following program: 15 min at 94˚C, 40 cycle of 15 s at 94˚C, 20 s at 60˚C, and 40 s at 72˚C. The specificity of the primer pair was verified by visualization of the PCR product on a 2% agarose gel and analysis of the qPCR melting curve. A tubulin gene (*Tc06g000360*, *TcTUB1*(Argout et al., 2011), *TcTUB1-5'*: GGAGGAGTCTCTATAAGCTTGCAGTTGG and *TcTUB1-3'*: ACATAAGCATAGCCAGCTAGAGCCAG) and a gene encoding an acyl-carrier protein (*Tc01g039970*, *TcACP1* (Argout, 2011 #12), *TcACP1-5'*: GGAAAGCAAGGGTGTCTCGTTGAA and *TcACP1-3'*: GCGAGTTGAAATCTGCTGTTGTTTGG) were used as reference genes.

Characterization of Arabidopsis mutants and Arabidopsis transformation

All Arabidopsis plants were grown in a Conviron growth chamber at 22°C with 16h light/8h dark cycle. Arabidopsis *ssi2* mutant was kindly donated by Dr. Kachroo (University of Kentucky), and homozygous *fab2* mutant (SALK_036854) was obtained from the Arabidopsis Biological Resource Center at Ohio State University (https://abrc.osu.edu/). The base pair mutation of *AtSSI2* in *ssi2* mutant was confirmed by sequencing the PCR product amplified by the following primers (LP1: TGAAGAAACCATTTACGCCAC; p3: CGTGTTGACATGAGGCAGATCG). The presence and homozygosity of T-DNA insertion in the *fab2* mutant line were confirmed by genotyping using the following primers (LB: CTTTGACGTTGGAGTCCAC; Up1F: TGAAACAGGTGCTAGTCCTACTTCA; Dn1R: CACCTGAAAGCCCGGTTAAGTC) (Schluter et al., 2011). Expression levels of *AtSSI2* in Col-0, *ssi2*, and *fab2* were examined by semi-quantitative RT-PCR using the following intron spanning primers (*AtSSI2-5'*: GGCCCCAAGGAGGTTGAGAG; *AtSSI2-3'*: ATCTGGAATGGATCCGCGGAC).

The full length coding sequence of *TcSAD1* was first amplified from leaf cDNA with the following primers (*TcSAD1-5'-XbaI*: GCTCTAGAATGGCTCTGAAATTGAATCCCAT and *TcSAD1-3'-XbaI*: TCTAGACTAGAGCTTCACTTCTCTATCAAAAATCCAAC). The resulting fragment was cloned into intermediate vector pE2113 using the *XbaI* site, driven under enhanced *E12-Ω CaMV 35S* promoter (Mitsuhara et al., 1996). The whole gene cassette was then cloned into binary pCAMBIA-1300 (Hajdukiewicz et al., 1994) using *EcoRI* and *HindIII* sites, resulting in plasmid pGZ14.1021 and was further transformed into *A. tumefaciens* strain *AGL1* by electroporation. Col-0, *ssi2*, and *fab2* Arabidopsis genotypes were transformed using the floral dip method (Clough and Bent, 1998). Transformants were selected on MS media containing hygromycin B (Harrison et al., 2006). Expression levels of *TcSAD1* in transgenic plants were confirmed by RT-PCR using the following primers: (*TcSAD1-RT-5'*: GCTCTGAAATTGAATCCCATCACTTCTCAA; *TcSAD1-RT-3'*: TGGCTCCAAAGAGGTTGAGAATGTC). Arabidopsis *Ubiquitin* (*AtUbi-5*': ACCGGCAAGACCATCACTCT; *AtUbi-3'*: AGGCCTCAACTGGTTGCTGT) was used as a reference gene.

3.4 Results and Discussion

The *T. cacao* contains eight highly conserved *TcSAD* isoforms in the genome

To identify the orthologous genes encoding SAD(s) in the cacao genome, the full-length amino acid sequence of AtSSI2 (At2g43710) was blasted against the predicted cacao proteome of the Belizean Criollo genotype (B97-61/B2) (http://cocoagendb.cirad.fr/ (Argout et al., 2011)) using blastp algorithm with E-value cut-off $1e^{-5}$ (Altschul et al., 1990). Eight putative *SAD* genes were identified in cacao genome with reliable sequencing data and predicted gene structures (Table 1). Interestingly, of these eight *SAD* isoforms, four genes are located on chromosomes 4, and three of them (*Tc04g017510*, *Tc04g017520*, *Tc04g017540*) are clustered except for one predicted transposable element gene in between, which is similar to in the Arabidopsis genome where three out of seven *SAD* isoforms (*At3g02610*, *At3g02620*, and *At3g02630*) are located in tandem on chromosome 3 (Kachroo et al., 2007). Multiple amino acid sequence comparison and phylogenetic analysis including SAD isoforms from cacao, castor and Arabidopsis revealed that Tc04g017510

(designated as TcSAD1) is closely clustered with RcSAD1 and AtSSI2 (Figure 3-1), both of which have significant impacts on oleic acid contents and fatty acid profiles in Arabidopsis and castor, respectively, suggesting a potential significant roles of TcSAD1 in cacao seed oil biosynthesis. Moreover, Tc05g012840 (designated as TcSAD2), Tc04g017520 (designated as TcSAD3) and Tc04g017540 (designated as TcSAD4) also exhibit high sequence similarities to RcSAD1, implying that they may also contribute to the activity of SAD in cacao.

The metabolic pathways in which SADs participate in are well known to occur in the lumen of plastids (Cahoon et al., 1997). Thus we performed bioinformatics analysis of transit peptides (TP) in the N-terminal sequences of the predicted cacao SAD proteins using the TargetP (Emanuelsson et al., 2000) and ChloroP (Emanuelsson et al., 1999) algorithms. AtSSI2 (At2g43710) in Arabidopsis and RcSAD1 (XP_002531889) in castor bean were also included in this analysis since the structures and the functions of these two proteins have been well-studied (Lindqvist et al., 1996; Cahoon et al., 1997; Kachroo et al., 2007; Whittle et al., 2008). High-confidence chloroplast transient peptides were successfully detected in AtSSI2 and RcSAD1, consistent with previous analysis (Kachroo et al., 2007; Schluter et al., 2011). Of eight cacao SAD isoforms, chloroplast transient peptides were identified in TcSAD1, TcSAD2, TcSAD5, and TcSAD7. A mitochondrial targeting peptide was predicted in TcSAD3 and a signal peptide related with the secretory pathway was detected in TcSAD4, suggesting that these isoforms may exhibit distinct functions. No transient peptides were detected in TcSAD6 and TcSAD8 (Table 2).

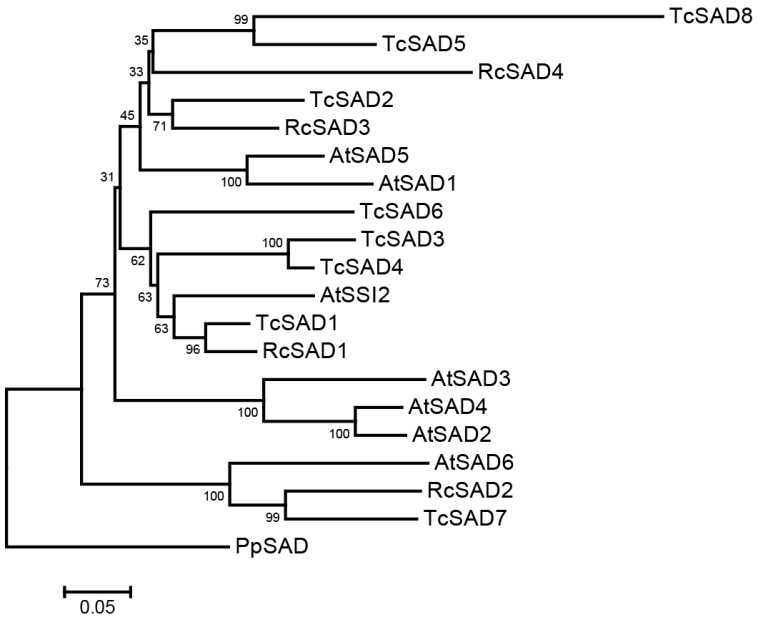


Figure 3-1 Neighbor-joining bootstrap tree phylogeny based on full-length amino acid sequences of stearoyl-acyl carrier protein-desaturase (SAD) gene families in Arabidopsis (AtSAD), castor (RcSAD), and cacao (TcSAD). The scale bar represents 0.05 estimated substitutions per residue and values next to nodes indicated bootstrap values from 1000 replicates. The phylogenetic tree was rooted by using the amino acid sequence of PpSAD (XP_001772953) from Physcomitrella patens.

Table 3-1 Blastp result using AtSSI2 (At2g43710) as the query.

Query	Locus ID	Designation	Expect	% Identity	Query Coverage	Hit Length
AtSSI2	Tc04g017510	TcSAD1	0	80.05	100	397
	Tc05g012840	TcSAD2	1E-177	80.91	87.53	394
	Tc04g017520	TcSAD3	1E-176	80.97	87.78	397
	Tc04g017540	TcSAD4	1E-174	80.86	87.28	366
	Tc04g005590	TcSAD5	1E-166	76.49	87.53	395

Tc09g024040	TcSAD6	1E-144	72.7	84.04	330
Tc08g012550	TcSAD7	1E-143	70.12	83.29	389
Tc01g009910	TcSAD8	2E-71	42.98	86.53	299

Table 3-2 Predicted protein characteristics of cacao, Arabidopsis, and castor SAD proteins. Transit peptide (TP) length and subcellular localization are indicated (C, Chloroplast, M, Mictochondrial, S, Secretory pathway).

Designation	Locus ID	pI (isoelectric point)	MW (kDa)	Length (a.a)	Score	TP-Length (a.a)	Predicted Localization
TcSAD1	Tc04g017510	6.05	45.39	396	0.564	33	C
TcSAD2	Tc05g012840	6.24	44.83	393	0.536	27	C
TcSAD3	Tc04g017520	5.87	45.31	396	0.508	33	M
TcSAD4	Tc04g017540	5.33	41.86	365	0.43	26	S
TcSAD5	Tc04g005590	6.51	44.76	394	0.546	28	C
TcSAD6	Tc09g024040	5.49	37.94	329	0.434	21	-
TcSAD7	Tc08g012550	6.39	44.08	388	0.538	38	C
TcSAD8	Tc01g009910	5.88	33.91	298	0.448	21	-
AtSSI2	At2g43710	6.61	45.65	401	0.587	35	C
RcSAD1	XP_002531889	6.19	45.37	396	0.557	33	C

SAD isozymes in *T. cacao* share conserved overall protein structures with noteworthy differences

The three-dimensional protein structure of RcSAD1 from castor seed has been extensively studied in detail as a model of Δ^9 desaturase in plants (Schneider et al., 1992; Fox et al., 1993; Lindqvist et al., 1996). To provide further insights into potential catalytic activities of SAD isozymes in *T. cacao*, a multiple amino acid sequence alignment was conducted including RcSAD1, AtSSI2 and eight cacao SADs, and the determinative amino acid residues that are highly associated with substrate binding activities and chain length specificities of the SADs were evaluated (Figure 3-2). As mentioned above, a high degree of variation was observed within putative transient signal peptides in N-terminal sequences. Consistent with the crystal structure of RcSAD1, cacao SADs also consist of eleven highly conserved α-helices, except that putative TcSAD8 lacks the region from α3b- to α5-helices, which might greatly compromise its normal functionality. Notably, a further detailed investigation of the determinative residues revealed that several prominent divergences exist within cacao SADs, implying that they might have distinct substrate preferences. For example, amino acid residues at positions of 117, 118, 189, and 206 (positions assigned according to castor RcSAD1 amino acid sequence, indicated in Figure 3-2) were reported to be crucial for the function of RcSAD1 as a Δ^9-18:0-ACP desaturase, and replacements of these residues converted RcSAD1 into an enzyme that either functioned as a Δ^6-16:0-ACP desaturase or converted 18:0-ACP into the allylic alcohol *trans*-isomer (E)-10-18:1-9-OH (Cahoon et al., 1997; Whittle et al., 2008). In this respect, TcSAD3 and TcSAD4 share the same amino acid residues at those positions which differ from RcSAD1, AtSSI2, and other cacao SADs, suggesting that TcSAD3 and TcSAD4 may exert different functions from the other cacao SADs. Likewise, at positions of 179 and 181, which are crucial for the substrate specificity of RcSAD1(Cahoon et al., 1997), TcSAD5, TcSAD6, and TcSAD8 contain unique varied amino acid residues compared to RcSAD1, AtSSI2, and the other cacao SADs.

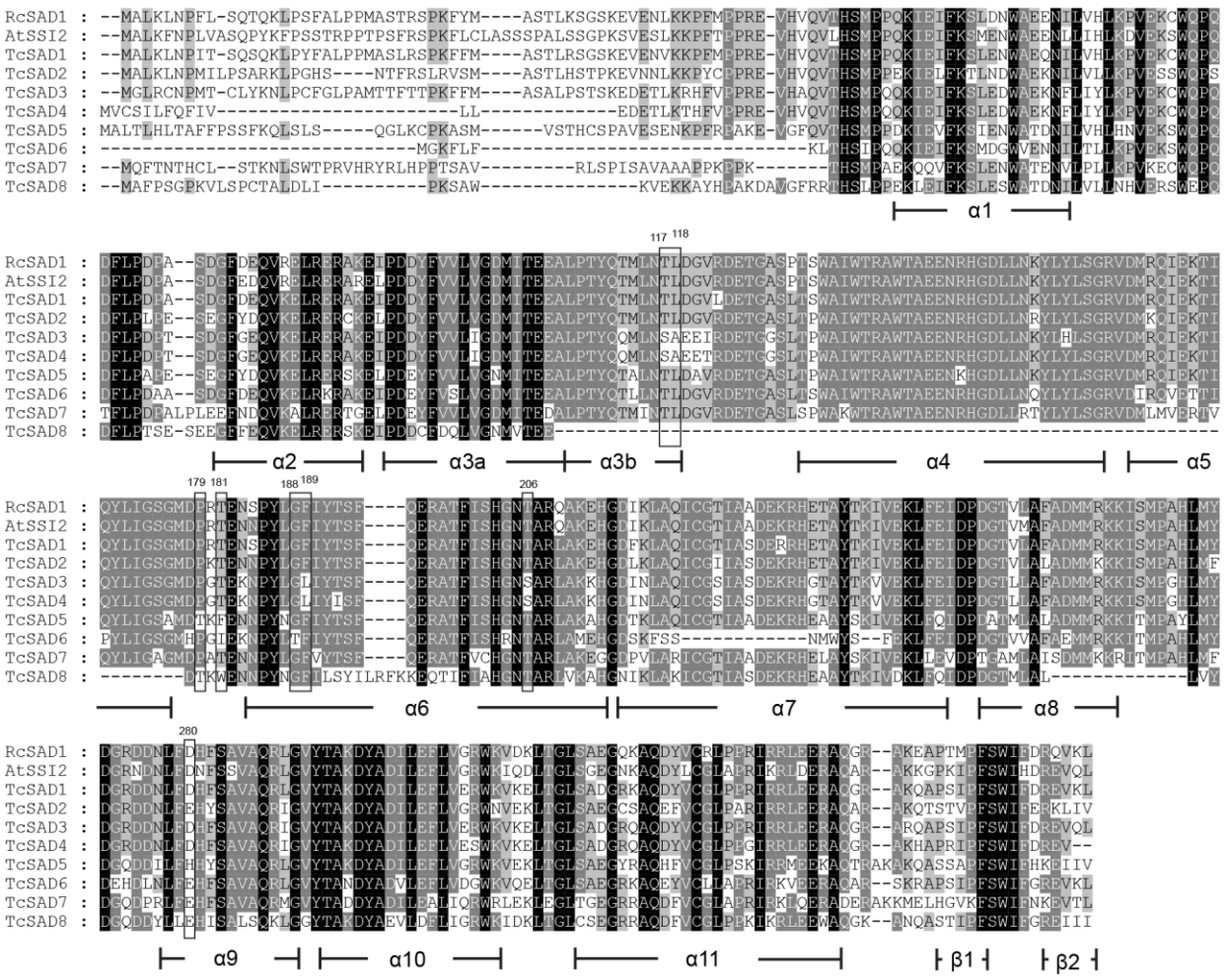


Figure 3-2 Multiple amino acid sequence alignment of TcSADs from cacao with SAD from castor bean (RcSAD1) and Arabidopsis (AtSSI2). Secondary protein structures of SAD were compared, deduced and annotated according to the crystal structure of RcSAD1. Shadings were performed by Genedoc software in a conservation mode. Residues in black represent 100% conservation of all sequences; residues in dark grey represent > 75% identity of all sequences; residues in light grey >50% identity of all sequences. Functional determinative amino acids were marked in rectangles and numbered according to the start codon of RcSAD1.

