THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF CHEMICAL ENGINEERING

HETEROLOGOUS EXPRESSION OF PHYTOCYANIN-LIKE ARABINOGALACTAN PROTEIN FROM DIOSCOREA ROTUNDATA AS A POTENTIAL TREATMENT TO PROMOTE SOMATIC EMBRYOGENESIS

EVA MEI SHOUSE
SUMMER 2018

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Microbiology
with honors in Chemical Engineering

Reviewed and approved* by the following:

Wayne Curtis
Professor of Chemical Engineering
Thesis Supervisor and Honors Advisor

Lorraine Santy
Professor of Biochemistry and Molecular Biology
Faculty Reader

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

Somatic embryogenesis is the conversion of a somatic cell to a embryogenic cell, effectively creating a true clone. The Phytocyanin-like arabinogalactan (PLA) protein has been shown to enhance somatic embryogenesis of plant tissue in multiple plan species. Although the details of how PLA is involved in somatic embryogenesis are not known, its effectiveness makes the PLA protein a potential treatment to facilitate embryo formation in plant tissue culture. Furthermore, the non-glycosylated form of the protein is still functional, which opens the door for utilizing bacteria as an expression system to produce PLA protein. This thesis documents an effort to express *Dioscorea rotundata* PLA proteins in *E. coli* in order to promote somatic embryogenesis in yam tissue. Yam is an important 'orphan crop' of sub-Saharan Africa, where this technology will contribute to propagation systems being developed for this plant. The work starts with the identification of *D. rotundata* PLA genes and progressed to initial attempts at expression of DrPLA proteins in *E. coli*. Appendices of this thesis also describe assistance with other projects including characterization of yam culture 'contaminants' and likely endosymbionts as well as several other side projects.

TABLE OF CONTENTS

LIST OF FIGURES iii

LIST OF TABLES iv

ACKNOWLEDGEMENTS v

Chapter 1 Preface 1

Chapter 2 Background and Introduction 2

Chapter 3 Materials and Methods 5

- 3.1 Organisms 5
- 3.2 Plasmids 5
- 3.3 Oligonucleotides 6
- 3.4 Molecular Cloning Methods 6
  - 3.4.1 Restriction Digest Cloning 6
  - 3.4.2 Fusion PCR 6
  - 3.4.2 pJET Blunt-End Cloning 7
  - 3.4.3 TA Cloning 8
  - 3.4.5 DNA Sequencing 8
  - 3.4.6 Heat Shock Transformation 8
- 3.5 Protein Methods 9
  - 3.5.1 Protein Expression 9
  - 3.5.2 Protein Isolation 9
  - 3.5.4 SDS-PAGE 9

Chapter 4 Results and Discussion 10

- 4.1 Identification of DrPLA Gene Candidates 10
- 4.2 Construction of the DrPLA Coding Sequence DNA 11
  - 4.2.1 Amplification of DrPLA gDNA Sequence 11
  - 4.2.2 Removal of Intron from DrPLA 13
  - 4.2.3 Correcting SNP in DrPLA3 gene 14
- 4.3 Cloning DrPLA Coding Sequences into Expression Vectors 14
  - 4.3.1 pET14b-EURO4 14
  - 4.3.2 pET-EI:X 14
- 4.4 Expression of DrPLA in *E. coli* 15

Chapter 5 Future Work 16

5.1 Summary of Current Project Status 16
5.2 qPCR to Measure DrPLA Expression *in vivo* 16
5.3 Purification of DrPLA Protein 17
5.4 Treatment of *Dioscorea rotundata* tissue with DrPLA Protein 18

Appendix A Other Work 19

A.1 Contamination Identification 19
A.1.1 Yam Contamination 19
A.1.2 Cassava Contamination 20
A.1.3 Sparge Train Contamination 20
A.2 Studying Transport between Whitefly Egg and Plant Tissue with Fluorescent Proteins 21
A.3 Construction of an Alcohol Inducible Vector 23

Appendix B Sequences 24

B.1 Primer Sequences 24
B.2 PLA Amino Acid Sequences 25
B.3 DrPLA DNA Sequences 26
B.4 Contamination 16s rDNA Sequences 27
B.5 Signal Peptide and mCherry Sequences 30
B.5.1 Amino Acid Sequences 30
B.5.2 DNA Sequences 31
B.6 Alcohol Inducible Vector Sequences 32

Appendix C Methods Details 34

BIBLIOGRAPHY 36

LIST OF FIGURES

Figure 1: Structure of PLA Protein 3

Figure 2: Diagram of fusion PCR..... 7

Figure 3: DaPLA3 versus DrPLA3 Sequence 12

Figure 4: DrPLA1 and DrPLA3 DNA Sequences..... 13

Figure 5: Alcohol Inducible Vector Map 23

LIST OF TABLES

Table 1: PLA primers designed from *D. alata* transcriptome 24

Table 2: DrPLA3 primers designed from *D. rotundata* genome..... 24

Table 3: Internal primers for removal of DrPLA intron via fusion PCR 25

Table 4: Mismatch primer to correct SNP in DrPLA3 25

Table 5: Universal 16s primers..... 25

ACKNOWLEDGEMENTS

Thank you Dr. Curtis. You have given me so many opportunities over my entire undergraduate career, and I can't thank you enough. This goes beyond writing my honors thesis (which I most certainly couldn't have done without you); you gave me the chance to get involved with research—really cool research, I might add—and I don't think I would have been able to fall in love with science without that chance.