Tissue specific expression pattern of *TcSAD* isoforms in *T. cacao*

To further characterize functions of the cacao *SAD* family, their expression profiles were examined by qRT-PCR in various cacao tissues including: leaves at developmental stages A, C, and E (defined in (Mejia et al., 2012)), unopened flowers, open flowers, roots, and zygotic seeds at 14, 16, 18, and 20 weeks after pollination (WAP). Since *SAD* isoforms are highly conserved in their functional domains (α-helices and β-sheets), isoform specific qRT-PCR primers were designed to preferably target at predicted 5' and 3' untranslated regions, and signal transient peptide regions, all of which exhibited low sequence similarities among SAD isoforms in cacao (Table 3). The specificities of each set of primers were determined by examining amplification products on the agarose gel and further confirmed by melting point analysis (Data not shown). Expression levels of *SAD* isoforms were normalized to *TcTUB1* (*Tc06g000360*), which was previously identified as the most stably expressed gene across various cacao tissues (unpublished data). Among them, transcripts of *TcSAD1* and *TcSAD2* were detected in all the examined tissues, in which the expression levels of *TcSAD1* were higher than those of *TcSAD2* (Figure 3-3). Interestingly, *TcSAD3* and *TcSAD4* were predominantly expressed in unopened and open flowers, where the expression levels of *TcSAD3* and *TcSAD4* were significantly higher than any other SAD isoforms; however, the transcripts of both were barely detectable in all the other tissues, indicating that they are flower specific genes. The expression of *TcSAD5* was primarily detected in roots; however, at levels lower than those of *TcSAD1*, *TcSAD2*, and *TcSAD7* in roots. *TcSAD7* constitutively expresses in all examined tissues with higher expression in roots and zygotic seeds at all developmental stage (Figure 3-3). Overall, on the basis of gene expression data, both *TcSAD1* and *TcSAD7* contribute to the major proportions to the total SAD transcripts in developing zygotic embryos, with TcSAD7 being higher expressed over all the stages, implying that the activities of both TcSAD1 and TcSAD7 are possibly involved in the synthesis and accumulation of 18:1 in cacao embryos. Notably, the transcripts of *TcSAD6* and *TcSAD8* were barely detected in any of the tissues analyzed, suggesting that they might be pseudogenes in cacao genome or expressed in a tissue, stage of development or induction condition not tested.

Profoundly, *TcSAD3* and *TcSAD4* appear to be exclusively and highly expressed in floral tissues, and this specific expression pattern is consistent with the analyses of transient signal peptide (Table 2) and multiple sequence alignment (Figure 3-2), which suggested that TcSAD3 and TcSAD4 may target at mitochondrial and the secreted pathway, respectively, and comprise the same altered amino acids at those key positions. Interestingly, a previous study in orchid revealed that the activities of SAD isoforms were involved in the biosynthesis of alkene species with different double bond positions, which result in the major odor differences among orchid species attracting differential pollinators (Schluter et al., 2011). Moreover, a more recent study in tobacco showed that the normal function of one isoform of stearoyl-ACP desaturase was critical to maintain membrane lipid composition during ovule development and thus affected female fertility (Zhang et al., 2014). Therefore, even though functions of TcSAD3 and TcSAD4 in cacao floral tissues remain largely unknown, it is reasonable to speculate that their activity might be associated with the synthesis of fatty acid derived metabolites that are vital to the pollination process and the development of cacao reproductive tissues.

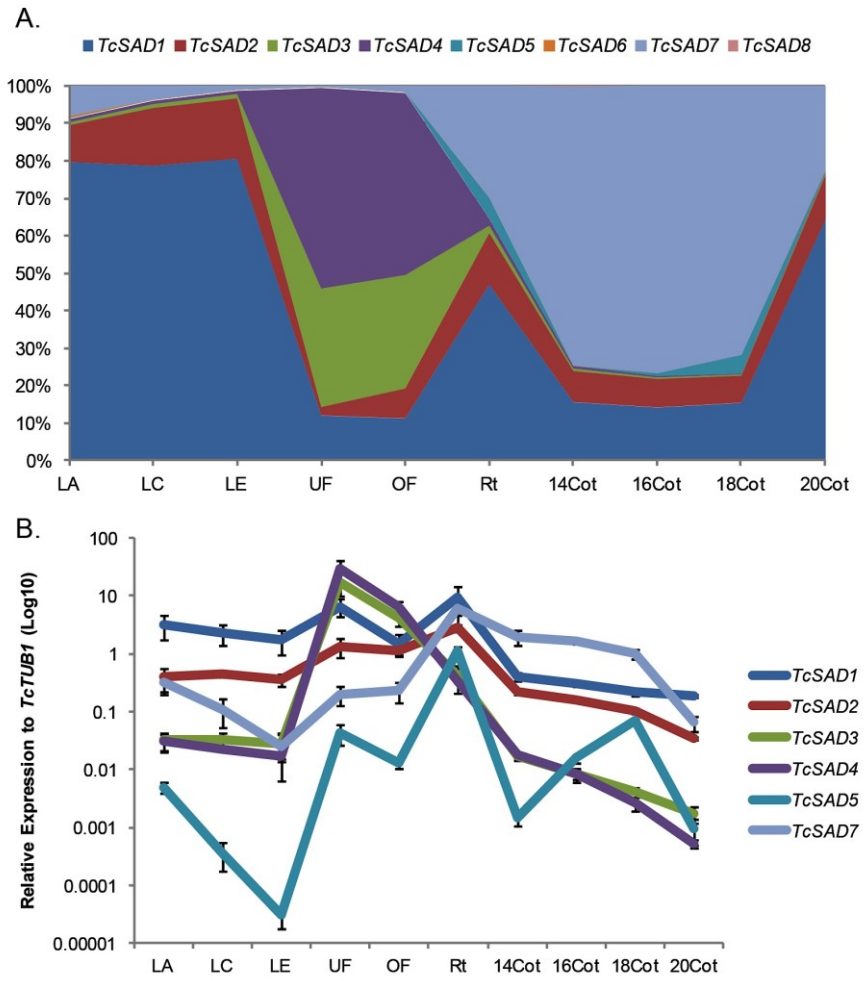


Figure 3-3 *TcSAD* expression patterns in various cacao tissues. Tissues include: leaves stage A, C and E (early, mid and late development, see (LA, LC, LE), developing flowers (UF) and mature flowers (OF), root (Rt), and zygotic cotyledons from seeds collected at 14, 16, 18, and 20 weeks after pollination (14Cot, 16Cot, 18Cot, and 20Cot). The expression levels were analyzed by qRT-PCR and normalized relative to that of *TcTUB1*. Three biological samples for each tissue were included in the analysis. (**A**) 100% stacked area plot representing the percentage of each *TcSAD* isoforms that contributes to the total expression levels of *TcSAD* isoforms in each tissue. (**B**) Line chart representing the relative expression level of each *TcSAD* isoform in various cacao tissues on the $\log_{10}$ scale. *TcSAD6* and *TcSAD8* were excluded because of their low expression levels.

Table 3-3 Summary of expression levels of TcSAD isoforms in various cacao tissues

Table 3-3 Summary of expression levels of *TcSAD* isoforms in various cacao tissues

	Expression Pattern	Predominance	LA	LC	LE	UF	OF	Rt	14Cot	16Cot	18Cot	20Cot
TcSAD1	Constitutive	Lower in embryos	3.1736 ± 1.4003	2.2747 ± 0.8641	1.7835 ± 0.7933	6.6441 ± 2.1348	1.5259 ± 0.6346	9.633 ± 4.888	0.4123 ± 0.0608	0.2997 ± 0.0191	0.2159 ± 0.0294	0.187 ± 0.0113
TcSAD2	Constitutive	Low	0.397 ± 0.1732	0.449 ± 0.0554	0.3587 ± 0.0806	1.3392 ± 0.465	1.1114 ± 0.1451	2.8314 ± 0.3767	0.2243 ± 0.0278	0.1643 ± 0.0041	0.1039 ± 0.0098	0.0345 ± 0.0017
TcSAD3	Flowers	~30% of total expression	0.0328 ± 0.0112	0.0328 ± 0.0096	0.0274 ± 0.0138	17.535 ± 7.4756	4.1346 ± 1.1218	0.4207 ± 0.2133	0.0172 ± 0.0023	0.0084 ± 0.0023	0.0042 ± 0.0007	0.0018 ± 0.0006
TcSAD4	Flowers	~40% of total expression	0.0309 ± 0.0102	0.0231 ± 0.0007	0.0174 ± 0.0016	29.6626 ± 11.4378	6.6223 ± 1.3677	0.351 ± 0.0339	0.0182 ± 0.0015	0.0085 ± 0.0017	0.0027 ± 0.0007	0.0005 ± 0.0001
TcSAD5	Root & Embryos	Low	0.005 ± 0.001	0.0004 ± 0.0002	0 ± 0	0.0434 ± 0.0164	0.0131 ± 0.0029	1.1332 ± 0.1956	0.0015 ± 0.0004	0.016 ± 0.0027	0.0711 ± 0.0026	0.0009 ± 0.0004
TcSAD6	---		0.0181 ± 0.0012	0.0019 ± 0.0005	0.0015 ± 0.0003	0.0064 ± 0.0029	0.0032 ± 0.0011	0.0042 ± 0.0007	0.0026 ± 0.0017	0.0008 ± 0.0003	0.0003 ± 0.0002	0.0003 ± 0.0001
TcSAD7	Constitutive	Higher in roots & embryos	0.3271 ± 0.1328	0.1102 ± 0.0555	0.0255 ± 0.019	0.1993 ± 0.0737	0.2312 ± 0.0889	6.1258 ± 3.4435	1.9744 ± 0.5346	1.6229 ± 0.1256	1.0066 ± 0.1921	0.0656 ± 0.0191
TcSAD8	---		0 ± 0	0 ± 0	0 ± 0	0.0001 ± 0.0001	0 ± 0	0.0001 ± 0.0001	0.0013 ± 0.0018	0.0001 ± 0.0001	0 ± 0	0 ± 0

Table 3-4 Primer sequences used in RT-qPCR for *TcSAD* expression pattern analysis.

Gene	Position of primers	Primer sequences	Amplicon size
TcSAD1	3'UTR	5'-GTCACTGTGTGGCAAGTGTCTTCT-3'	136 bp
	3'UTR	5'-CCACACACCTCCCAAACAATTCCT-3'	
TcSAD2	3'UTR	5'-TGAGGAGATGGTTGCAGTAGGTCA-3'	142 bp
	3'UTR	5'-ACGAGACAGGAAACGCTTACCGA-3'	
TcSAD3	CDS	5'-AATTAGAAGGCTGGAGGAGAGAGC-3'	146 bp
	3'UTR	5'-GAGAAGACACTTGCAGGGAGAAGA-3'	
TcSAD4	3'UTR	5'-TTTCCTCTCCCTGCAAGTGTCTTC-3'	82 bp
	3'UTR	5'-GGATCTATTCTTTGTGGGTTGACCAG-3'	
TcSAD5	CDS	5'-ACAGGTCTTTCTGCCGAGGGTTAT-3'	136 bp
	CDS	5'-AAATCCAGCTGAAAGGAGCGCTTG-3'	
TcSAD6	5'UTR	5'-CAAGCGGCCAATTAAAGGTTGAG-3'	92 bp
	CDS	5'-TGTTGTGGGATGGAGTGAGT-3'	
TcSAD7	CDS	5'-ACGAGCGGGCCAAGAAGAT-3'	99 bp
	3'UTR	5'- AAGCCGCCTTGCCTCCTTCATTT-3'	
TcSAD8	CDS	5'-TCCAAAGTCTGCATGGAAGGTGGA-3'	81 bp
	CDS	5'-AGGGAATGAGTCCTTCGAAACCCA-3'	

Oleic acid accumulation and *SAD* isoform expression during cacao seed development

Notably, the maturation of cacao seeds initiates from the axial ends and gradually towards the cotyledonary ends. Seeds during this rapid transition are ideal to evaluate the correlation between oleic acid contents and the activities of *TcSAD*s since biological variations are greatly minimized. To explore the role of each TcSAD isoform in seed lipid

biosynthesis, maturing cacao seeds were isolated from fruit at 12 WAP (Figure 3-4A), and the entire cacao seed was transversely dissected into six segments along the developmental gradient, with the least developed segment (segment 1, color coded as dark green, Figure 3-4A) at the cotyledonary end and the most developed segment (segment 6, color coded as dark red, Figure 3-4A) at the axial end. The analysis of fatty acid composition in each segment indicated that the percentage of $18:1^{n-9}$ gradually increased from 12% in segment 1 to 35% in segment 6 (Figure 3-4B), which is consistent with previous reports of fatty acid composition changes during cacao seed development over time (Patel et al., 1994). Thereafter, the expression of two major *TcSAD* isoforms in seeds (*TcSAD1* and *TcSAD7*, Figure 3-3) was examined in each segment by qRT-PCR, and the correlation between the expression levels of *TcSAD*s and fatty acid composition was evaluated by linear regression (Figure 3-4C). The expression levels of *TcSAD*s were normalized to *TcACP1* (*Tc01g039970*), which was previously identified as the most stably expressed gene during cacao seed development (unpublished data). In general, the levels of *SAD* gene expression shown in Figure 3-4C suggested that *TcSAD1* expression (left panel) was highly correlated with gradual alterations of fatty acid composition during development. However, expression of *TcSAD7* did not correlate well with fatty acid composition changes (right panel). Specifically, *TcSAD1* expression was positively correlated with the molar percentage of $18:1^{n-9}$ (r = 0.799) and 18:0 (r = 0.522), and negatively correlated with the molar percentage of $18:2^{n-6}$ (r = -0.638); however, poor correlations were observed between the expression of *TcSAD7* and these fatty acid species. Additionally, the expression of *TcSAD1* was also positively correlated with the level of total molar percentage of unsaturated fatty acids ($18:1^{n-9}$ + $18:2^{n-6}$ + $18:3^{n-3}$) (r = 0.438) and the ratio of unsaturated fatty acid (as above) over saturated fatty acid (16:0 + 18:0) (r = 0.464). Again, the expression of *TcSAD7* was poorly correlated with these two aspects. Taken together, the evidence suggested that the increased expression of *TcSAD1* during the 12 WAP transition period of embryo development is a critical factor determining the increased rate of 18:1 synthesis and accumulation. Based on the correlation analysis, we could not rule out the significant role of TcSAD7 in 18:1 synthesis and accumulation throughout the maturation of the embryo; however, its level does not increase specifically rapidly during the transition period of embryo development at 12 WAP.

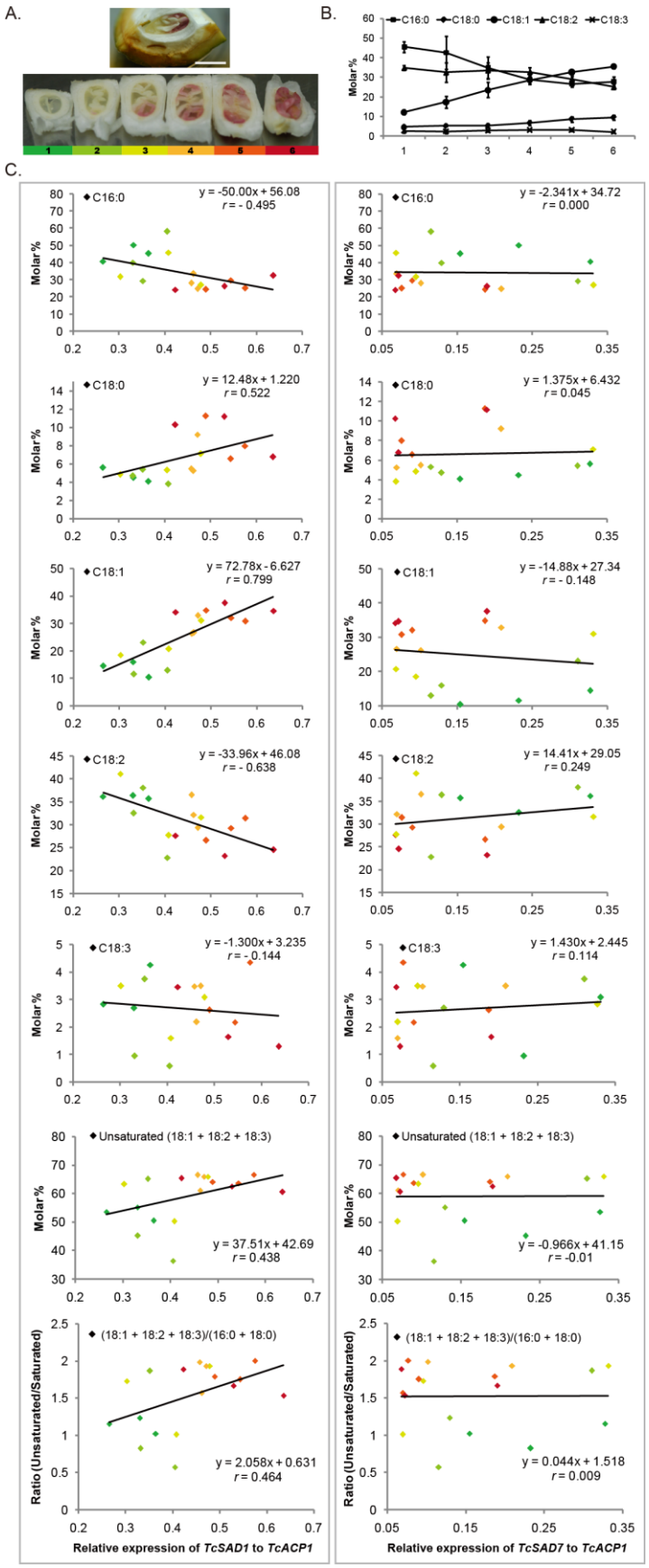

A.
1
2
3
4
5
6
B.
C16:0
C18:0
C18:1
C18:2
C18:3
Molar%
C.
C16:0
y = -50.00x + 56.08
r = - 0.495
y = -2.341x + 34.72
r = 0.000
C18:0
y = 12.48x + 1.220
r = 0.522
y = 1.375x + 6.432
r = 0.045
C18:1
y = 72.78x - 6.627
r = 0.799
y = -14.88x + 27.34
r = - 0.148
C18:2
y = -33.96x + 46.08
r = - 0.638
y = 14.41x + 29.05
r = 0.249
C18:3
y = -1.300x + 3.235
r = - 0.144
y = 1.430x + 2.445
r = 0.114
Unsaturated (18:1 + 18:2 + 18:3)
y = 37.51x + 42.69
r = 0.438
y = -0.966x + 41.15
r = -0.01
(18:1 + 18:2 + 18:3)/(16:0 + 18:0)
Ratio (Unsaturated/Saturated)
y = 2.058x + 0.631
r = 0.464
y = 0.044x + 1.518
r = 0.009
Relative expression of TcSAD1 to TcACP1
Relative expression of TcSAD7 to TcACP1

Figure 3-4 *TcSAD* relative expression and fatty acid composition in maturing cacao seeds. (**A**) (Upper) Mid-stage cacao seed sliced longitudinally, within ovule and exocarp tissue (pod shell) 12 weeks after pollination. The embryo root axial meristem is to the right, expanding cotyledonary tissues to the left. (Lower) Six radial serial thick sections dissected from a single cacao seed oriented as in the top image, then rotated 90 degrees to display the cross sections. Each section is sequentially numbered from left to right and color coded by green, light green, yellow, orange, light red, red, respectively, from youngest cotyledonary end (left) to the most mature axial end (right). Scale bar = 1cm. (**B**) Molar percentages of fatty acid compositions in each seed section. (n= 3, mean ± SE). (**C**) The correlation of *TcSAD1* and *TcSAD7* gene expression and molar percentage of fatty acid composition in various cross sections of developing cacao seeds. Left column, *TcSAD1;* right column, *TcSAD7*. Color-coded data points represent different seed sections as labeled in (**A**). Three biological replicates from each seed section were analyzed. Each independent measurement is shown as a separate data point.

TcSAD1 is functionally equivalent to AtSSI2

To further investigate *in vivo* TcSAD1 and TcSAD7 protein functions, *TcSAD1* and *TcSAD7* were overexpressed in Arabidopsis *ssi2* mutant background, driven by *Cauliflower mosaic virus* 35S promoter. Owing to the poor transformation rate and mortality rate of the *ssi2* mutant (Schluter et al., 2011), only six transgenic lines overexpressing *TcSAD1* were identified; however, no transgenic plant overexpressing *TcSAD7* was obtained from three Arabidopsis transformation attempts. Notably, two out of six *35S::TcSAD1* (ssi2) transgenic lines were morphologically indistinguishable from wild type Col-0 plants (Figure 3-5A), suggesting that the function of TcSAD1 successfully rescued the dwarf phenotype of *ssi2* mutant. RT-PCR analysis confirmed that transcripts of *TcSAD1* were only detected in the transgenic lines, but not in wild type or *ssi2* mutant plants (Figure 3-5B). Fatty acid composition analysis on wild type Col-0, *ssi2* mutants and the *35S::TcSAD1* (*ssi2*) transgenic lines revealed that overexpression of *TcSAD1* restored the wild type-like level of 18:0 (Figure 3-5C). Another dramatic phenotype of the Arabidopsis *ssi2* is the constitutive

expression of *PR* genes and spontaneous cell death (Shah et al., 2001). In this respect, *35S::TcSAD1* (*ssi2*) plants exhibited no visible cell death on the leaves and contained only basal level of *PR-1* expression (Figure 3-5B). Taken together, we concluded that the overexpression of *TcSAD1* was sufficient to functional complement the reduced activity of SSI2 in the *ssi2* Arabidopsis mutant, demonstrating that *TcSAD1* is a functional ortholog of *AtSSI2*. However, the function of TcSAD7 could not be evaluated because of the lack of the transgenic plant.

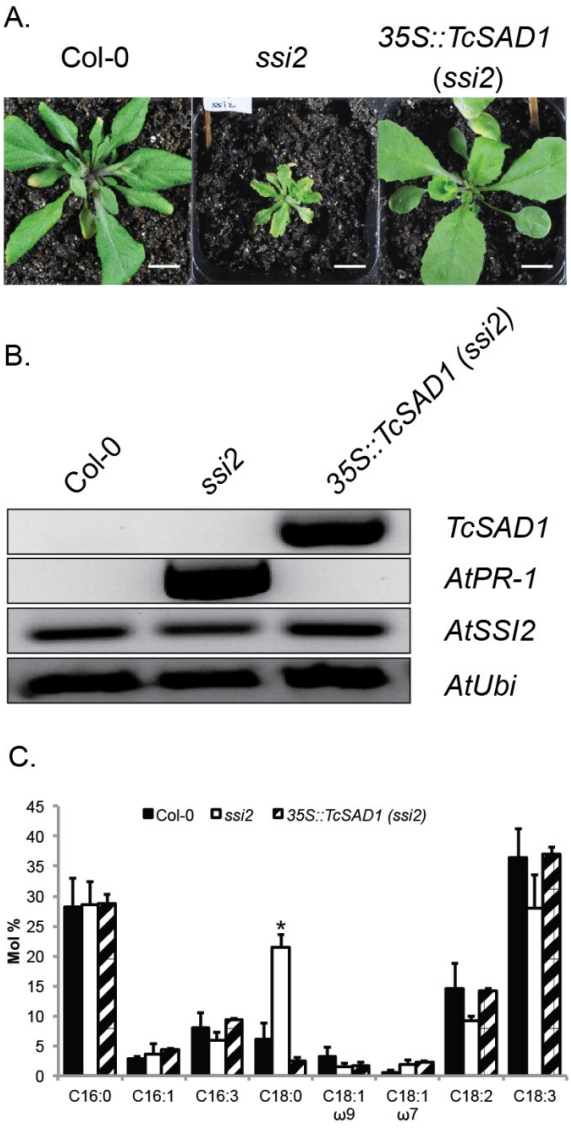


Figure 3-5 Transgenic complementation of the Arabidopsis *ssi2* mutation with the cacao *TcSAD1* gene. Comparison of morphological, molecular, and fatty acid composition

phenotypes of *ssi2* and *35S::TcSAD1* (*ssi2*) plants. (**A**) Morphological comparison of wild-type Col-0, *ssi2*, and *35S::TcSAD1* (*ssi2*) at 4-week-old plants. Overexpression of *TcSAD1* in Arabidopsis *ssi2* successfully rescued dwarf phenotype and eliminated hypersensitive responses. Scale bar = 1cm. (**B**) Overexpression of *TcSAD1* eliminated constitutive overexpression of *AtPR1* (At2g14610) in *ssi2*, but did not affect the level of *AtSSI2*. Arabidopsis *Ubiquitin* was used as the reference gene in RT-PCR analysis. (**C**) Fatty acid compositions of wild-type Col-0, *ssi2*, and *35S::TcSAD1* (*ssi2*). (n= 6 for Col-0 and *ssi2*, n=2 for *35S::TcSAD1*, mean ± SD).

3.5 Conclusion

We identified eight putative SAD isoforms in the cacao genome with high amino acid sequence similarities and functional domains conservation with SAD genes from other species. In the present study, we concluded that the expression level of *TcSAD1* is highly associated with the alteration of fatty acid composition in cacao developing embryos, which can be utilized as a biomarker for screening and breeding for novel genotypes with desirable fatty acid compositions. *TcSAD1* is universally expressed across all the tissues. The activity of TcSAD1 successfully rescued all AtSSI2 related phenotypes in Arabidopsis *ssi2* mutant background, further supporting the functional orthology between these two genes. Diverse tissue expression patterns of these eight SAD isoforms were observed as well. Notably, *TcSAD3* and *TcSAD4* appear to be exclusively and highly expressed in flowers, however, functions of which remain unknown.

3.6 References

1. Argout X, Salse J, Aury JM, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN *et al*: **The genome of Theobroma cacao**. *Nat Genet* 2011, **43**(2):101-108.
2. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool**. *J Mol Biol* 1990, **215**(3):403-410.
3. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput**. *Nucleic Acids Res* 2004, **32**(5):1792-1797.
4. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0**. *Mol Biol Evol* 2007, **24**(8):1596-1599.
5. Gasteiger E, Hoogland C, Gattiker A, Duvaud Se, Wilkins M, Appel R, Bairoch A: **Protein Identification and Analysis Tools on the ExPASy Server**. In: *The Proteomics Protocols Handbook.* Edited by Walker J: Humana Press; 2005: 571-607.
6. Emanuelsson O, Nielsen H, von Heijne G: **ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites**. *Protein Sci* 1999, **8**(5):978-984.
7. Emanuelsson O, Nielsen H, Brunak S, von Heijne G: **Predicting subcellular localization of proteins based on their N-terminal amino acid sequence**. *J Mol Biol* 2000, **300**(4):1005-1016.
8. Argout X, Salse J, Aury JM, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN *et al*: **The genome of Theobroma cacao**. *Nature Genetics* 2011, **43**(2):101-108.
9. Schluter PM, Xu S, Gagliardini V, Whittle E, Shanklin J, Grossniklaus U, Schiestl FP: **Stearoyl-acyl carrier protein desaturases are associated with floral isolation in sexually deceptive orchids**. *Proc Natl Acad Sci U S A* 2011, **108**(14):5696-5701.
10. Mitsuhara I, Ugaki M, Hirochika H, Ohshima M, Murakami T, Gotoh Y, Katayose Y, Nakamura S, Honkura R, Nishimiya S *et al*: **Efficient promoter cassettes for**

enhanced expression of foreign genes in dicotyledonous and monocotyledonous plants. *Plant Cell Physiol* 1996, **37**(1):49-59.

11. Hajdukiewicz P, Svab Z, Maliga P: **The small, versatile pPZP family of Agrobacterium binary vectors for plant transformation**. *Plant Molecular Biology* 1994, **25**(6):989-994.
12. Clough SJ, Bent AF: **Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana**. *Plant J* 1998, **16**(6):735-743.
13. Harrison SJ, Mott EK, Parsley K, Aspinall S, Gray JC, Cottage A: **A rapid and robust method of identifying transformed Arabidopsis thaliana seedlings following floral dip transformation**. *Plant Methods* 2006, **2**:19.
14. Kachroo A, Shanklin J, Whittle E, Lapchyk L, Hildebrand D, Kachroo P: **The Arabidopsis stearoyl-acyl carrier protein-desaturase family and the contribution of leaf isoforms to oleic acid synthesis**. *Plant Mol Biol* 2007, **63**(2):257-271.
15. Cahoon EB, Lindqvist Y, Schneider G, Shanklin J: **Redesign of soluble fatty acid desaturases from plants for altered substrate specificity and double bond position**. *Proc Natl Acad Sci U S A* 1997, **94**(10):4872-4877.
16. Lindqvist Y, Huang W, Schneider G, Shanklin J: **Crystal structure of delta9 stearoyl-acyl carrier protein desaturase from castor seed and its relationship to other di-iron proteins**. *EMBO J* 1996, **15**(16):4081-4092.
17. Whittle EJ, Tremblay AE, Buist PH, Shanklin J: **Revealing the catalytic potential of an acyl-ACP desaturase: tandem selective oxidation of saturated fatty acids**. *Proc Natl Acad Sci U S A* 2008, **105**(38):14738-14743.
18. Schneider G, Lindqvist Y, Shanklin J, Somerville C: **Preliminary crystallographic data for stearoyl-acyl carrier protein desaturase from castor seed**. *J Mol Biol* 1992, **225**(2):561-564.
19. Fox BG, Shanklin J, Somerville C, Munck E: **Stearoyl-acyl carrier protein delta 9 desaturase from Ricinus communis is a diiron-oxo protein**. *Proc Natl Acad Sci U S A* 1993, **90**(6):2486-2490.

20. Mejia LC, Guiltinan MJ, Shi Z, Landherr L, Maximova SN: **Expression of Designed Antimicrobial Peptides in Theobroma cacao L. Trees Reduces Leaf Necrosis Caused by Phytophthora spp.** *Acs Sym Ser* 2012, **1095**:379-395.
21. Zhang J, Li J, Garcia-Ruiz H, Bates PD, Mirkov TE, Wang X: **A stearoyl-acyl carrier protein desaturase, NbSACPD-C, is critical for ovule development in Nicotiana benthamiana**. *The Plant Journal* 2014:n/a-n/a.
22. Patel VK, Shanklin J, Furtek DB: **Changes in fatty-acid composition and stearoyl-acyl carrier protein desaturase expression in developing Theobroma-cacao L embryos**. *Planta* 1994, **193**(1):83-88.
23. Shah J, Kachroo P, Nandi A, Klessig DF: **A recessive mutation in the Arabidopsis SSI2 gene confers SA- and NPR1-independent expression of PR genes and resistance against bacterial and oomycete pathogens**. *Plant Journal* 2001, **25**(5):563-574.

CHAPTER 4: APPLYING GLYCEROL AS A FOLIAR SPRAY ACTIVATES THE DEFENSE RESPONSE AND ENHANCES DISEASE RESISTANCE OF *THEOBROMA CACAO*

Yufan Zhang[a, b], Philip Smith[c], Siela Maximova[a, b], and Mark Guiltinan[a,b,1]

[a]Huck Institute of Life Sciences, The Pennsylvania State University, University Park, PA 16802

[b]The Department of Plant Science, The Pennsylvania State University, University Park, PA 16802

[c]Metabolomics Core Facility, The Pennsylvania State University, University Park, PA 16802

[1]Corresponding author

4.1 Summary

Previous work has implicated glycerol-3 phosphate (G3P) as a mobile inducer of systemic immunity in plants. We tested the hypothesis that exogenous application of glycerol as a foliar spray might enhance disease resistance of *Theobroma cacao* through modulation of endogenous G3P levels. We found that exogenous application of glycerol to cacao leaves over a period of 4 days elevated the endogenous level of G3P and lowered the level of 18:1 (oleic acid). Reactive oxygen species (ROS) production (a marker of defense activation) and expression of many pathogenesis related genes were also induced. Notably, the effects of glycerol application on G3P, 18:1 fatty acid content and on gene expression levels in cacao leaves were dosage dependent. A 100 mM glycerol spray application was sufficient to stimulate defense response without causing any observable damage and resulting in significant lowered lesion formation by the cacao pathogen *Phytophthora capsici*, but a 500 mM glycerol treatment led to chlorosis and cell death. The effects of glycerol treatment on the level of 18:1 and ROS were constrained to the locally treated leaves without affecting

distal tissues. The mechanism of glycerol mediated defense response in cacao and the potential its use as part of sustainable farming systems are discussed.

4.2 Introduction

The current knowledge of the strategies of cacao disease control and the roles of fatty acid derived signaling pathway in defense responses were earlier reviewed in **Section 1.4**. In the present study, taking advantage of a rapid leaf disc pathogen screening bioassay developed in our lab, we explored the potential of applying glycerol as a foliar spray to enhance disease resistance of cacao against the oomycete pathogen *Phytophthora capsici* (Fulton, 1989; Evans, 2007). Our results demonstrate that foliar glycerol application resulted in induction of the plant defense response in cacao and suggests that it has potential as an environmentally safe means to fight important plant diseases in the field.

4.3 Results

Glycerol application increased endogenous G3P levels and decreased oleic acid levels

According to the proposed glycerol metabolism pathway in plants (Eastmond, 2004; Chanda et al., 2008), exogenously applied glycerol is absorbed by the plant cell and further phosphorylated to G3P by glycerol kinase (EC 2.7.1.30) in the cytosol (Figure 3-1B, adapted from (Chanda et al., 2008)). Given the fact that G3P can be transported into the plastids and acylated with oleic acid (18:1-ACP) on glycerolipid biosynthesis pathway (Chanda et al., 2008), increased endogenous levels of G3P may lead to reduction of oleic acid levels (Figure 3-1B). An increase of G3P and reduction of oleic acid levels are thought to be associated with induction of the defense pathways in various plant species.