Chapter 1

Preface

This thesis describes the work I have performed in CurtisLab over the course of my undergraduate career. I was in and out of the lab since the summer before my freshman year of college in 2012. I first began by working on developing a method for the preparation of electrocompetent *Rhodobacter capsalatus*. I left at the end fall 2012, only to come back in the fall of 2015; this time, I worked on expressing squid ring tooth protein in cyanobacteria. After leaving once more, I came back for a much longer stay May 2017. This is when I began working on the project on which the bulk of my thesis will consist: production of *Dioscorea rotundata* PLA protein in *E. coli* for the enhancement of somatic embryogenesis in yam tissue. However, I have also had the opportunity to become involved in numerous other projects. This other work will be discussed in Appendix A. It includes the identification of contaminating microbes around the lab as well as likely endophytic bacteria from yam cultures. Another project I assisted with was the generation of transgenic plants expressing fluorescent protein (mCherry) in order to study transport between plant tissue and whitefly eggs, and the construction of an alcohol inducible vector for the controlled expression of viral replicases in plants.

Chapter 2

Background and Introduction

Agriculture was the key step in humanity that allowed our complex civilizations to develop. Although it can be easy to forget when most people buy their food at a grocery store, our lives depend on the ability of farms to continually grow crops and raise animals. Plant propagation is thus an old technology that retains its importance, although the technology has changed over the years. The oldest and most obvious form of plant propagation is via seeds. However, tissue culture propagation, also called micropropagation, is an alternative method that is often used in research but also used in agriculture for crops and ornamentals that are difficult to proliferate via seeds, seedless sterile hybrids, or clonal propagation is required for phenotype such as flower. Tissue culture allows the propagation of genetically identical plants, whereas seeds produce offspring that, while similar to their parents, are not clones. In addition, new plants can be generated without having to wait for the plants to reach maturity. This is especially advantageous in plants that take a long time to produce seeds (perennials, trees, e.g. coffee, chocolate). Tissue culture also can be used to prevent the accumulation of viruses and generally work with clean plant material for interstate & international transport.

The key components of plant tissue propagation are the hormones and growth regulators added to the media. These compounds promote the desired behavior of the cells, and can induce various changes such as somatic embryogenesis – however this very species specific and determined emperically. Current techniques for the regeneration of plants heavily depend on these hormones and growth regulators; however, the process takes a long time, and is often hit or

miss. Researchers are always looking for new and improved methods for initiating embryogenesis, as it will not only facilitate the generation of transgenic plants but may also provide better methods for propagation of plants in agricultural settings.

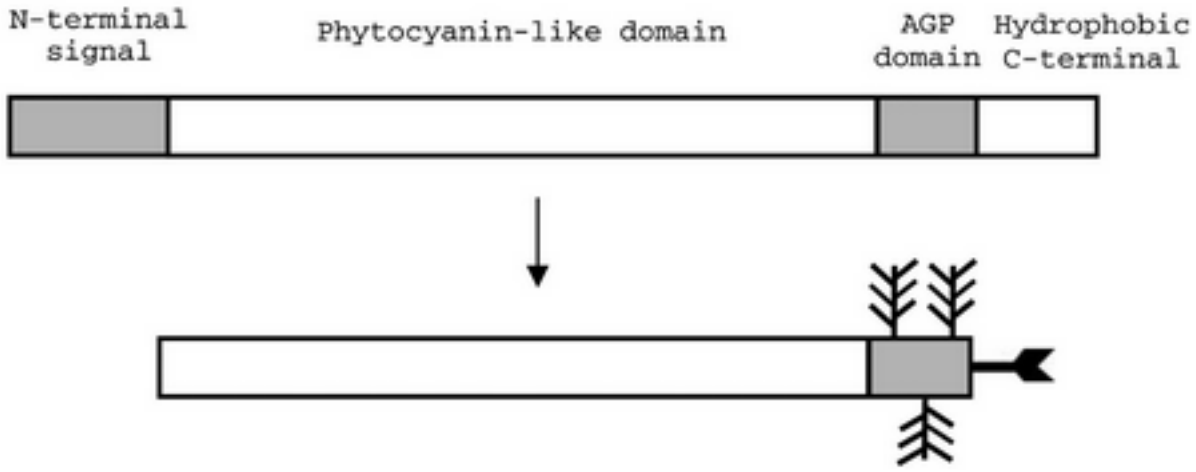

Figure 1: Structure of PLA Protein. Figure taken from Poon et al 2012.