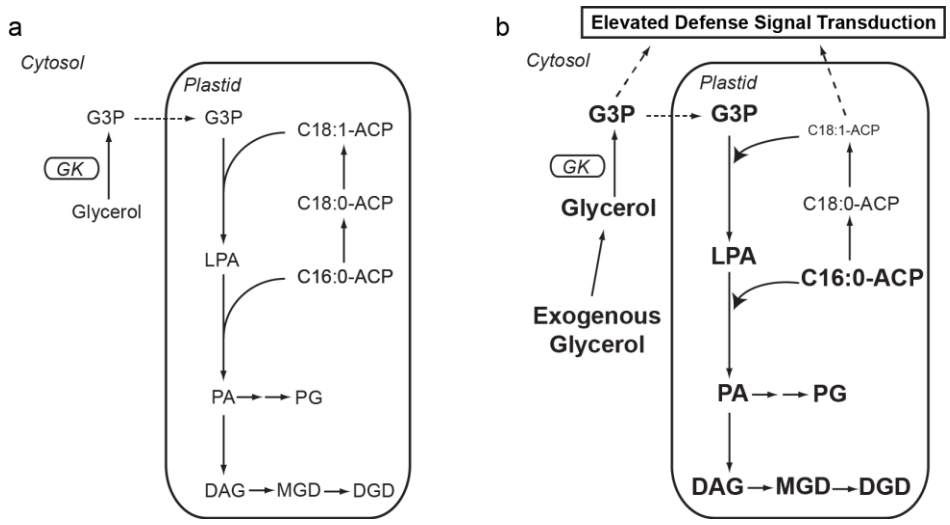


Figure 4-1 Abbreviated scheme of glycerol metabolism and glycerolipid biosynthesis pathway (adapted from (Chanda et al., 2008)) in cacao leaves and proposed effects of application of exogenous glycerol on metabolic pools. (a) Glycerol is phosphorylated to G3P by glycerol kinase (GK) in the cytosol. G3P can translocate into plastids, where G3P is acylated with C18:1-ACP (stearoyl-acyl carrier protein) to yield LPA. LPA can be further acylated with C16:0-ACP (palmitoyl-acyl carrier protein) to yield PA. (b) Upon application of exogenous glycerol, metabolic flow on plastidial glycerolipid pathway is increased through G3P, LPA, PA and DAG pools (indicated as larger font sizes in bold), resulting in a reduction in C18:1-ACP pools (indicated as reduced font sizes) and an elevation of C16:0-ACP pools (indicated as larger font sizes in bold). Abbreviations: GK, glycerol kinase; G3P, glycerol-3-phosphate; LPA, lysophosphatidic acid; PA, phosphatidic acid; PG, phosphatidylglycerol; DAG, diacylglycerol; MDG, monogalactosyldiacylglycerol; DGD, digalactosyldiacylglycerol.

To explore the possibility of using glycerol as a foliar spray for protection of cacao against pathogens, glycerol solutions at different concentrations were prepared in water and applied to SCA6 cacao, stage C leaves on three to four individual plants growing in a greenhouse (Figure 2A). Water without any glycerol was also applied as a negative control on separate trees (panel ii). The selected individual leaves were repeatedly treated with water or glycerol once every 24h for 3 consecutive days and effects were evaluated 24h

after the last treatment. In water-treated leaves, no visible phenotypic changes were observed, except for the normal developmental leaf expansion and hardening identical to that seen with untreated leaves (Figure 4-2A, i and ii). Leaves treated with glycerol at concentrations of 10 mM up to 300 mM also exhibited no significant phenotypic difference compared to untreated or water-treated leaves (Figure 4-2A, iii-x). However, the application of glycerol higher than 300 mM caused chlorosis and yellowing of leaf sections between subveins that were first observable three days after treatment (Figure 4-2A, xi-xiv). The chlorosis increased in intensity with higher concentrations of glycerol and over time. Moreover, cell death was observed on the curled tips of these leaves two days after high concentration glycerol treatment (Figure 4-2A, xi-xiv).

The effects of exogenous application of glycerol on the endogenous levels of G3P within the treated leaves were evaluated by measuring G3P levels by UHPLC-MS (Figure 4-2B). Interestingly, untreated SCA6 cacao leaves contained significant higher level of G3P compared to water or 100 mM glycerol treated leaves ($p < 0.05$), and no significant difference of the levels of G3P was observed between water-treated and 100 mM glycerol treated leaves. Notably, the 500 mM glycerol application led to a massive increase of G3P of approximately 100-fold in comparison to the water control ($p < 0.001$) and 100 mM glycerol-treated leaves ($p < 0.001$). The results suggested that glycerol application at a low concentration (100 mM) did not result in the accumulation of G3P. However, the accumulation of G3P after treatment with higher concentrations of glycerol (500 mM) implied that the normal functions of plastids associated with lipid metabolism might be disrupted, which could be related with the phenotypic changes presented above (Figure 4-2A).

To further explore effects of exogenously applied glycerol on the metabolic pools that feed into the plastidial glycerolipid biosynthesis pathway, fatty acid profiles were determined by GC/MS (Figure 4-2C). Among all fatty acid species detected in our cacao leaf samples, 16:0, 18:0, 18:1 (n-9), 18:2 (n-6) and 18:3 (n-3) were the major contributors to the fatty acid pools. Water treatment resulted in no significant difference in fatty acid composition compared to that of untreated leaves, except that 16:0 did accumulate a slightly higher molar proportion in water-treated leaves than untreated leaves (p-value <

0.05). Glycerol application had significant impacts on the levels of 16:0, 18:1, and 18:3, starting from the lowest concentration of glycerol tested (10mM). Moreover, progressive reductions of 16:0 and 18:1 levels and increment of 18:3 level were highly correlated with increased concentrations of glycerol applied (*correlation coefficient r* = -0.84, -0.96, and 0.76, respectively, Figure 4-2C). As indicated by the correlation coefficient, the level of 18:1 had the strongest linear correlation with the concentration of exogenously applied glycerol, which was consistent with the scheme of the glycerolipid biosynthesis pathway (Figure. 1) in which 18:1 serves as the direct substrate for G3P acylation reaction in plastids.

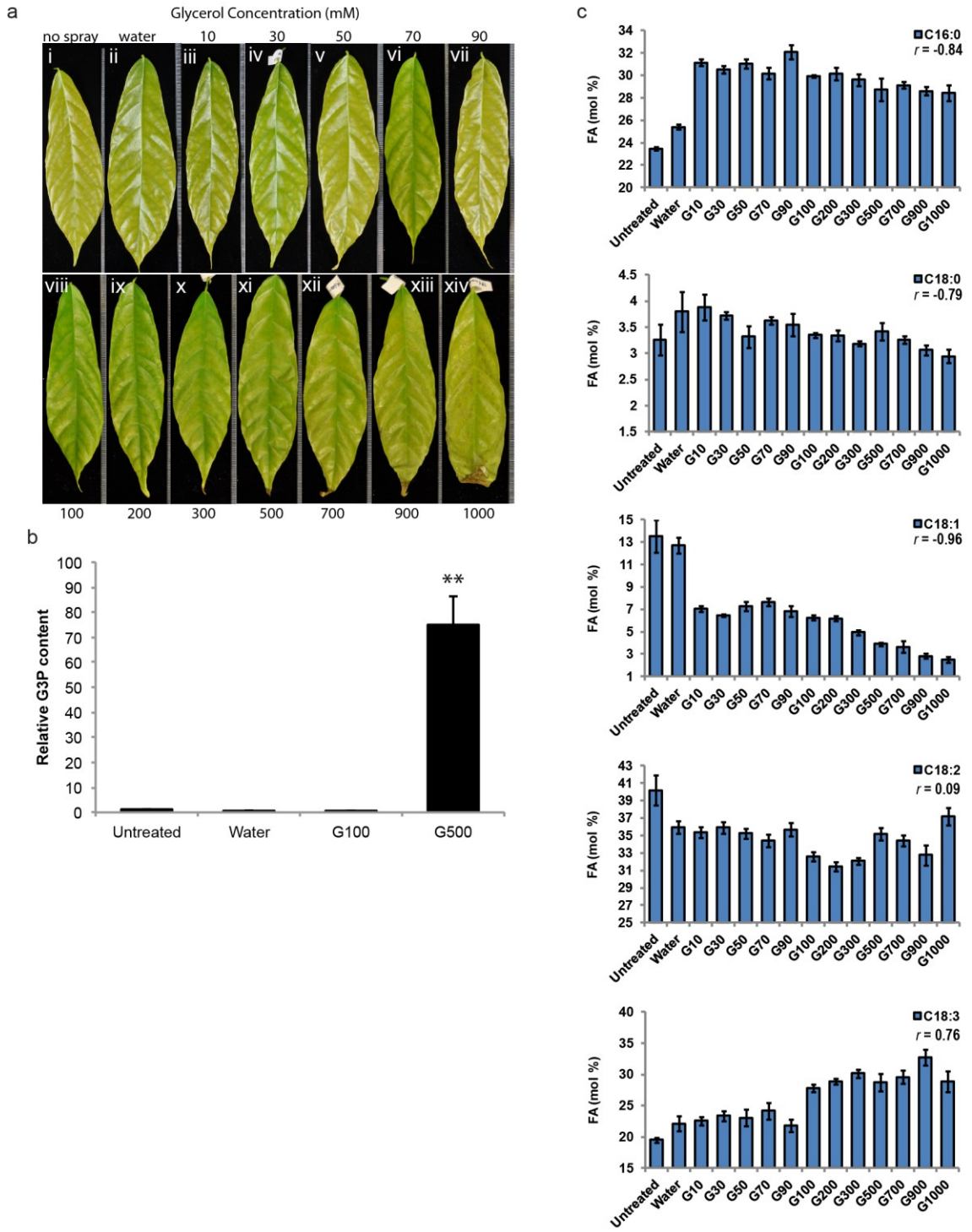


Figure 4-2 Effects of glycerol application on leaf morphology, G3P levels and fatty acid levels in cacao leaves. (a) Morphology of cacao leaves after exogenous glycerol treatments with various concentrations (i) untreated, (ii) water treated, (iii – xiv) glycerol treated with concentrations: 10 mM, 30 mM, 50 mM, 70 mM, 90 mM, 100 mM, 200 mM, 300

mM, 500 mM, 700 mM, 900 mM and 1000 mM. (b) Endogenous G3P levels measured four days after water or glycerol treatments (n=7, mean ± SE). (c) Fatty acid levels after indicated treatments. Fatty acid levels are presented as mol% (n=4, mean ± SE). Significance was determined by t-test (** represents p-value < 0.001, *r* represents the correlation between the concentrations of glycerol applied and the mol% of fatty acids).

Glycerol application induced ROS accumulation and elevated MDA content in cacao leaves

It has been well documented that the production and accumulation of ROS in infected plant tissues plays a significant role in early pathogen recognition and defense related gene regulation (Torres et al., 2006). Additionally, ROS can also directly lead to cross-linking of plant cell walls, lipid peroxidation, and membrane disruption (Zurbriggen et al., 2010). In this respect, the correlation between the decreased level of 18:1 induced by glycerol and the generation of ROS has been reported in soybean (Kachroo et al., 2008) and rice (Jiang et al., 2009). Therefore, global levels of ROS in untreated, water and glycerol-treated SCA6 cacao leaves were measured by DCF staining method, where a bright green fluorescence is an indication of the presence of ROS (Maxwell et al., 1999). Both 100 mM and 500 mM glycerol were tested considering that 100 mM was able to lower the 18:1 level without causing severe leaf chlorotic phenotypes (Figure 4-2A). There was no significant difference in the fluorescence detected between the untreated and water-treated SCA6 cacao leaves (Figure 4-3A, B and E). However, both 100 mM and 500 mM glycerol application resulted in high levels of fluorescence, indicating significantly higher ROS accumulation in these treatments compared to either untreated or water-treated leaves (p-value < 0.05) (Figure 4-3C, D and E). No significant difference was observed between 100 mM and 500 mM glycerol treatments. These results indicated that glycerol application induced global ROS accumulation.

One downstream effect of ROS accumulation is elevated levels of lipid peroxidation products such as malondialdehyde (MDA) (Zoeller et al., 2012). Thus, MDA was measured as a proxy of lipid peroxidation in control and glycerol treated leaves (100 mM and 500

mM). As depicted in Figure 4-3F, there was no significant difference in MDA concentration among untreated, water-treated and 100 mM glycerol treated cacao leaves. However, the 500 mM glycerol treatment resulted in a further increase in MDA content, significantly higher than untreated and water treated leaves (p-value < 0.05).

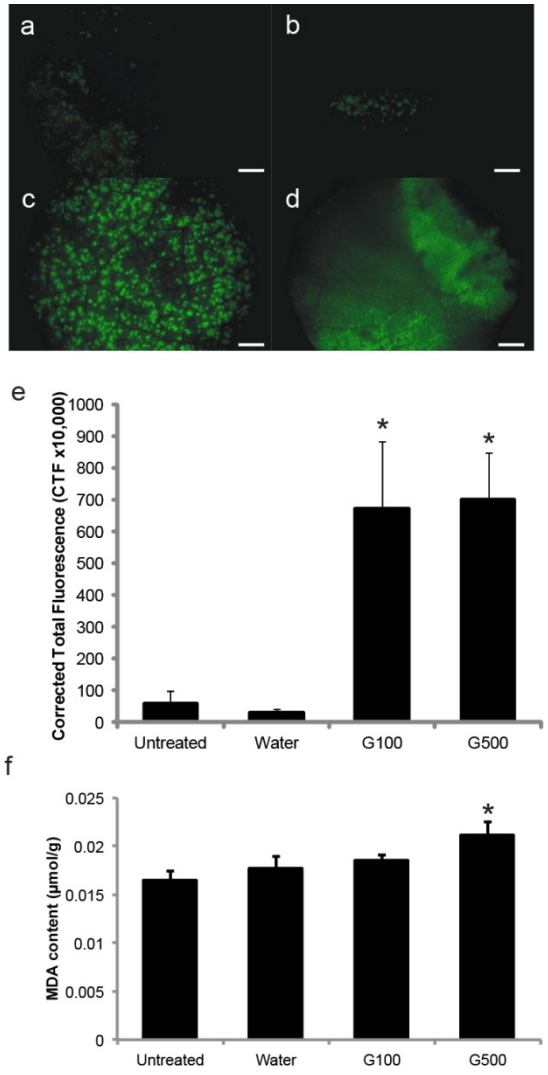


Figure 4-3 Effects of glycerol application on ROS accumulation and MDA content in cacao leaves. Representative microscope images of cacao leaves infiltrated with DCF to stain total ROS. (a) untreated, (b) water treated, (c) 100 mM (G100) and (d) 500 mM (G500) glycerol treated cacao leaves. Scale bars = 0.5 mm. (e) Quantification of ROS accumulation in cacao leaves determined by corrected total fluorescence from DCF using

ImageJ. (f) MDA content in cacao leaves. (n=7, mean ± SE). Significance was determined by t-test (* represents p-value < 0.05).

Glycerol induction of defense gene expression in cacao leaves

Since the generation and accumulation of ROS is highly correlated with defense related gene regulation (Torres et al., 2006; Kachroo et al., 2008), the expression levels of five classes of *PR* genes (Chanda et al.), *PR-1* to *PR-5* were measured in leaves after glycerol applications. To select the orthologous genes in cacao, the full-length amino acid sequences of *PR-1* to *PR-5* from tobacco (Queries summarized in Table 1) were blasted against the predicted cacao proteome of the Belizian Criollo genotype (B97-61/B2) (http://cocoagendb.cirad.fr/ (Argout et al., 2011)) using blastp algorithm with E-value cut-off of $1e^{-5}$ (Altschul et al., 1990). One hundred and five genes were identified (Table S1) and further screened for their inducibility by *Colletotrichum* and *Phytothphora* using available cacao microarray data (Guiltinan lab, unpublished data). As a result, 14 candidate genes were selected and their expression levels were evaluated by RT-qPCR after treatments with water, 100 mM or 500 mM glycerol (Table 1). Two additional chitinase genes (*TcChi1*, *Tc02g003890* (Maximova et al., 2006); *TcChi4*, *Tc04g018110* (unpublished data)) from PR-3 family were also included in this study as overexpression of these genes in cacao conferred resistance against certain pathogens in our previous study. *TcTUB1* (*Tc06g000360*), which was previously identified as the most stable gene during leaf development, was used as the reference gene. The relative expression of *PR* genes was plotted as $\log_{10}$ scale (Figure 4-4).

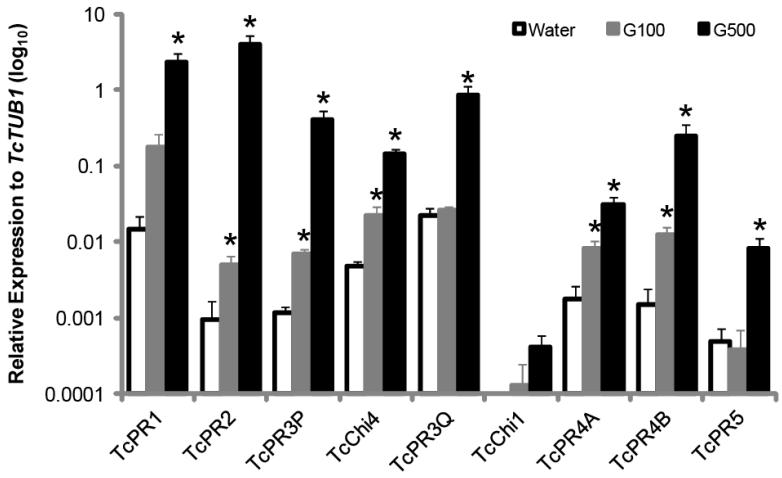


Figure 4-4 Transcript levels of *PR* genes measured by RT-qPCR. The relative expression levels of *PR* genes were quantified and compared after treatment with water, 100 mM and 500 mM glycerol. The expression levels of *PR* genes were normalized to that of *TcTUB1* and plotted as $\log_{10}$ scale (n=3, mean ± SE). Significance was determined by t-test (* represents p-value < 0.05).

Table 4-1 Summary of the changes in PR gene expression levels in response to 100 mM and 500 mM glycerol application measured by RT-qPCR. Expression levels were normalized to TcTUB1. Fold changes were calculated as a ratio of relative expression levels between glycerol and water treated leaves (n=3).

PR Family	Identifier in cacao genome database	Annotation in cacao genome database	Average Fold change	
			G100/Water	G500/Water
PR-1	Tc02g002410	Pathogenesis-related protein 1	9.87	156.46
PR-2	Tc04g029300	Glucan endo-1,3-beta-glucosidase, basic vacuolar isoform	3.31	4174.80
PR-3P	Tc01g000770	Endochitinase	3.56	348.71
Chi4	Tc04g018110	Endochitinase PR4	4.60	30.05
PR-3Q	Tc04g018160	Endochitinase PR4	0.89	37.51

Chi1	Tc02g003890	Endochitinase 1	nd	nd
PR-4A	Tc05g027220	Pathogenesis-related protein P2	4.70	17.65
PR-4B	Tc05g027210	Pathogenesis-related protein PR-4B	3.44	164.63
PR-5	Tc03g026990	Thaumatin-like protein	0.97	16.42

Every gene evaluated showed a high levels of induction in the 500 mM treated leaves, up to a 4,174-fold increase. At 100 mM treatment, all but two of the genes showed a more modest increase in expression, up to 10-fold. In general, the results clearly indicated that glycerol application significantly up-regulated the expression of a set of *PR* genes (Table 1). At 100 mM glycerol treatment, two genes, *PR-3Q* and *PR-5*, were not significantly induced (Figure 4-4). The expression of *TcChi1* was significantly induced under both 100 mM and 500 mM glycerol applications even though its expression was not detectable under the specific RT-qPCR conditions in water-treated leaves. Moreover, expression of *TcChi4* was up-regulated by about 5- and 30- fold by 100 mM and 500 mM glycerol applications, respectively, compared to water-treated leaves (Table 1).

Glycerol application enhanced disease resistance of cacao leaves against *Phytothphora capsici*

Previously it was reported that overexpression of *TcChi1* enhanced resistance of cacao leaves against *Colletotrichum gloeosporioides* (Maximova et al., 2006) and that PR-1, PR-2, PR3, PR-4, and PR-5 have direct antifungal and/or anti-oomycete activities in many plant species, such as tobacco and wheat (Yun, 1996; van Loon et al., 2006; Campos et al., 2007). Thus, impacts of glycerol application on disease resistance of cacao leaves against pathogens were investigated using an *in vivo* pathogenicity assay. Stage C leaves of SCA6 were treated with glycerol or water for 3 consecutive days, leaf discs were excised 24h after the last treatment and infected with cacao pathogen *Phytophthora capsici*. Untreated leaf discs were also infected with *P. capsici* as a control. Leaves were photographed (Figure 2-5 A-E) and lesion sizes were measured three days after infection. There were no

significant differences in average sizes of the lesions between untreated and water-treated cacao leaves (Figure 2-5F). Notably, all three glycerol treatments drastically decreased the average lesion size compared to untreated or water-treated leaves (p-value < 0.05). There were no significant differences in the average sizes of lesions among the 100 mM, 500 mM and 1000 mM glycerol treatments (Figure 2-5F), indicating that 100 mM glycerol was sufficient to induce disease resistance against *P. capsici*, and that increased glycerol concentration did not contribute to further elevated resistance. Glycerol treatments at concentration lower than 100 mM were also evaluated (50, 70 and 90 mM), which did not result in consistent enhanced resistance response (data not shown).

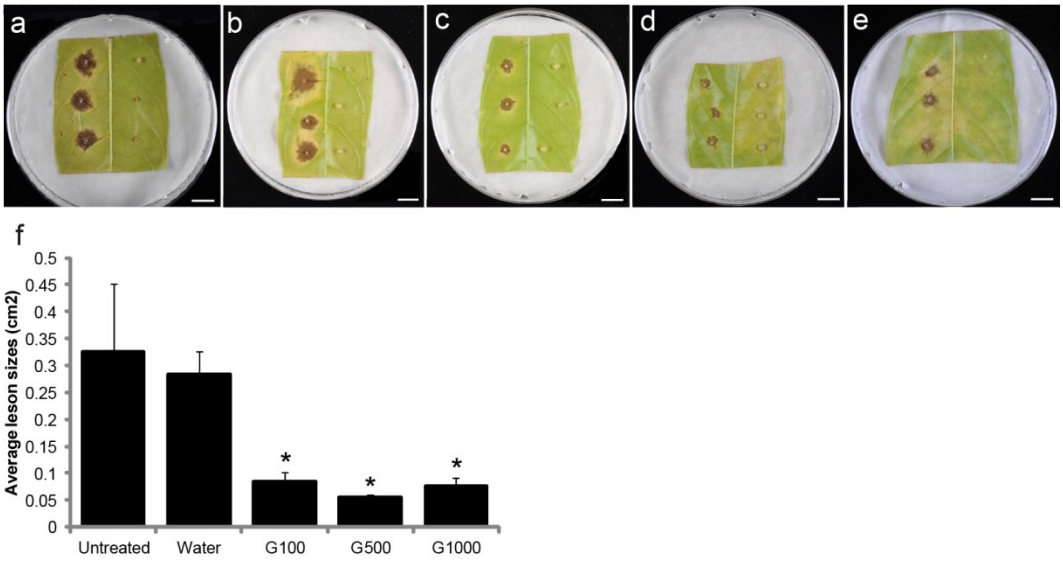


Figure 4-5 Glycerol enhanced disease resistance of cacao leaves against *P. capsici*. (a-e) Representative images of leaf lesions three days after inoculation with *P. capsici*. (a) Untreated leaves; (b) Leaves were treated with water; (c) 100 mM glycerol; (d) 500 mM glycerol; (e) 1000 mM glycerol. Glycerol treatments at the designated concentration were applied once per day every 24 h for 3 consecutive days prior to *P. capsici* infection. (f) Average lesion sizes were by measured by ImageJ three days after inoculation with *P. capsici* (n=7, mean ± SE). Significance was determined by t-test (* represents p-value < 0.05). Scale bars = 1 cm.

Local glycerol treatments did not affect G3P, ROS, and 18:1 levels in distal leaves

We tested if glycerol treatment would act systemically to alter the levels of G3P, 18:1 and ROS levels in distal non-treated leaves. Water or 500 mM glycerol were applied to five leaves of the same branch on two separate cacao plants. Treated leaves and leaves from distal branches were collected and G3P, 18:1 and ROS levels were examined. Consistent with previous results, glycerol application significantly increased G3P and ROS levels and decreased 18:1 levels in the treated leaves compared to water-treated leaves (p-value < 0.05) (Figure 4-6A-C). However, no significant differences were recorded for the G3P, ROS and 18:1 levels in non-treated distal leaves on glycerol-treated plants compared to the water controls (Figure 4-6A-C), suggesting that under the conditions we tested, local glycerol treatment did not produce a mobile signal that would induce a similar response in distal leaves, which was consistent with the result indicating that exogenous application of G3P alone was not sufficient to induce the translocation of G3P to the distal tissues (Chanda et al., 2011).

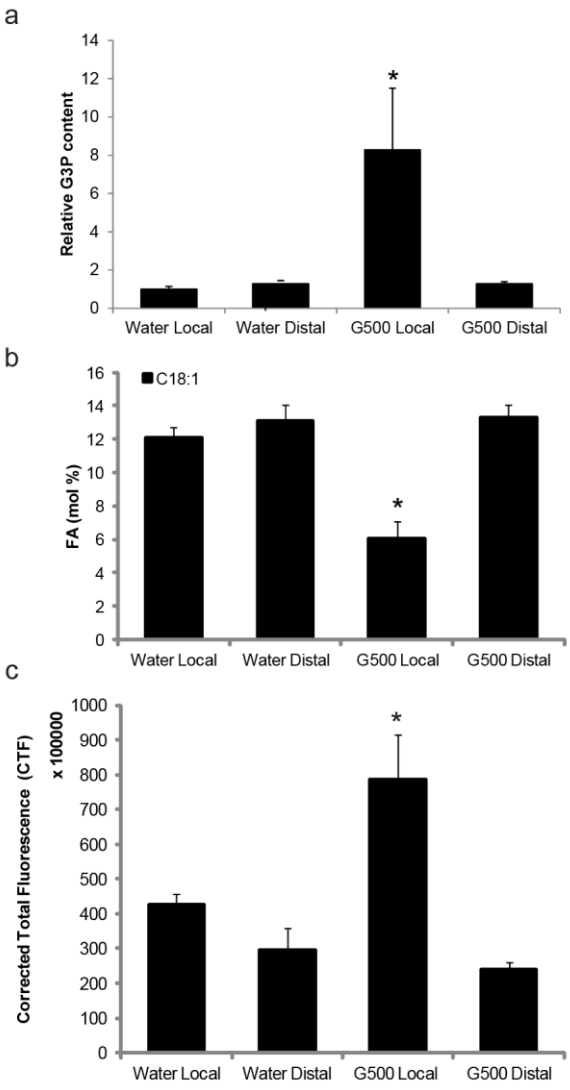


Figure 4-6 G3P, 18:1 and ROS levels in distal leaves in response to local glycerol treatment. (a) G3P levels in local and distal water or glycerol treated cacao leaves. (n=4, mean ± SE). (b) 18:1 levels in local and distal water or glycerol treated cacao leaves. (n=5, mean ± SE). (c) ROS levels in local and distal water or glycerol treated cacao leaves. (n=3, mean ± SE). Significance was determined by t-test (* represents p-value < 0.05).

4.4 Discussion

The mechanism of glycerol mediated defense response in cacao

The present study provides proof of concept of applying glycerol to enhance disease resistance in cacao. We observed dose responses to glycerol treatment in the cacao leaf

physiological and molecular states, which has not been previously reported. A strong correlation (r = -0.96) between the concentrations of the exogenously applied glycerol and the oleic acid content of treated leaves indicated that the effects of glycerol application on disease resistance are likely to be transduced through a fatty acid (18:1, oleic acid) mediated signaling pathway. In this respect, it is well documented that a lowered level of 18:1 in the Arabidopsis *ssi2* mutant induces ROS production, *PR* gene expression, hypersensitive response and spontaneous cell death, and that exogenous application of glycerol can mimic these phenotypes (Kachroo et al., 2003; Kachroo et al., 2004). Thus, it is likely, and consistent with our data, that glycerol application induces defense responses in cacao through a similar mechanism. The fact that glycerol application significantly up-regulates a large set of *PR* genes in cacao leaves provides further insights into the mechanism of the induced defense response. For example, it has been previously demonstrated by our group that overexpression of *TcChi1* enhances resistance of cacao to *Colletotrichum gloeosporioides* (Maximova et al., 2006) and *Phytophthora capsici* (unpublished data), which is consistent with our observation that induced elevated level of *TcChi1* by glycerol does enhance resistance to *Phytophthora capsici*. Although detailed functional characterization of other PR genes is still needed, it is likely that glycerol application is able to confer broad-spectrum resistance since the direct activities of these PR genes against various fungi and oomycete pathogens have been well illustrated in cacao and other plant species (Campos et al., 2007; Sels et al., 2008). Moreover, in general, induction of *PR* genes by glycerol application was observed to be dose-dependent, which could result from rising levels of transduction signals involved in an 18:1 pathway, such as nitric oxide (Mandal et al., 2012) and/or salicylic acid (SA) (Shah et al., 2001; Kachroo et al., 2003). Together, our data suggest that glycerol application might enhance disease resistance of cacao leaves via an 18:1 mediated defense-signaling pathway.

As an initial acylation substrate for glycerolipid biosynthesis in plastids, the level of 18:1 is tightly modulated by the level of glycerol derivative G3P (Chanda et al., 2008). Compelling evidence reveals the significant roles of G3P in contributing to the basal resistance of Arabidopsis against *Colletotrichum higginsianum* (Chanda et al., 2008) and as a critical immune signal generated following the local pathogen infection to confer SAR in

the distal tissues (Chanda et al., 2011). However, it is noteworthy that 100 mM glycerol application conferred sufficient resistance in cacao leaves against *Phytophthora capsici* without significant alteration of the endogenous G3P level, suggesting that G3P does not seem to be involved in the defense response at this level (Figure 4-2B), and it is more likely that G3P serves as the intermediate for the acylation of 18:1 instead of the accumulation in plastids. However, our data cannot rule out the possibility that elevated G3P levels might contribute to enhanced resistance of cacao to pathogens since 500 mM glycerol application did lead to massive accumulation of G3P and result in enhanced resistance (Figure 2-5A and B) as well. Alternatively, another explanation for the accumulation of G3P could be that high concentration glycerol (500 mM) causes phytotoxic damages to cacao leaves (Figure 4-2B) and disrupts normal functions of plastids, which affects the glycerolipid biosynthesis pathway and thus causes the accumulation of G3P as an intermediate. This hypothesis is further supported by the fact that 500 mM glycerol application increases ROS production (singlet oxygen $^1O_2$ and free radicals, (Jambunathan, 2010; Zoeller et al., 2012)) and elevates the level of lipid peroxidation measured by MDA content (Figure 4-3F), which could potentially destroy the integrity of membranes (Folden et al., 2003) and lead to cell death and prominent symptoms of hypersensitive response that have never been described in cacao (Figure 4-2A).

It was observed that the local glycerol treatment of cacao leaves did not lead to the changes in G3P, fatty acid profiles and ROS production levels in distal tissues. Consistent with this observation, in Arabidopsis the exogenous application of G3P alone did not result in the translocation of G3P into distal tissue (Chanda et al., 2011). Interestingly, DIR1, a lipid transfer protein, was required to assist the translocation of G3P to confer SAR in distal tissue. Thus it seems plausible that the glycerol applications we tested in cacao did not induce DIR1 accumulation to aid the movement of G3P. However, the elevated level of G3P in locally treated tissue could still be beneficial to induce enhanced SAR when plants are exposed to pathogens and the expression of *DIR1* is induced (Chanda et al., 2011). Meanwhile, it was recently discovered that NO/ROS → Azelaic acid (AzA) → G3P-derived signaling pathway operates in parallel with SA pathways to induce SAR, and that ROS triggers SAR in a concentration dependent manner (Yu et al., 2013; Wang et al., 2014).

Therefore, it is equally possible that glycerol-induced accumulation of ROS might negatively regulate SAR, which could explain why local increment of G3P did not induce similar effects in distal leaf tissues. Together, it remains a formal possibility that either or both pathways are involved in the mechanisms of glycerol induced defense response in cacao.

The potential of applying glycerol as a foliar spray in cacao plantations

As mentioned earlier, cacao is very susceptible to many severe diseases that cause significant yield loss in cacao plantations (Evans, 2007). A low cost and effective strategy for disease control is thus greatly needed. Remarkably, in this aspect, glycerol is an environmentally friendly chemical excessively generated as a low cost byproduct from biodiesel industry. Our results provide strong evidence suggesting that glycerol has a great potential to be applied as a substitute or additive into current agricultural chemicals, which also satisfies the urgent need of exploration for a value added utilization of glycerol (Yang et al., 2012). Further research is needed to bring this potential technology to commercialization. For example, the dosage, timing and intervals of the application need to be characterized under natural tropical conditions. In addition, given that the effects of glycerol application in our studies were constrained to locally treated tissues, a broader spray on a whole plant scale might be considered to achieve a better performance. Further investigations with more cacao pathogens are needed to determine the potential efficacy of glycerol treatment for different diseases and in different cacao growing regions.

4.5 Material and Methods

Glycerol treatment

Fully expanded leaves, developmental stage C (Shi, 2010) of greenhouse grown SCA6 were treated with glycerol (MP, Cat. 800688) solutions at various concentrations prepared in sterile water containing 0.04% Silwett L-77 (Fisher Scientific Cat. NC0138454). Water containing 0.04% Silwett L-77 was used as the control treatment. Leaves were completely

immersed in 100 ml solutions for 10 s every 24h for 3 consecutive days and collected 24h after the last treatment for pathogen bioassay and fatty acid analysis.

G3P extraction and quantification by UHPLC-MS

To determine G3P levels, 1 g of cacao leaf tissue was frozen in liquid nitrogen and ground into fine powder. About 30 mg of ground tissues were weighed and extracted with 5 ml of 80% (v/v) ethanol, and centrifuged at 5520 g for 5 min at 4 °C. The supernatants were collected and 10 ul aliquots analyzed by UHPLC-MS (Thermo Ultimate 3000 pump and Thermo Exactive Plus Orbitrap). The method and instrumentation was based on an ion pairing reversed phase negative ion Orbitrap method (Lu 2010), a smaller particle size C18 column was used (Titan 100 x 2.1mm 1.9 um particle size, Sigma) and the mass spectrometer was scanned from 85-1000 m/z at a resolution of 140,000 throughout the entire run. Under these conditions, glycerol-3-phosphate bis salt (Sigma, Cat. G7886) eluted as a single peak at m/z 171.0063 with a retention time of 7.27 min. The integrated areas of the reconstructed ion chromatograms for this ion were used to measure the amount of glycerol-3-phosphate in the extracts.