The phytocyanin-like arabinogalactan protein, abbreviated as the PLA protein, was initially discovered to promote somatic embryogenesis within cotton (1). The PLA protein belongs to the arabinogalactan (AGP) family of proteins, which are known to be involved in somatic embryogenesis. However, even within the same plant, different tissues express different types of AGP's, and treating cells with these different AGP's has been shown to alter the growth behavior of the cells (2). In their experiments in cotton, Poon et al showed that the glycosylation was not necessary for functionality of the GhPLA1 protein. This opened the door for heterologous expression of the PLA protein in bacteria.

This research was conducted in an effort to test the versatility of the PLA protein as a method for promoting somatic embryogenesis in plants besides cotton. Previous work in the lab had been done with the PLA protein found in cacao and rice. When we received an NSF grant

for the development of new technologies for propagation of yam, the yam PLA protein became the next logical choice.

Chapter 3

Materials and Methods

3.1 Organisms

The molecular cloning work was performed in Top10 *E. coli* as the host organism. The protein expression work was performed in BL21 DE3 *E. coli*.

The species of yam from which the PLA genes were obtained is *Dioscorea rotundata*. However, the *Dioscorea alata* yam transcriptome DNA sequence was used for the initial *Dr*PLA protein bioinformatic searches before the *Dioscorea rotundata* genomic DNA sequence was available.

3.2 Plasmids

The pET14b vector was used for expression of the DrPLA proteins with and without a poly-Histidine tag. However, the actual plasmid used was pET14b-EURO4 from a previous lab project (Sergio Flores), which is the pET14b vector with the EURO4 gene in the multiple cloning site (3). This plasmid was used instead of native pET14b because the native pET14b plasmid in our lab was reported to be corrupted (Tina Lai).

The pET-EI:X vector was used for expressing the DrPLA protein with the intein solubility tag. (4)

Two cloning vectors were used throughout the molecular work in order to maintain DNA sequences: pJET2.1 from the CloneJET Kit and pCR2.1 from the TA Cloning Kit.

3.3 Oligonucleotides

All primers were synthesized by IDT. The full sequences are denoted in Appendix B.

3.4 Molecular Cloning Methods

3.4.1 Restriction Digest Cloning

Traditional cloning using restriction endonucleases and ligation was used to construct the expression vectors according to the restriction cloning methods in Sambrook (5).

3.4.2 Fusion PCR

Introns were removed from the DrPLA genes via fusion PCR, a method also known as splicing by overlap-extension PCR. (6) Two non-adjacent pieces of DNA can be fused together in only three PCR reactions with high-fidelity polymerase. The fusion PCRs performed in this work were done according to the “Fusion PCR Protocol” written by Katie Legenski and Victoria DeLeo, which can be found on the CurtisLab Google Drive.

7

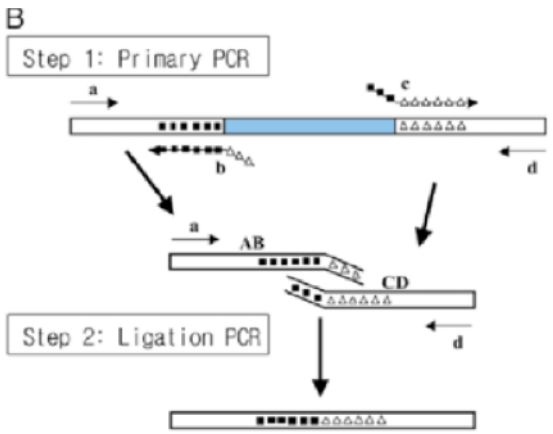

Figure 2: Diagram of fusion PCR. Taken from Lee et al 2004.

Both DrPLA1 and DrPLA3 had a single intron in the center of the gene. Primers were designed (Appendix B) that bracketed both regions of coding sequence, with the internal primers containing overhangs complementary to the other region.

3.4.2 pJET Blunt-End Cloning

The CloneJET PCR Cloning Kit from ThermoFisher was used to maintain DNA sequences within the pJET1.2 cloning vector. DNA from a PCR reaction using high fidelity polymerases is ligated into the pJET1.2 plasmid. Since the ends of the DNA fragments are blunt, there is no control over the orientation of the insert within the pJET1.2 vector. However, this does not matter when the vector is used as a way to maintain DNA. Furthermore, the pJET1.2 vector has positive selection that greatly reduces the number of false positive transformations. Overall, this has become the preferred cloning vector in our lab.

3.4.3 TA Cloning

The TA Cloning Kit from Invitrogen was used to maintain DNA sequences within the pCR2.1 vector. DNA directly from a PCR reaction using Taq polymerase is ligated into the pCR2.1 plasmid. Unlike the pJET cloning system, the TA Cloning Kit takes advantage of the Adenine nucleotides Taq adds on to the end of all PCR products. This allows the PCR products to be inserted into the pCR2.1 vector in a known orientation. However, the DNA must be from a PCR reaction using Taq polymerase no more than 24 hours old. Otherwise, the Adenine nucleotides begin to fall off the end of the DNA fragments. It is possible to add Adenine groups to the end of a DNA fragment by incubating the DNA with dATP and Taq polymerase. However, once we began using the CloneJET kit in our lab, I stopped using the TA Cloning Kit altogether.

3.4.5 DNA Sequencing

All plasmids were sequence verified via Sanger sequencing at the PSU DNA Sequencing center. The sequences were then uploaded to Benchling for analysis.

3.4.6 Heat Shock Transformation

All plasmids were transformed into *E. coli* via heat shock transformation according to the Heat Shock Transformation – Gold Standard SOP, which can be found on the CurtisLab Google Drive.

3.5 Protein Methods

3.5.1 Protein Expression

DrPLA1 protein was expressed in BL21 DE3 cells driven by the T7 promoter in the pET14b expression vector. Cells were grown up in Autoinduction media (Appendix C) with for 48 hours at 37C (7).

3.5.2 Protein Isolation

A crude protein extract was obtained from cells via freeze-thaw lysis. Cell cultures were spun down, and the pellets were resuspended in lysis buffer (see Appendix C for more details). Lysozyme was not used because at ~14.3 KDa it is too similar in size to the PLA protein (which is in the 10-15 KDa range) to be properly differentiated from PLA protein in SDS-PAGE. The resuspended cells were then exposed to six cycles of a three minute incubation in an ethanol-dry ice bath followed by a three-minute incubation in a 37C water bath. Once six freeze-thaw lysis cycles were completed, the cells were then spun down again. The supernatant was taken as the soluble protein isolate. The pellets were resuspended in lysis buffer, and this was taken as the insoluble protein isolate.

3.5.4 SDS-PAGE

Protein samples were analyzed via SDS-PAGE according to the SDS-PAGE Golden SOP written by Ryan Jones, which can be found on the CurtisLab Google Drive.

Chapter 4

Results and Discussion

4.1 Identification of DrPLA Gene Candidates

The PLA proteins native to *Dioscorea rotundata* were identified using the same method used by Tina Lai to identify PLA proteins in *Theobroma cacao* (8). The amino acid sequence for the *Gossypium hirsutum* (cotton) PLA protein GhPLA1 was used to search the *D. rotundata* genome via the NCBI tBLASTn search engine. However, since the *D. rotundata* genome had not yet been fully sequenced at the time, the GhPLA amino acid sequence was searched against the closely related *Dioscorea alata* transcriptome, which had been sequenced as part of a plant pathogen study (9). Several candidates' DNA sequences were identified using this method. Each candidate sequence was named according to similarity (homology) to the GhPLA1 protein based on E value: DrPLA1 was the most similar, DrPLA2 was the second most similar, and DrPLA3 was the third most similar, etc. The search was repeated by myself and Brielle Hohne with the *Dioscorea rotundata* genome when its sequence became available to determine the final numbering.

The two genes chosen for cloning work were DrPLA1 and DrPLA3. DrPLA2 was not chosen due to experiments with the PLA protein in *T. cacao* (chocolate tree). The first and third most homologous proteins in cacao, called TcPLA1 and TcPLA3, were found to be more effective than TcPLA2. Notably, DrPLA1 was most similar to TcPLA1, and DrPLA3 was most similar to DrPLA3, which further supports that these homologues might represent the most functional versions of the PLA protein. Collaborators at IITA Kenya are currently performing

qPCR to determine which DrPLA is the most expressed during embryogenesis, another indicator of functionality.

4.2 Construction of the DrPLA Coding Sequence DNA

4.2.1 Amplification of DrPLA gDNA Sequence

Once the DrPLA gene candidates were identified, primers were designed to amplify the DNA from *D. rotundata* genomic DNA. The primers were designed with restriction sites on either end to facilitate cloning the genes into a pET expression vector. Three sets of primers were designed for each gene; one to clone the PLA gene with no purification tags into pET14b, one to clone the PLA gene with a poly-Histidine tag into pET14b, and one to clone the gene into the pET-EI:X vector. The primer sequences can be found in Appendix B.

The *D. rotundata* genomic DNA was extracted from leaf tissue using the Promega Wizard Genomic DNA Purification kit. The gDNA was then used as template for PCR reactions to amplify the DrPLA genes. The following primer sets were used to amplify DrPLA1 from *D. rotundata* gDNA: DaPLA1 His(-) Fwd + DaPLA1 Rev, DaPLA1 His(+) + DaPLA1 Rev, DaPLA1 intein Fwd + DaPLA1 Rev. All three sets worked. The following primers sets were used to amplify DrPLA3 from *D. rotundata* gDNA: DaPLA2 His(-) Fwd + DaPLA2 Rev, DaPLA2 His(+) Fwd + DaPLA2 Rev, DaPLA2 intein Fwd + DaPLA2 intein Rev. However, none of the primer sets amplified the DrPLA3 gene. When the *D. rotundata* sequence became available, it was discovered that there were several key nucleotides that differed between *D. rotundata* and *D. alata* within the sequence used to design the DrPLA3 primers.