Fatty acid profiling by GC-MS

Plant tissue from glycerol treated and control leaves (four biological replicates) were ground in liquid nitrogen and fatty acid methyl esters (FAME) were prepared from each sample using approximately 30 mg of tissue per sample. Briefly, 1 ml of a MeOH/fuming HCl/Dichloromethane (10:1:1 v/v) solution was added to each tissue sample and incubated without shaking at 80 °C for 2h. Fatty acid methyl esters were re-extracted in 1 ml buffer H_2O/Hexane/Dichloromethane (5/4/1, v/v) with vortexing for 1min. The hexane (upper phase) was separated by centrifugation at 1500 g for 5 min, transferred to glass GC vials (Agilent) and evaporated to dryness under vacuum. The FAMEs were then dissolved in 500 µl hexane for GC-MS analysis. Pentadecanoic acid (C15:0) (Sigma, Cat. P6125) was used as the internal standard added prior to the extraction and methyl nonadecanoate (C19:0-methyl ester) (Sigma, Cat. N5377) was used as the spike control, added into the sample prior to the GC injection. Samples were analyzed on an Agilent 6890N gas chromatograph coupled to a Waters GCT time of flight mass spectrometer. Mass spectra were acquired in

electron ionization mode (70 eV) from 45 to 500 Da at a rate of one scan/second. The samples were separated on an Omegawax® 250 Capillary GC column (30 m x 0.25mm 0.25 uM phase thickness, Sigma, Cat. 24136) using helium at a constant flow of 1.0 ml/min. The initial oven temperature was 100 °C held for 1 min then increased at 15 °C/min to a temperature of 150 °C, and then increased at 4 °C/min to a final temperature of 280 °C. Samples (1ul) were injected onto the column using a split/splitless injector maintained at 240 °C with a split ratio of 50/1.

DCF staining of ROS production

Fully expanded, stage C leaves of genotype SCA6 were detached from cacao seedling after treatments and vacuum infiltrated for 2 min with 20 μM 2,7′-dichlorofluorescin diacetate (HDCF-DA; Sigma, 20 mM stock in DMSO). To avoid saturation of the fluorescent signal, samples were photographed exactly 2 min after infiltration. Images were obtained at 7.5x magnification using a Nikon SMZ-4 stereo microscope equipped with an epi-fluorescence attachment, a 100 W mercury light source and a 3 CCD video camera system (Optronics Engineering, Goleta, CA). The fluorescence imaging filter used was a 450-490 nm excitation and emission a 520-560 nm emission. Images were analyzed by ImageJ Software and corrected total fluorescence (CTF) of DCF was calculated by the formula: CTF = Integrated Density – (Area of selected cell X Mean fluorescence of background readings).

Lipid peroxidation assay

Lipid peroxidation in untreated and glycerol treated cacao leaves were measured by determining the malondialdehyde (MDA) content using thiobarbituric acid reactive substances (TBARS) assay. After treatments, cacao leaves were rapidly frozen in liquid nitrogen and ground into fine powder (30 mg per sample). 5 ml of 0.1% trichloroacetic acid (TCA) was added to each sample and were centrifuged at 5520 g (max speed) for 5 min. 1 ml of the supernatant was collected per sample and mixed with 2 ml of TBARS reagent consisting of 15% TCA and 0.375% 2-thiobarbituric acid (TBA) in 0.25M HCl. The mixture was boiled for 15 min, cooled on ice for 10 min and centrifuged at 3320 g for 5 min. 200 μl of supernatant were collected and absorbance was read at 532 nm. The MDA standard curve was prepared using 1,1,3,3-tetraethoxypropane.

RNA extraction and RT-qPCR analysis of gene expression

Plant tissues were first ground in liquid nitrogen. Total RNA was extracted using Plant RNA Purification Reagent (Life Technologies, Cat. 12322-012, following manufactures protocol). The concentration of RNA was measured using a Nanodrop 2000c (Thermo Scientific). RNA was further treated with RQ1 RNase-free DNase (Promega, Cat. M6101) to remove potential genomic DNA contamination (following the manufacturer's protocol). 500 ng of treated RNA was reverse-transcribed by M-MuLV Reverse Transcriptase (New England Biolabs) with oligo-(Dann et al.)$_{15}$ primers to obtain cDNA. RT-qPCR was performed in total reaction volume of 10 µl containing 4 µl diluted-cDNA (1:50), 5 µl SYBR Green PCR Master Mix (Takara), 0.2 µl Rox, and 0.4 µl each 5 µM primers. Each reaction was performed in duplicates using Roche Applied Biosystem Step One Plus Realtime PCR System under the following program: 15 min at 94˚C, 40 cycle of 15 s at 94 °C, 20 s at 60 °C, and 40 s at 72 °C. The specificity of the primer pair was examined by PCR visualized on a 2% agarose Gel and dissociation curve. A tubulin gene in cacao (*Tc06g000360*, *TcTUB1(Argout et al., 2011)*) *TcTUB1-5'*: GGAGGAGTCTCTATAAGCTTGCAGTTGG and *TcTUB1-3'*: ACATAAGCATAGCCAGCTAGAGCCAG) was used as a reference gene.

Leaf disk pathogen bioassay of cacao leaves

Fully expanded, stage C leaves of genotype SCA6 were collected after treatments and detached leaf pathogen bioassays were performed as previously described in (Mejia et al., 2012). Briefly, leaves were detached, dissected and the middle sections of the leaves were transferred abaxial side up to a new Petri dish on top of 3M filter paper soaked with 10 ml sterile water to maintain humidity. The right side of each leaf, delineated by the midvein, was inoculated with 3 agar plugs (3 mm diameter core bore) of *Phytophthora capsici* mycelium. Sterile agar plugs were placed at the left side of each leaf as negative controls. Inoculated leaves were then incubated at 27 °C and 12:12 (L: D) light cycle for three days before the evaluation of disease symptoms. At least five biological replicates from each treatment were infected and images were taken for the measurements of average lesion sizes by ImageJ software. Three lesions on the same leaf disc were taken as technical

replicates. Average lesion sizes are calculated from biological replicates and significance is determined by t-test.

Authors' contributions

YZ conceived and performed most of the experiments, such as the glycerol treatments, G3P and fatty acid extractions, gene expression analysis, pathogen assays, and drafted the manuscript. PS participated in developing the G3P extraction and quantification protocol, HPLC-MS and GC-MS analysis, and reviewed the manuscript. SNM was involved in designing and directing the experiments, and revising the manuscript. MJG contributed to the conception of the study, gave advice on experiments, and finalized the manuscript.

Acknowledgements

We would like to thank Lena Sheaffer and Brian Rutkowski for the technical assistance in maintenance of our greenhouse. We are also grateful to Andrew Fisher for valuable comments to improve the manuscript. This work was supported in part by The Pennsylvania State University, College of Agricultural Sciences, The Huck Institutes of Life Sciences, the American Research Institute Penn State Endowed Program in the Molecular Biology of Cacao and a grant from the National Science Foundation BREAD program (NSF0965353), which supported the development of the transient gene expression assay for cacao.

4.6 Supplemental Data

Table S1 Putative PR1-PR5 genes identified in cacao genome by blastp algorithm

Query	Expect	% Identity	Locus ID	Function
PR1a_CAA31008.1_Tobacco	3E-53	66.19	Tc02_p002400	Tc02_g002400 Pathogenesis-related protein 1
PR1a_CAA31008.1_Tobacco	3E-53	67.63	Tc02_p002380	Tc02_g002380 Pathogenesis-related protein 1C
PR1a_CAA31008.1_Tobacco	3E-53	68.35	Tc02_p002410	Tc02_g002410 Pathogenesis-related protein 1
PR1a_CAA31008.1_Tobacco	7E-50	64.14	Tc02_p002420	Tc02_g002420 Basic form of pathogenesis-related protein 1
PR1a_CAA31008.1_Tobacco	2E-48	58.7	Tc02_p010380	Tc02_g010380 Hypothetical protein
PR1a_CAA31008.1_Tobacco	2E-46	62.59	Tc02_p002390	Tc02_g002390 Pathogenesis-related protein 1
PR1a_CAA31008.1_Tobacco	9E-41	49.24	Tc01_p034430	Tc01_g034430 Putative Pathogenesis-

				related protein 1C
PR1a_CAA31008.1_Tobacco	6E-40	57.14	Tc02_p002430	Tc02_g002430 Basic form of pathogenesis-related protein 1
PR1a_CAA31008.1_Tobacco	3E-39	47.26	Tc05_p005530	Tc05_g005530 Putative Pathogenesis-related protein PR-1
PR1a_CAA31008.1_Tobacco	3E-38	54.14	Tc09_p016590	Tc09_g016590 Pathogenesis-related protein PR-1
PR1a_CAA31008.1_Tobacco	2E-35	49.64	Tc09_p016580	Tc09_g016580 STS14 protein, putative
PR1a_CAA31008.1_Tobacco	9E-35	45.52	Tc09_p000720	Tc09_g000720 STS14 protein, putative
PR1a_CAA31008.1_Tobacco	1E-26	41.67	Tc01_p003940	Tc01_g003940 STS14 protein, putative
PR1a_CAA31008.1_Tobacco	2E-23	37.68	Tc10_p000980	Tc10_g000980 Hypothetical protein
PR2_AAA34103.1_Tobacco	1E-102	56.13	Tc09_p024130	Tc09_g024130 Lichenase
PR2_AAA34103.1_Tobacco	4E-97	53.21	Tc04_p029300	Tc04_g029300 Glucan endo-1,3-beta-glucosidase, basic vacuolar isoform
PR2_AAA34103.1_Tobacco	1E-96	55.49	Tc09_p024150	Tc09_g024150 Glucan endo-1,3-beta-glucosidase, basic vacuolar isoform
PR2_AAA34103.1_Tobacco	2E-95	55.45	Tc09_p024140	Tc09_g024140 Glucan endo-1,3-beta-glucosidase, basic vacuolar isoform
PR2_AAA34103.1_Tobacco	6E-90	52.92	Tc09_p023540	Tc09_g023540 Lichenase
PR2_AAA34103.1_Tobacco	2E-86	50	Tc00_p083950	Tc00_g083950 Glucan endo-1,3-beta-glucosidase
PR2_AAA34103.1_Tobacco	1E-80	49.21	Tc05_p016070	Tc05_g016070 Glucan endo-1,3-beta-glucosidase
PR2_AAA34103.1_Tobacco	3E-68	45.17	Tc03_p029620	Tc03_g029620 Putative glucan endo-1,3-beta-glucosidase GVI (Fragment)
PR2_AAA34103.1_Tobacco	1E-65	42.19	Tc02_p023780	Tc02_g023780 Putative Glucan endo-1,3-beta-glucosidase, basic isoform
PR2_AAA34103.1_Tobacco	4E-63	38.12	Tc09_p008700	Tc09_g008700 Glucan endo-1,3-beta-glucosidase 13
PR2_AAA34103.1_Tobacco	3E-61	36.88	Tc09_p011610	Tc09_g011610 Putative Glucan endo-1,3-beta-glucosidase 11
PR2_AAA34103.1_Tobacco	1E-58	41.46	Tc08_p009900	Tc08_g009900 Putative glucan endo-1,3-beta-glucosidase GVI (Fragment)
PR2_AAA34103.1_Tobacco	3E-58	37.19	Tc06_p001730	Tc06_g001730 Predicted protein (Fragment)
PR2_AAA34103.1_Tobacco	1E-56	36.42	Tc00_p054290	Tc00_g054290 Glucan endo-1,3-beta-glucosidase 14
PR2_AAA34103.1_Tobacco	5E-55	37.46	Tc04_p012520	Tc04_g012520 Glucan endo-1,3-beta-glucosidase 2
PR2_AAA34103.1_Tobacco	6E-55	35.93	Tc01_p017650	Tc01_g017650 Putative Glucan endo-1,3-beta-glucosidase 3
PR2_AAA34103.1_Tobacco	6E-55	36.45	Tc02_p007080	Tc02_g007080 Glucan endo-1,3-beta-glucosidase 7
PR2_AAA34103.1_Tobacco	2E-54	37.69	Tc08_p001980	Tc08_g001980 Glucan endo-1,3-beta-glucosidase 7
PR2_AAA34103.1_Tobacco	2E-53	35.38	Tc03_p022650	Tc03_g022650 Glucan endo-1,3-beta-glucosidase 11
PR2_AAA34103.1_Tobacco	2E-52	37.58	Tc09_p031660	Tc09_g031660 Glucan endo-1,3-beta-glucosidase
PR2_AAA34103.1_Tobacco	2E-52	37.19	Tc03_p021200	Tc03_g021200 Glucan endo-1,3-beta-glucosidase
PR2_AAA34103.1_Tobacco	3E-52	36.12	Tc07_p002650	Tc07_g002650 Putative Glucan endo-1,3-beta-glucosidase 12
PR2_AAA34103.1_Tobacco	1E-51	34.8	Tc06_p012580	Tc06_g012580 Glucan endo-1,3-beta-glucosidase 11
PR2_AAA34103.1_Tobacco	9E-51	34.57	Tc02_p028070	Tc02_g028070 Glucan endo-1,3-beta-

				glucosidase 3
PR2_AAA34103.1_Tobacco	7E-50	33.02	Tc06_p019320	Tc06_g019320 Glucan endo-1,3-beta-glucosidase 1
PR2_AAA34103.1_Tobacco	2E-49	32.62	Tc01_p021070	Tc01_g021070 Probable glucan endo-1,3-beta-glucosidase A6
PR2_AAA34103.1_Tobacco	4E-48	35.4	Tc09_p013490	Tc09_g013490 Putative Glucan endo-1,3-beta-glucosidase
PR2_AAA34103.1_Tobacco	2E-47	36.48	Tc10_p003270	Tc10_g003270 Putative Glucan endo-1,3-beta-glucosidase
PR2_AAA34103.1_Tobacco	6E-47	33.95	Tc08_p005240	Tc08_g005240 Putative Glucan endo-1,3-beta-glucosidase 13
PR2_AAA34103.1_Tobacco	2E-46	33.93	Tc00_p034720	Tc00_g034720 Glucan endo-1,3-beta-glucosidase 6
PR2_AAA34103.1_Tobacco	2E-46	32.21	Tc01_p010310	Tc01_g010310 Glucan endo-1,3-beta-glucosidase 8
PR2_AAA34103.1_Tobacco	2E-45	33.13	Tc05_p028370	Tc05_g028370 Glucan endo-1,3-beta-glucosidase 8
PR2_AAA34103.1_Tobacco	2E-45	32.63	Tc04_p020310	Tc04_g020310 Probable glucan endo-1,3-beta-glucosidase A6
PR2_AAA34103.1_Tobacco	2E-44	31.17	Tc09_p021600	Tc09_g021600 Glucan endo-1,3-beta-glucosidase 4
PR2_AAA34103.1_Tobacco	2E-42	31.37	Tc09_p010240	Tc09_g010240 Glucan endo-1,3-beta-glucosidase 8
PR2_AAA34103.1_Tobacco	2E-41	32.6	Tc04_p023620	Tc04_g023620 Predicted protein
PR2_AAA34103.1_Tobacco	7E-41	30.3	Tc09_p034460	Tc09_g034460 Glucan endo-1,3-beta-glucosidase 9
PR2_AAA34103.1_Tobacco	1E-40	31.52	Tc01_p014030	Tc01_g014030 Glucan endo-1,3-beta-glucosidase 5
PR2_AAA34103.1_Tobacco	2E-40	32.27	Tc01_p010320	Tc01_g010320 Putative Glucan endo-1,3-beta-glucosidase 8
PR2_AAA34103.1_Tobacco	3E-40	31.6	Tc09_p000150	Tc09_g000150 Glucan endo-1,3-beta-glucosidase 5
PR2_AAA34103.1_Tobacco	5E-34	31.82	Tc09_p006080	Tc09_g006080 Glucan endo-1,3-beta-glucosidase 5
PR2_AAA34103.1_Tobacco	2E-31	28.86	Tc10_p002190	Tc10_g002190 Putative Glycoside hydrolase%2C family 17
PR2_AAA34103.1_Tobacco	0.0000006	23.94	Tc01_p006420	Tc01_g006420 Putative Probable glucan endo-1,3-beta-glucosidase A6
PR3P_P17514.1_endochitinase	3E-73	54.96	Tc02_p003890	Tc02_g003890 Endochitinase 1
PR3P_P17514.1_endochitinase	5E-73	53.85	Tc01_p000800	Tc01_g000800 Chitinase 1
PR3P_P17514.1_endochitinase	1E-53	41.44	Tc01_p032950	Tc01_g032950 Chitinase 10
PR3P_P17514.1_endochitinase	1E-31	34.76	Tc04_p018160	Tc04_g018160 Endochitinase PR4
PR3P_P17514.1_endochitinase	6E-28	29.64	Tc04_p029180	Tc04_g029180 Putative Endochitinase
PR3P_P17514.1_endochitinase	1E-27	33.6	Tc04_p018100	Tc04_g018100 Endochitinase PR4
PR3Q_116341_endochitinase	4E-89	66.67	Tc01_p000770	Tc01_g000770 Endochitinase
PR3Q_116341_endochitinase	3E-61	52.44	Tc01_p010350	Tc01_g010350 Basic endochitinase
PR3Q_116341_endochitinase	6E-34	32.41	Tc06_p000490	Tc06_g000490 Putative Endochitinase
PR3Q_116341_endochitinase	2E-31	36.6	Tc04_p018090	Tc04_g018090 Endochitinase PR4
PR3Q_116341_endochitinase	3E-27	36.73	Tc04_p018110	Tc04_g018110 Endochitinase PR4
PR4A_CAA41437.1_Tobacco	2E-47	70.59	Tc00_p012980	Tc00_g012980 Pro-hevein
PR4A_CAA41437.1_Tobacco	1E-44	67.8	Tc05_p027320	Tc05_g027320 Pro-hevein
PR4A_CAA41437.1_Tobacco	2E-28	47.7	Tc05_p027220	Tc05_g027220 Pathogenesis-related protein P2