>org_DaPLA3
ATGGATACTGTTAAAGCTTTTGCTTGGCTCCTCTTCATGCTCCAAATTA
TGTCCTGCTCTTTAGCATATGAATTCCAGGTAGGTGGGAAAGATGGTTG
GGTTTTGAACCCTCATGAGAGTTACTCTCAATGGTCTGGCAGGAACCGG
TTCCAAGTCCATGACAAGCTTGGTATGAATTTCTTTCAAAACTTTCTTT
TATGAACTATATATGTGTTTTTAATCTGTGTTATATCAATGCAGTGTTT
AAGTACAAGAAGGAAGAGGACTCTGTTTTGGTGGTGAGCAAAGAGGACT
ATGACAAGTGCAATGTGAGCAACCCCATCAAGAAGTTTGATGATGGGAA
CTCAGTGTTTGAGTTTGATAAAACTGGTCCTTTCTTCTTCATCAGTGGA
GCTTCAGGGAAGTGCAATCAGAACCAAAAGCTTAGTGTGATTGTGCTGT
CTTTGAAGACTAAGAAACCCAGCACATCCCCTGCTCCTTCTCCAACCTC
CTCTGTTTCATCTCCACCACCATCATCATCATCATCACCATCACTATCA
CCATCATCTGCAGCTTCATCTCCATCATCATCACCATCATCTGCAGCTT
CATCTCCATCATCATCACCAGAAGTATCTCCTACTCCATCCATATCTAT
TTCTCCATCTGCAACAACGGGCTCCCCTGTTTCATCTCAATCACCATCA
CCATCTTCTGAAACTTCATCTGGGGTGCCTTCTTCATTGTCTCCAGGTC
AATCTAGCGGATCAACAAATGCAGACAGCTCTCCTGGAAGTTCACCTGC
AGGTCCACAGTCTCCTGCTCCATCTTTGGCTGCCACTTGGATGTCCTTT
GGATGGGTCACCTTGGTCTTTTGTGGCCTCTTTTTAGGCTGA

>Pseudo_Chromosome_20_P1_20_P2_17_stt_9079676_end
_9080586_len_911_reverse
ATGGATACTGTTAAAGCTTTTGCTTGGCTCCTCTTCATGCTCCAAATCA
TGTCCTGCTCTTTAGCATATGAATTCCAGGTGGGTGGGAAAGATGGTTG
GATTTTGAACCCTCATGAGAGTTACTCTCAATGGTCTGGCAGGAACCGG
TTCCAAGTCCATGACAAGCTTGGTATGAATTTCTTTCACAACTTTCTTT
TATGAACTATGTATATGTTTT.AATCTGTGTTATATCAATGCAGTGTTT
AAGTACAAGAAGGAAGAGGACTCTGTTTTGGTGGTGAGCAAAGAGGACT
ATGACAAGTGCAATGTGAGCAACCCCATCAAGAAGTTTGATGATGGGAA
CTCAGTGTTTGAGTTTGATAAAACTGGTCCTTTCTTCTTCATCAGTGGA
GCTTCAGGGAAGTGCAATCAGGGTCAAAAGCTCAGTGTGATTGTGCTGT
CTTTGAAGACTAAGAAACCCAGCACATCCCCTGCTCCTTCTCCAACCTC
CTCTGTTCCATCTCCACCACCATCATCATCATCGTCACCACCACCACCA

Figure 3: DaPLA3 versus DrPLA3 Sequence. The sequence highlighted yellow is the PLA domain being cloned. The sequence highlighted green is the intron. The sequence highlighted red are the SNP's between the *D. alata* and *D. rotundata* sequences.

By contrast, there were no SNP's between DaPLA1 and DrPLA1 in the primer binding sites. The primers were redesigned with the *D. rotundata* sequence, and these primers successfully amplified DrPLA3 from the *D. rotundata* genomic DNA (gDNA).

The amplicons were then gel extracted and ligated into the pCR2.1 cloning vector from the TA Cloning Kit, which was subsequently transformed into Top Ten *E. coli*. The DrPLA genes were then sent for sequencing within the pCR2.1 plasmid as DP1-TOPO and DP3-TOPO.

4.2.2 Removal of Intron from DrPLA

DrPLA1

GAGTTCTTGGTCGGAGGAGCTGCSGGAGCATGGAAAGTCCCATCATCA
CCTTCAGAATCACTAAACAAATGGGCCGAGATCAACCGTTTCCAAGTC
GGAGATTCTCTCGGTATATTTATTTACTCATTCTATATATCTTTGATTAT
AAAGTATTCGGCAAGATAAGAGACCCAAGATTTATGTGGTTTATATAA
TGTCTCTTTATTTATTTWWTTTTTTTATTGCAGTGTTTAAGTTTGCCGAC
AAGAAGGATTCAGTACTAAAAGTAAGCAGAGAAGATTACTTGGCTTGC
CGGATATCGGCGCCGTTGGAGGAGCACAAGGATGATGAGAAGAATGCC
CTTGTGAAGTTGGATAAGTCTGGTGCTTTTTACTTTGTAAGTGGAAACC
AGAAAGCTTGT

DrPLA3

GAATTCCAGGTGGGTGGGAAAGATGGTTGGATTTTGAACCCTCATGAG
AGTTACTCTCAATGGTCTGGCAGGAACCGGTTCCAAGTCCATGACAAG
CTTGGTATGAATTTCTTTCACAACTTTCTTTTATGAACTATGTATATGTT
TTAATCTGTGTTATATCAATGCAGTGTTTAAGTACAAGAAGGAAGAGG
ACTCTGTTTTGGTGGTGAGCAAAGAGGACTATGACAAGTGCAATGTGA
GCAACCCCATCAAGAAGTTTGATGATGGGAACTCAGTGTTTGAGTTTGA
TAAAACTGGTCCTTTCTTCTTCATCAGTGGAGCTTCAGGGAAGTGCAAT
CAGGGTCAAAAGCTCAGTGTGATTGTGCTGTCTT

Figure 4: DrPLA1 and DrPLA3 DNA sequences. Sequence highlighted in yellow is the coding sequence for the domain of interest. Sequence highlighted green is the intron to be removed via fusion PCR.

A single intron was identified in both DrPLA1 and DrPLA3. However, since bacteria cannot remove introns from genes, the intron had to be removed before the protein could be

expressed in *E. coli*. Fusion PCR was used to remove the introns from both genes. The intron-free DrPLA genes were then cloned into the pCR2.1 cloning vector.

4.2.3 Correcting SNP in DrPLA3 gene

After the DrPLA3 coding sequence was cloned into pCR2.1, an SNP was identified when the gene was sequenced. The SNP was corrected by using a mismatch primer with the correct nucleotide and Q5 high fidelity polymerase.

4.3 Cloning DrPLA Coding Sequences into Expression Vectors

4.3.1 pET14b-EURO4

Once the DrPLA1 and DrPLA3 coding sequences were confirmed in the pCR2.1 cloning vector, the genes were moved to the pET14b expression vector via restriction digest cloning. The plasmid was sequence verified before transforming into BL21 for protein expression. The vectors were named pET14b-DrPLA1 His+ and pET14b-DrPLA3 His+. The sequence information can be found on the Benchling files of the same name.

4.3.2 pET-EI:X

The sequence-verified DrPLA1 and DrPLA3 coding sequences were also cloned into pET-EI:X via restriction digest cloning. However, sequencing results showed that there were mutations relative to the reported vector sequence. Furthermore, these mutations were located

within the coding sequence for the intein protein, and one mutation resulted in an amino acid change from a Histidine to an Arginine. Because of the ambiguity of the source of the mutation (possibly in the provided vector) further work was put on hold at this stage, because the self-cleaving activity of the intein protein is required for the removal of the intein-ELP tag from the purified protein. The intein with the correct sequence has recently been obtained from a different vector containing IntF/2A (10), which would provide a logical path forward to complete this approach.

4.4 Expression of DrPLA in *E. coli*

In order to express protein, the pET14b-DrPLA1 His+ and pET14b-DrPLA3 His+ plasmids were transformed into the BL21 DE3 *E. coli*, which is designed for protein expression. The cells were then grown up in Autoinduction Media in 5mL cultures at 37C for 48 hours. A crude protein extract was then isolated from the cells via freeze-thaw lysis. The soluble and insoluble fractions were run on SDS-PAGE to check for expression of DrPLA. However, due to inexperience with SDS-PAGE, the results from the gels were inconclusive. It is still unclear whether or not the pET14b-DrPLA1 His+ and pET14b-DrPLA3 His+ plasmids successfully drive expression DrPLA1 and DrPLA3 respectively.

Chapter 5

Future Work

5.1 Summary of Current Project Status

As of right now, the *Dioscorea rotundata* PLA homologs, DrPLA1 and DrPLA3 are both in two different expression vectors: pET14b and pET-EI:X. These plasmids are in BL21 cells, and are therefore ready for protein production. However, expression of the DrPLA proteins has not yet been verified. Ongoing collaboration to identify highly expressed PLA homologs in embryogenic yam tissue will provide an excellent step forward to utilize this thesis work for production of the most promising PLA proteins for use as an additive to yam propagation bioreactor systems. The goal of creating a low-cost production of PLA using a self-cleaving intron / solubility tag fusion was stalled due to a sequence inconsistency that can now be corrected.