PR4A_CAA41437.1_Tobacco	6E-28	48.51	Tc05_p027230	Tc05_g027230 Pathogenesis-related protein P2
PR4A_CAA41437.1_Tobacco	7E-26	50.93	Tc05_p027250	Tc05_g027250 Pathogenesis-related protein P2
PR4B_CAA41438.1_Tobacco	3E-55	83.9	Tc05_p027210	Tc05_g027210 Pathogenesis-related protein PR-4B
PR4B_CAA41438.1_Tobacco	5E-31	53.19	Tc10_p011130	Tc10_g011130 Pathogenesis-related protein P2
PR4B_CAA41438.1_Tobacco	3E-11	62.22	Tc05_p027340	Tc05_g027340 Putative Predicted protein
PR5_BAA74546.2_Tobacco	5E-72	58.64	Tc03_p005540	Tc03_g005540 Thaumatin-like protein
PR5_BAA74546.2_Tobacco	2E-70	58.74	Tc01_p001580	Tc01_g001580 Thaumatin-like protein 1
PR5_BAA74546.2_Tobacco	4E-69	56.64	Tc08_p003730	Tc08_g003730 Thaumatin-like protein 1
PR5_BAA74546.2_Tobacco	7E-68	55.16	Tc02_p005190	Tc02_g005190 Pathogenesis-related protein 5
PR5_BAA74546.2_Tobacco	9E-67	54.75	Tc08_p001670	Tc08_g001670 Pathogenesis-related protein 5
PR5_BAA74546.2_Tobacco	2E-66	54.67	Tc01_p001590	Tc01_g001590 Thaumatin-like protein 1a
PR5_BAA74546.2_Tobacco	3E-66	55.56	Tc00_p056050	Tc00_g056050 Thaumatin-like protein 1a
PR5_BAA74546.2_Tobacco	7E-64	52.65	Tc00_p056070	Tc00_g056070 Thaumatin-like protein 1
PR5_BAA74546.2_Tobacco	4E-62	54.59	Tc08_p010190	Tc08_g010190 Thaumatin-like protein
PR5_BAA74546.2_Tobacco	9E-62	52.89	Tc00_p056060	Tc00_g056060 Thaumatin-like protein 1a
PR5_BAA74546.2_Tobacco	2E-61	53.15	Tc04_p020870	Tc04_g020870 Thaumatin-like protein 1
PR5_BAA74546.2_Tobacco	2E-60	50.45	Tc02_p005200	Tc02_g005200 Thaumatin-like protein 1
PR5_BAA74546.2_Tobacco	7E-60	48.2	Tc05_p004240	Tc05_g004240 Thaumatin-like protein
PR5_BAA74546.2_Tobacco	7E-60	53.78	Tc08_p003740	Tc08_g003740 Thaumatin-like protein 1
PR5_BAA74546.2_Tobacco	4E-57	49.55	Tc10_p002890	Tc10_g002890 Thaumatin-like protein
PR5_BAA74546.2_Tobacco	8E-53	49.32	Tc02_p003020	Tc02_g003020 Putative Pathogenesis-related protein 5
PR5_BAA74546.2_Tobacco	8E-52	42.6	Tc04_p008530	Tc04_g008530 Putative Thaumatin-like protein
PR5_BAA74546.2_Tobacco	7E-46	45.05	Tc03_p027030	Tc03_g027030 Osmotin-like protein OSM34
PR5_BAA74546.2_Tobacco	4E-43	40.18	Tc09_p031980	Tc09_g031980 Osmotin-like protein
PR5_BAA74546.2_Tobacco	2E-40	41.89	Tc03_p026960	Tc03_g026960 Thaumatin-like protein
PR5_BAA74546.2_Tobacco	9E-38	38.46	Tc00_p060970	Tc00_g060970 Thaumatin-like protein
PR5_BAA74546.2_Tobacco	5E-37	42.34	Tc03_p027010	Tc03_g027010 Thaumatin-like protein
PR5_BAA74546.2_Tobacco	2E-34	39.91	Tc03_p026990	Tc03_g026990 Thaumatin-like protein
PR5_BAA74546.2_Tobacco	1E-27	32.59	Tc00_p056110	Tc00_g056110 Putative Thaumatin-like protein 1
PR5_BAA74546.2_Tobacco	5E-26	33.94	Tc03_p027000	Tc03_g027000 Thaumatin-like protein
PR5_BAA74546.2_Tobacco	3E-22	39.19	Tc03_p026980	Tc03_g026980 Thaumatin-like protein
PR5_BAA74546.2_Tobacco	2E-10	30.22	Tc09_p016370	Tc09_g016370 Putative Thaumatin-like protein
PR5_BAA74546.2_Tobacco	0.0000008	24.31	Tc04_p014480	Tc04_g014480 Putative Osmotin-like protein
PR5_BAA74546.2_Tobacco	0.000005	60.98	Tc03_p027020	Tc03_g027020 Putative Thaumatin%2C pathogenesis-related

Table S2 Primer sequences for SYBR green assay of PR genes analyzed by RT-qPCR

PR Family	Identifier in cacao genome database	Primer Sequences
TcPR-1	Tc02g002410	5’-CCTCAATGCTCACAACACGGCTC-3’
		5'-CGCTGCTCATTGCAAGGTTCTC-3'
TcPR-2	Tc04g029300	5'-CCTTGCTAACCCTTCCAATGCACAG-3'
		5'-CCAAGGCAGGCAAAACAAATTGAGC-3'
TcPR-3P	Tc01g000770	5'-GCACAACCGGTGACCTTACTACC-3'
		5'-CTGTTCCCTTATAAAGCAATATCCCCATGC-3'
TcChi4	Tc04g018110	5’-CGGCAACTGAATGTTGCAGTCGAT-3’
		5’-TGGTCCACGGCCATAGTAACCTTT-3’
TcPR-3Q	Tc04g018160	5'-GTTGCCAGTATCGTGTCACCTGC-3'
		5'-GGCATGTGCAAAAAAAGCAGCAATCTC-3'
TcChi1	Tc02g003890	5’-TTTCTCACGCACAGACAGTAGCGA-3’
		5’-TTGTAGCGAAGGCAGGGAAAGACT-3’
TcPR-4A	Tc05g027220	5'-GTGACCAACACTGCCACTGGAG-3'
		5'-GCATAGTCCACAGTAAGGTGACCTTG-3'
TcPR-4B	Tc05g027210	5'-CGCTTCCAATGTGAGAGCTACTTACC-3'
		5'-GACCACAAAAGGCCGTCCATCC-3'
TcPR-5	Tc03g026990	5'-CTGGACAGAGGTGGAGTCTGGAAC-3'
		5'-CATAGGCTTGGCACTGGAGGAG-3'

4.7 References

1. Fulton R: **The cacao disease trilogy: black pod, Monilia pod rot, and witches'-broom**. *Plant Disease* 1989:601.
2. Evans HC: **Cacao diseases - The trilogy revisited**. *Phytopathology* 2007, **97**(12):1640-1643.
3. Eastmond PJ: **Glycerol-insensitive Arabidopsis mutants: gli1 seedlings lack glycerol kinase, accumulate glycerol and are more resistant to abiotic stress**. *Plant J* 2004, **37**(4):617-625.
4. Chanda B, Venugopal SC, Kulshrestha S, Navarre DA, Downie B, Vaillancourt L, Kachroo A, Kachroo P: **Glycerol-3-Phosphate Levels Are Associated with Basal Resistance to the Hemibiotrophic Fungus Colletotrichum higginsianum in Arabidopsis**. *Plant Physiology* 2008, **147**(4):2017-2029.
5. Torres MA, Jones JDG, Dangl JL: **Reactive oxygen species signaling in response to pathogens**. *Plant Physiology* 2006, **141**(2):373-378.
6. Zurbriggen MD, Carrillo N, Hajirezaei MR: **ROS signaling in the hypersensitive response: when, where and what for?** *Plant Signal Behav* 2010, **5**(4):393-396.
7. Kachroo A, Fu DQ, Havens W, Navarre D, Kachroo P, Ghabrial SA: **An oleic acid-mediated pathway induces constitutive defense signaling and enhanced resistance to multiple pathogens in soybean**. *Molecular Plant-Microbe Interactions* 2008, **21**(5):564-575.
8. Jiang CJ, Shimono M, Maeda S, Inoue H, Mori M, Hasegawa M, Sugano S, Takatsuji H: **Suppression of the rice fatty-acid desaturase gene OsSSI2 enhances resistance to blast and leaf blight diseases in rice**. *Mol Plant Microbe Interact* 2009, **22**(7):820-829.
9. Maxwell DP, Wang Y, McIntosh L: **The alternative oxidase lowers mitochondrial reactive oxygen production in plant cells**. *Proc Natl Acad Sci U S A* 1999, **96**(14):8271-8276.

10. Zoeller M, Stingl N, Krischke M, Fekete A, Waller F, Berger S, Mueller MJ: **Lipid profiling of the Arabidopsis hypersensitive response reveals specific lipid peroxidation and fragmentation processes: biogenesis of pimelic and azelaic acid**. *Plant Physiol* 2012, **160**(1):365-378.
11. Argout X, Salse J, Aury JM, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN *et al*: **The genome of Theobroma cacao**. *Nature Genetics* 2011, **43**(2):101-108.
12. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic Local Alignment Search Tool**. *Journal of Molecular Biology* 1990, **215**(3):403-410.
13. Maximova SN, Marelli JP, Young A, Pishak S, Verica JA, Guiltinan MJ: **Over-expression of a cacao class I chitinase gene in Theobroma cacao L. enhances resistance against the pathogen, Colletotrichum gloeosporioides**. *Planta* 2006, **224**(4):740-749.
14. Yun D-J: **Plant Antifungal Proteins**. In., vol. 14; 1996: 39-88.
15. van Loon LC, Rep M, Pieterse CM: **Significance of inducible defense-related proteins in infected plants**. *Annu Rev Phytopathol* 2006, **44**:135-162.
16. Campos MA, Rosa DD, Teixeira JÉC, Targon MLPN, Souza AA, Paiva LV, Stach-Machado DR, Machado MA: **PR gene families of citrus: their organ specific-biotic and abiotic inducible expression profiles based on ESTs approach**. *Genetics and Molecular Biology* 2007, **30**:917-930.
17. Chanda B, Xia Y, Mandal MK, Yu K, Sekine KT, Gao QM, Selote D, Hu Y, Stromberg A, Navarre D *et al*: **Glycerol-3-phosphate is a critical mobile inducer of systemic immunity in plants**. *Nature Genetics* 2011, **43**(5):421-427.
18. Kachroo A, Venugopal SC, Lapchyk L, Falcone D, Hildebrand D, Kachroo P: **Oleic acid levels regulated by glycerolipid metabolism modulate defense gene expression in Arabidopsis**. *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**(14):5152-5157.
19. Kachroo A, Lapchyk L, Fukushige H, Hildebrand D, Klessig D, Kachroo P: **Plastidial fatty acid signaling modulates salicylic acid- and jasmonic acid-mediated defense pathways in the Arabidopsis ssi2 mutant**. *Plant Cell* 2003, **15**(12):2952-2965.

20. Sels J, Mathys J, De Coninck BMA, Cammue BPA, De Bolle MFC: **Plant pathogenesis-related (PR) proteins: A focus on PR peptides**. *Plant Physiology and Biochemistry* 2008, **46**(11):941-950.

21. Mandal MK, Chandra-Shekara AC, Jeong RD, Yu KS, Zhu SF, Chanda B, Navarre D, Kachroo A, Kachroo P: **Oleic Acid-Dependent Modulation of NITRIC OXIDE ASSOCIATED1 Protein Levels Regulates Nitric Oxide-Mediated Defense Signaling in Arabidopsis**. *Plant Cell* 2012, **24**(4):1654-1674.

22. Shah J, Kachroo P, Nandi A, Klessig DF: **A recessive mutation in the Arabidopsis SSI2 gene confers SA- and NPR1-independent expression of PR genes and resistance against bacterial and oomycete pathogens**. *Plant Journal* 2001, **25**(5):563-574.

23. Jambunathan N: **Determination and Detection of Reactive Oxygen Species (ROS), Lipid Peroxidation, and Electrolyte Leakage in Plants**. *Plant Stress Tolerance: Methods and Protocols* 2010, **639**:291-297.

24. Folden DV, Gupta A, Sharma AC, Li SY, Saari JT, Ren J: **Malondialdehyde inhibits cardiac contractile function in ventricular myocytes via a p38 mitogen-activated protein kinase-dependent mechanism**. *Br J Pharmacol* 2003, **139**(7):1310-1316.

25. Yu K, Soares Juliana M, Mandal Mihir K, Wang C, Chanda B, Gifford Andrew N, Fowler Joanna S, Navarre D, Kachroo A, Kachroo P: **A Feedback Regulatory Loop between G3P and Lipid Transfer Proteins DIR1 and AZI1 Mediates Azelaic-Acid-Induced Systemic Immunity**. *Cell Reports* 2013, **3**(4):1266-1278.

26. Wang C, El-Shetehy M, Shine MB, Yu K, Navarre D, Wendehenne D, Kachroo A, Kachroo P: **Free radicals mediate systemic acquired resistance**. *Cell Rep* 2014, **7**(2):348-355.

27. Yang F, Hanna MA, Sun R: **Value-added uses for crude glycerol--a byproduct of biodiesel production**. *Biotechnol Biofuels* 2012, **5**:13.

28. Shi Z: **Functional analysis of Non Expressor of PR1 (NPR1) and its paralog NPR3 in Theobroma cacao and Arabidopsis thaliana** University Park: Pennsylvania State University; 2010.

29. Dann E, Diers B, Byrum J, Hammerschmidt R: **Effect of treating soybean with 2,6-dichloroisonicotinic acid (INA) and benzothiadiazole (BTH) on seed yields and the level of disease caused by Sclerotinia sclerotiorum in field and greenhouse studies**. *European Journal of Plant Pathology* 1998, **104**(3):271-278.

30. Mejia L, Guiltinan M, Shi Z, Landherr L, Maximova S: **Expression of designed antimicrobial peptides in *Theobroma cacao* L. trees reduces leaf necrosis caused by *Phytophthora* spp**. In: *Acs Sym Ser.* vol. 1095: American Chemical Society; 2012: 379-395.

CHAPTER 5: CONCLUSIONS AND PERSPECTIVES

5.1 Major Conclusions

5.1.1 TcLEC2 is the Functional Ortholog of Arabidopsis AtLEC2 and Functions in the Control of Embryogenesis in Cacao

In my thesis project, I functionally characterized the significant roles of TcLEC2 during cacao SE and established its orthology to AtLEC2. Both encode B3 domain transcription factors that are involved in a genetic regulatory network with many other embryo specific genes that coordinate the development of zygotic and somatic embryos. Several key lines of evidence support this conclusion:

1) The expression of *TcLEC2* was higher in dedifferentiated cells competent to somatic embryogenesis (embryogenic calli), compared to non-embryogenic calli.

2) The overexpression of *TcLEC2* in cacao explants greatly increased the frequency of regeneration of stably transformed somatic embryos.

3) *TcLEC2* overexpressing cotyledon explants exhibited a very high level of embryogenic competency and when cultured on hormone free medium, a reiterating process coined an “embryonic chain-reaction”.

My study provides a deeper knowledge of the regulatory networks controlling embryogenesis and will contribute insight to the improvement of the SE process and acceleration of breeding programs.