5.2 qPCR to Measure DrPLA Expression *in vivo*

The sequence search was able to identify DrPLA gene candidates, but it did not identify which proteins were expressed within *D. rotundata*. It is not necessarily true that the DrPLA protein most homologous to the one functional in cotton is the one most functional in yam. To measure actual expression of the various DrPLA gene candidates, working with Brielle Hohne. the sequences were sent to a collaborator to design qPCR experiments to measure expression of

the candidates within embryogenic *D. rotundata* tissue. The results from this experiment will show which DrPLA mRNA are most abundantly expressed within the organism, which is an indicator of protein functionality in tissues that are in the developmental state of somatic embryogenesis. While it is likely that DrPLA1 and DrPLA3 will be embryogenic candidates, the effort of this thesis has established the cloning approach and methods to rapidly progress alternative DrPLAx genes that are identified.

5.3 Purification of DrPLA Protein

There are two methods of protein purification being pursued. The first is the more standard His-tag purification. The DrPLA1 and DrPLA3 genes within pET14b will have a poly-Histidine tag at the C-terminal end of the protein that can be used to purify the DrPLA protein from the crude protein extract via absorption on a metal affinity resin. The other method uses the intein solubility tag, which consists of a self-cleaving intein protein fused to an ELP (elastin-like protein) at the N-terminal end of the protein. The ELP region allows for temperature dependent solubility of the protein, whereas the cleavage of the intein is controlled by pH. The result is a purification method that relies solely on centrifugation coupled with changes in temperature and pH (4). Following up this work is very worthwhile because it represents a very low-cost method for protein production, and while the purity of the protein would not be as high as affinity purification methods, such purity is not needed for an application like plant propagation.

5.4 Treatment of *Dioscorea rotundata* tissue with DrPLA Protein

Once the DrPLA1 and DrPLA3 protein has been obtained, it can be tested as a potential enhancer of somatic embryogenesis. After the proteins are produced heterologously, the purified protein can be applied to tissue culture where embryo production can be monitored and recorded.

Appendix A

Other Work

A.1 Contamination Identification

Bacterial contamination is a common problem in any biological lab. However, the repeated appearance of contamination in the same places presents an opportunity for study. In a supposedly axenic line of *Dioscorea cayenensis*, the frequent appearance of a white contamination suggested the possibility that the plants had retained endophytic bacteria. While this contaminant would not grow extensively in tissue culture, it would grow on permissive R2A media (11). In addition, bacterial growth in sterile water used to humidify air that was feeding the yam propagation bioreactors was interesting for another reason; the ability to grow with no added nutrients may be suggestive of the presence of a nitrogen-fixing bacteria. Without a carbon source it must be either using CO2 or vaporized organics from the air compressors. Contaminating bacteria were identified by their 16s rDNA sequences (12). Colony PCR was performed on the cultures using Universal 16s Primers (see Appendix B).

The resulting amplicons were then purified and sent for sequencing. The sequences were BLAST'd in order to identify their families, and in some cases, their species.

A.1.1 Yam Contamination

Several types of contamination were found in *Dioscorea cayenensis* cultures within our lab. However, the recurring nature of the contamination suggested the possibility that the contamination was endophytic symbiotic bacteria. Three species of bacteria were identified:

Bacillus aryabatthai, Bacillus ginsengihumi, and *Hyphomicrobium facile.* Bacillus species are commonly found associated with plants and the soil around them. Both identified *Bacillus* species are documented as plant growth promoting bacteria.

Additionally, *H. facile* is a methylotrophic species of bacteria. This exact species was independently identified twice from the *D. cayenensis* plants. Methylotrophs are common endophytes, as plants produce methanol as a biochemical waste product of pectin formation making it an idea carbon source for endosymbiotic microorganisms.

A.1.2 Cassava Contamination

At one point, a bright pink contamination was isolated from a cassava culture within the lab. The bacteria was streaked onto R2A plates, and colony PCR was performed on colonies from these plates. After 16s identification, it was determined to be within the *Methylobacterium* family. As suggested by the name, *Methylobacterium* are often methylotrophs. Furthermore, these bacteria are often associated with plants (13) for reasons noted above for the yam endosymbionts. The bacteria were inoculated into liquid R2A cultures and grown at room temperature. The liquid culture was cryopreserved with 50% glycerol at -80C for later study. A very profuse orange slimy ‘contaminant’ of musa (banana) and likely ensysmbiont (that would appear in stressed tissue) was not able to give 16s amplification.

A.1.3 Sparge Train Contamination

Two types of bacteria were isolated from contamination in the sparge train. The contaminated water was spread on an R2A plate, and the bacteria were streaked onto separate

plates based on the color. The bacteria were named based on their colors: “Sparge Beige” and “Sparge Yellow”. The sequence results from the 16s rDNA PCR showed that both species were from the *Mesorhizobium* family of bacteria. However, the difference in color does seem to reflect a difference in species. This class of bacteria is noted to be nitrogen fixing as was hypothesized based on its ability to grow in the sterile water.