5.1.2 Activity of TcSAD1 is Associated with Fatty Acid Composition in Cacao Seeds

In this study, I surveyed a large gene family in cacao that encodes the stearoyl-acyl carrier protein-desaturases (SAD) that convert stearic acid into oleic acid. The activities of this

gene family are critical not only for the central fatty acid metabolism, but also in determining the final fatty acid composition in cocoa butter. Our data suggested that:

1) The cacao genome contains eight putative SAD isoforms with high amino acid sequence similarities and functional domains conserved with SAD genes from other species.
2) *TcSAD1* is universally expressed across all the tissues. Particularly in cacao developing seeds, its expression was highly associated with the alteration of fatty acid composition.
3) The activity of TcSAD1 successfully rescued all AtSSI2 related phenotypes in Arabidopsis *ssi2* mutant background, further supporting the functional orthology between these two genes.
4) *TcSAD3* and *TcSAD4* appear to be exclusively and highly expressed in flowers; however, their functions there remain unknown.

My study revealed the significant roles of TcSAD1 in determining the final fatty acid composition of cocoa butter and its potential as a biomarker for screening of cacao varieties with desirable fatty acid profiles.

5.1.2 Applying Glycerol as a Foliar Spray to Enhance Disease Resistance of Cacao

In the present study, I studied the effects of glycerol on fatty acid biosynthesis and defense responses in cacao leaves, and explored the potential of applying glycerol as a foliar spray to enhance disease resistance of *Theobroma cacao*. Key findings include:

1) Foliar application of glycerol elevated the endogenous level of G3P and lowered the level of 18:1.
2) Glycerol application induced ROS production (a marker of defense activation) and expression of many pathogenesis related genes.
3) The effects of glycerol application on G3P, 18:1 fatty acid content and on gene expression levels in cacao leaves were dosage dependent. Treatments of glycerol higher than 300mM led to chlorosis and cell death.
4) 100 mM glycerol application was sufficient to stimulate defense responses without causing any observable damage and resulting in significant enhanced resistance against the cacao pathogen *Phytophthora capsici*.

5) The effects of glycerol treatment were constrained to the locally treated leaves without affecting distal tissues.

Our study demonstrated that glycerol has a great potential to become as disease control agent as part of worldwide sustainable farming system.

5.2 Implications for Future Research

5.2.1 Use of TcLEC2 to Improve the Efficiency of Cacao SE

Somatic embryogenesis (SE) is a favorable propagation system for many crop species because of its inherent high multiplication rate and the potential to generate genetically identical, disease-free plant materials. However, researchers realized that embryogenic potentials vary dramatically among different types of cacao somatic cells and different cacao genotypes although theoretically every somatic plant cell has all the genetic materials needed to dedifferentiate and re-differentiate into a whole new plant (totipotency). Notably, TcLEC2 appears to be a master regulator of SE, determining the competencies of plant cells to enter the somatic embryogenesis developmental pathway. Transgenic embryos overexpressing *TcLEC2* exhibited much higher SE efficiency and longer embryogenic capacity compared to control explants (Figure 3-6), implying that it is possibly to increase the embryogenic potential and efficiency of SE by elevating the activity of TcLEC2. However, because the constitutive overexpression of TcLEC2 in the transgenic embryos caused abnormal phenotypes and the development of most embryos ceased at globular stage, it is equally important to precisely control the activity of TcLEC2.

5.2.1.1 Regulated Expression of TcLEC2

Glucocorticoid receptor (GR) inducible system seems to be a suitable technique that enables us to control the subcellular localization of TcLEC2 protein. In absence of glucocorticoid, GR locates in the cytosol in the form of a protein complex (Aoyama and Chua, 1997). After binding with a glucocorticoid, such as dexamethasone (DEX), GR-glucocorticoid complex will translocate into nucleus. Therefore, the activity of TcLEC2_GR fusion protein can be induced by adding DEX into media and shut down by removing DEX from the media. According to the expression pattern of *TcLEC2* during cacao SE (Figure 3-

4), we can try to induce the activity of TcLEC2 either at around 40 days during the embryo initiation stage to increase embryogenesis efficiency, or earlier than the natural expression pattern to speed up the SE process. Cacao explants could be cultured on the DEX-containing media to maintain the elevated level of TcLEC2 and high embryogenic efficiency. However, developing embryos probably need to be cultured on DEX-free media to shut down the activity of TcLEC2 and thus achieve a normal embryo development. For those non-embryogenic cacao genotypes, elevating the activity of TcLEC2 might also be helpful to enhance their embryogenic potential and induce SE.

5.2.1.2 Use of TcLEC2 to Program Embryogenesis in New Tissue Explants

We also found that transient overexpression of *TcLEC2* in cacao stage C leaves is sufficient to activate the expression of many seed specific and embryogenic related genes, such as ABI3, LEC1, and AGL15 (Figure 3-3). It inspires us to think of the possibility to induce SE from other cacao tissues, such as leaves. In the current protocol, only petals, staminodes, and somatic cotyledons, are embryogenic, which limits the availability of the sources for the explants. To test this hypothesis, leaves from stable transgenic cacao plant overexpressing *TcLEC2* could be used as the source of the explant to induce primary embryogenesis under the current lab protocol (Maximova et al., 2005). Special attention should be paid on the concentration of auxin in the culture media since the interaction of TcLEC2 activity and auxin biosynthesis and accumulation has been widely reported. Hormone free SE induction media should also be considered if necessary. Furthermore, if the hypothesis is supported that somatic embryos could be induced from leaf explants, leaves from transgenic plant overexpressing *TcLEC2_GR* should be used to precisely control the activity of TcLEC2 to avoid the abnormal phenotypes of somatic embryos. Thus, taking the advantage of TcLEC2, if a suitable regeneration system could be developed, leaves are abundant source of explants for initiation of SE cultures.

5.2.1.3 Interaction between TcLEC2 and Other Embryogenic Specific Genes

To extent the scope of the knowledge of regulatory network of SE in cacao, comprehensive functional characterization of embryogenic specific genes could be conducted through a similar research pipeline. Meanwhile, taking advantage of GR inducible system, many TcLEC2 downstream target genes could be identified, which may specify the master

regulatory functions of TcLEC2 in different biological and metabolic pathways. Characterization of these embryogenic related genes will not only provide us with more biomarkers for the breeding program, but also offer the possibility to precisely control the SE process of caccao.

5.2.2 Functional Characterization of *TcSAD* Isoforms in Cacao

5.2.2.1 Develop *TcSAD1* as a Biomarker for Fatty Acid Composition in Cocoa Butter

It is well noted that cocoa butter originated from different growing regions may have different fatty acid compositions. In order to develop *TcSAD1* as a useful gene marker associated with fatty acid compositions of cocoa butter, it is necessary to conduct a comprehensive survey on a representative cacao population to evaluate the correlation between the expression level and gene sequence of *TcSAD1*, and fatty acid profiles of each cacao variety. This correlation could also be evaluated in the F1 or F2 progenies from the cross of two cacao varieties with distinct fatty acid profiles. Furthermore, it has been well noted that cocoa butter obtained from different geographic regions may has different fatty acid profiles and hardness characteristics. Regarding this, some research has revealed the effect of temperature on the expression level of *SAD* genes and fatty acid compositions in developing soybean (Byfield and Upchurch, 2007). Likewise, evaluating effects of temperature on the expression levels of *TcSAD1* could help us better understand the mechanism controlling the fatty acid composition of cocoa butter. Meanwhile, we could not rule out the possibility that other upstream or/and downstream enzymes on the fatty acid biosynthesis pathway also contribute to determine the final fatty acid composition of cocoa butter, such as ketoacyl-ACP synthase II (*KASII*), fatty acyl-thioeasterase (*FatA* and *FatB*), and delta12-fatty acid desaturase (*FAD2*), all of which have been previously identified in cacao genome (Argout et al., 2011).

5.2.2.2 Functions of *TcSAD* Isoforms in Cacao Flowers

Even though eight *TcSAD* isoforms share highly conserved amino acids sequences, there are significantly differences in their expression patterns across various cacao tissues. Both unopened and open flowers contain significantly high expression levels of *TcSAD* genes,

and both *TcSAD3* and *TcSAD4* appear to be flower specific *TcSAD* isoforms. However, it is still unknown why cacao floral tissues require such a high level of TcSAD activities, what chemical compound is synthesized through the activities of TcSAD isoforms, and what the functions of those compounds are in cacao flowers. Some recent studies shed some lights on how *SAD* genes are involved in ovule development in tobacco (Zhang et al., 2014) and pollinator attraction in orchids (Schluter et al., 2011). Thus, it is reasonable to speculate that high expression levels of *TcSAD* isoforms in cacao flowers might also be associated with pollination process, since most of cacao varieties are self-incompatible and the involvement of certain pollinators are essential to successfully complete the pollination. Notably, it was discovered that volatile organic compounds produced by cacao flowers were rich in 1-pentadecene and 7-tetradecene, even though the specific roles of both compounds in pollinator attraction have not been investigated (Young and Severson, 1994). In this aspect, the activity of TcSAD isoforms might contribute to produce both 1-pentadecene and 7-tetradecene, since the desaturation step is required in the biosynthesis pathway of alkene species. All in all, identification of the key visual and scent cues during cacao pollination process could assist us overcome the limitation of low pollination efficiency and potentially increase cacao yields.

5.2.3 The Potential of Applying Glycerol as a Disease Control Reagent in Fields

5.2.3.1 Field Study of Glycerol Effects on Cacao Plants

Like many other economically important crop species, yield loss caused by several severe diseases of cacao is significant. In this thesis, we explored the possibility of applying glycerol as a foliar spray to enhance disease resistance in cacao leaves. It is necessary to extend our research from petri dish pathogen assay to a field study to further prove that glycerol spray could provide sufficient resistance against pathogens in more complicated environments. More fungal pathogens, such as *Phytophthora megakarya*, and *palmivora*, *Moniliophthora perniciosa* and *roreri*, could be included in this study according to the local environments of cacao plantations. In our study, we showed that 100 mM is sufficient to induce the defense response without cause any visible damages to cacao leaves. More

detailed characterization of physiological changes of those leaves after glycerol spray could be conducted, such as photosynthetic rate measurement, to evaluate the effect of exogenous glycerol spray on the productivity of the plants. Last but not least, since several major severe diseases in cacao are pod related diseases, it will be also worthy to study the diffusion efficiency of glycerol into the pods and effects of glycerol on inducing defense responses in pods.

5.2.3.2 Defense Signaling Pathway in Cacao and SAR

In Arabidopsis, exogenous glycerol application can induce the defense responses via a fatty acid derived signaling pathway (Shah et al., 2001; Kachroo and Kachroo, 2009). In this regard, levels of those important defense related phytohormones could be measured after glycerol spray, such as SA, JA, and ET, which will help us better understand the mechanism since glycerol-3-phosphate (G3P), a bioactive form of glycerol in plant cells, is able to serve as a systemic signal to induce systemic acquired resistance (SAR)of the protection provided by glycerol and to assist the estimation of the effectiveness of glycerol against different types of pathogens. Additionally, to protect the plants from further pathogen infection (Chanda et al., 2011), further efforts should be made in characterize the SAR possibly induced by glycerol application in cacao. Although no induction of G3P in the distal leaves was observed in our study, it is most likely due to the absence of pathogen infection in our greenhouse during the glycerol application, which is necessary to facilitate the translocation of G3P and thus confer the SAR. Therefore, glycerol application in the field offers us a great opportunity to evaluate the SAR since those plants have been exposed to a complex environment with multiple pathogens.

5.3 Concluding Remarks

This thesis contributes to the knowledge of three major areas of cacao research: somatic embryogenesis, cocoa butter biosynthesis and disease resistance. Taking advantage of the genome resources and a translational biology approach, I identified the key genes on different regulatory networks and metabolic pathways, and detailed characterized the functions of these genes, such as *TcLEC2* and *TcSAD1*. This knowledge not only assists us

elucidate the complicated biological mechanisms, but also offers us great opportunities to further improve cacao varieties through the breeding programs.

5.4 References

Aoyama T, Chua N-H (1997) A glucocorticoid-mediated transcriptional induction system in transgenic plants. The Plant Journal **11:** 605-612

Argout X, Salse J, Aury JM, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN, Abrouk M, Murat F, Fouet O, Poulain J, Ruiz M, Roguet Y, Rodier-Goud M, Barbosa-Neto JF, Sabot F, Kudrna D, Ammiraju JS, Schuster SC, Carlson JE, Sallet E, Schiex T, Dievart A, Kramer M, Gelley L, Shi Z, Berard A, Viot C, Boccara M, Risterucci AM, Guignon V, Sabau X, Axtell MJ, Ma Z, Zhang Y, Brown S, Bourge M, Golser W, Song X, Clement D, Rivallan R, Tahi M, Akaza JM, Pitollat B, Gramacho K, D'Hont A, Brunel D, Infante D, Kebe I, Costet P, Wing R, McCombie WR, Guiderdoni E, Quetier F, Panaud O, Wincker P, Bocs S, Lanaud C (2011) The genome of Theobroma cacao. Nat Genet **43:** 101-108

Byfield GE, Upchurch RG (2007) Effect of Temperature on Delta-9 Stearoyl-ACP and Microsomal Omega-6 Desaturase Gene Expression and Fatty Acid Content in Developing Soybean Seeds. Crop Sci. **47:** 1698-1704

Chanda B, Xia Y, Mandal MK, Yu K, Sekine K, Gao Q-m, Selote D, Hu Y, Stromberg A, Navarre D, Kachroo A, Kachroo P (2011) Glycerol-3-phosphate is a critical mobile inducer of systemic immunity in plants. Nature Genetics **43:** 421-427

Kachroo A, Kachroo P (2009) Fatty Acid-derived signals in plant defense. Annu Rev Phytopathol **47:** 153-176

Maximova SN, Young A, Pishak S, Miller C, Traore A, Guiltinan MJ (2005) Integrated System for Propagation of Theobroma cacao L. *In* SM Jain, P Gupta, eds, Protocol for Somatic Embryogenesis in Woody Plants, Vol 77. Springer Netherlands, pp 209-227

Schluter PM, Xu S, Gagliardini V, Whittle E, Shanklin J, Grossniklaus U, Schiestl FP (2011) Stearoyl-acyl carrier protein desaturases are associated with floral isolation in sexually deceptive orchids. Proc Natl Acad Sci U S A **108:** 5696-5701

Shah J, Kachroo P, Nandi A, Klessig DF (2001) A recessive mutation in the Arabidopsis SSI2 gene confers SA- and NPR1-independent expression of PR genes and resistance against bacterial and oomycete pathogens. Plant Journal **25:** 563-574

Young A, Severson D (1994) Comparative analysis of steam distilled floral oils of cacao cultivars (Theobroma cacao L., Sterculiaceae) and attraction of flying insects: Implications for aTheobroma pollination syndrome. Journal of Chemical Ecology **20:** 2687-2703

Zhang J, Li J, Garcia-Ruiz H, Bates PD, Mirkov TE, Wang X (2014) A stearoyl-acyl carrier protein desaturase, NbSACPD-C, is critical for ovule development in Nicotiana benthamiana. The Plant Journal**:** n/a-n/a

VITAE

Yufan Zhang

EDUCATION

2009 – 2014 Ph.D. in Plant Biology, Huck Institutes of the Life Sciences,
Minor in Statistics, Department of Statistics,
The Pennsylvania State University, PA, USA.

2005 – 2009 B.S. in Biological Sciences, Life Science Department,
Fudan University, Shanghai, China.

PUBLICATIONS

Yufan, Zhang; Philip, Smith; Siela, N, Maximova; Mark, J, Guiltinan. (2014) Applying glycerol as a foliar spray activates the defense response and enhances disease resistance of Theobroma cacao. Molecular Plant Pathology DOI: 10.1111/mpp.12158

Yufan, Zhang; Adam, Clemens; Siela, N, Maximova; Mark, J, Guiltinan. (2014) The Theobroma cacao B3 domain transcription factor TcLec2 plays a duel role in control of embyro development and maturation. BMC Plant Biology 14:106.

Yufan, Zhang; Siela, N, Maximova; Mark, J, Guiltinan. (2014) Characterization of a stearoyl-acyl carrier protein desaturase gene family from chocolate tree, Theobroma cacao L. (Manuscript in preparation)

Siela, N, Maximova; Sergio, Florez; Xiangling, Shen; Nicolas, Niemenak; **Yufan, Zhang**; Wayne, Curtis; Mark, J, Guiltinan. (2014) Genome-wide analysis reveals divergent patterns of gene expression during zygotic and somatic embryo maturation of Theobroma cacao L., The chocolate tree. BMC Plant Biology 14:185.

Luis, C, Mejía; Edward, Allen; Herre, Jed; P. Sparks; Klaus, Winter; Milton, N, García1; Sunshine, A, Van Bael; Joseph, Stitt; Zi, Shi; **Yufan, Zhang**; Mark, J, Guiltinan; Siela, N, Maximova. (2014) Pervasive effects of an endophytic fungus on host genetic and phenotypic expression in a tropical tree. Front. Microbiol.5:479.

Zi, Shi; **Yufan, Zhang**; Siela, N, Maximova; Mark, J, Guiltinan. (2013) TcNPR3 from Theobroma cacao functions as a repressor of the pathogen defense response. BMC Plant Biology 13: 204.

Argout, X.; Salse, J.; Aury, J.M.; Guiltinan, M.J.; Droc, G.; Gouzy, J.; Allegre, M.; Chaparro, C.; Legavre, T.; Maximova, S.N.; et al. (another 51 authors). (2011) The genome of Theobroma cacao. Nat Genet 43:101-108.