A.2 Studying Transport between Whitefly Egg and Plant Tissue with Fluorescent Proteins

Whiteflies lay their eggs on the leaves of plants. A special structure, called the pedicel, has been shown to mediate transport of water and other solutes from the plant tissue into the egg (14). As yet, the exact range of particles that can be transported in this manner is not fully known. However, if protein can move from plant to whitefly egg in this manner, this provides a potential method for genetically modifying whiteflies via CRISPR-Cas9.

I assisted David Krum in the creation of transgenic plants expressing the mCherry fluorescent protein to visually track the movement of protein from tomato plant into whitefly eggs. Since the transported particles come from the apoplastic space of the plant, the protein must be excreted from the cell to reach the whitefly egg. Dr. Curtis examined various signal peptides responsible for targeting protein for extracellular transport. This search included the signal peptide for the cotton embryogenic protein GhPLA1. The signal peptides were rated using the SignalP tool, and the results are denoted in Table 1 below.

Table 1: Signal Peptide Search

Gene	AA Length	Y-score	D-score	Comment (all tomato sequences)

LePLA1	24	0.868	0.887	Embryogenic protein
Pgal-b	27	0.588	0.759	Ripening solubilization
Xgluc-hy	22	0.497	0.727	Cell wall construction
STIG-1	23	0.495	0.659	Stigma exudate

The signal peptides were rated by the Y-score (combined cleavage score), which predicts whether the region is actually a cleavage site, and the D-score (discrimination score), which discriminates between cleavage sites and non-cleavage sites. The PLA protein signal peptide had the highest rating and was therefore used for fusion to mCherry fluorescent protein. The Xgluc-hy signal peptide was also chosen as a back-up candidate. Although PgaI-b had a higher rating, Xgluc-hy was chosen instead based on its function within the plant. Whereas cell wall construction occurs all over the plant, ripening solubilization is most likely restricted to regions the fruit of the plant.

My role within this project was to fuse the chosen signal peptides to the mCherry coding sequence and then clone the fused gene into a binary vector. The vector chosen for this work was pEAQHT, as we had already been using it for other work in lab. Furthermore, we had a pEAQHT construct that also had P19 anti-gene silencing constitutively expressed under a 35S promoter. Fusion PCR was used to fuse the signal peptides to mCherry. Each signal peptide was fused to the 5’ end of the mCherry gene. The fused genes were then cloned into pEAQHT via restriction digest cloning.

A.3 Construction of an Alcohol Inducible Vector

As part of the DARPA project, it was desirable to make a transgenic tomato plant to express a viral replication associated protein (REP) that would complement a deconstructed virus without REP (REP-). As I was training in molecular biology methods with the finishing graduate student Tina Lai, I took on some effort to continue with this project. This was challenging since the original work with this vector was over 10 years ago with a minimally annotated vector map.

Figure 5: Alcohol Inducible Vector Map. The alcR gene is constitutively expressed under the 35S promoter. In the presence of ethanol, the alcR transcription factor is activiated and binds to the alcA promoter region to initiate transcription of its corresponding gene—in this case, C1/C2. The tobacco etch virus 5' UTR is placed between the alcA promoter and the gene to be expressed increases transcription. Figure taken from Zhang et al 2006.

Initially, the plan was to replace the C1/C2 gene in the pSRN-ARPR plasmid with the REP gene.(15) However, problems continued to arise in the course of working with this construct. Sam Kariuki and David Krum attempted to determine the content of the plasmid via sequencing, only to find that the sequencing results did not match with the reported vector map. As a result, we decided to shift gears and construct the alcohol inducible system in a new vector.

I created a cloning plan to move the alcohol inducible system into the pEAQHT binary vector. The AlcR gene would be expressed under a 35s promoter, while the alcA promoter would be used to regulate the expression of the desired viral replicase.

Appendix B

Sequences

B.1 Primer Sequences

Table 2: PLA primers designed from *D. alata* transcriptome

Primer Name	Primer Sequence (5' → 3')
DaPLA1 His(-) Fwd	ATACCATGGGAGAGTTCTTGGTCGGAGGA
DaPLA1 His(+) Fwd	ATTACATATGGAGTTCTTGGTCGGAGGA
DaPLA1 Rev	ATAGGATCCTTAACAAGCTTTCTGGTTTCCAC
DaPLA1 Intein Fwd	TGTTGTACACAACGAGTTCTTGGTCGGAGA
DaPLA1 Intein Rev	ATATCGGCTCAGCCCTTGTTAACAAGCTTTCTGG
DaPLA2 His(-) Fwd	ACACCATGGGAATTCCAGGTAGGTG
DaPLA2 His(+) Fwd	AGGACATATGGAATTCCAGGTAGGTG
DaPLA2 Rev	GGTGGATCCTTAACAAGCTTTCTGGTTTCCAC
DaPLA2 Intein Fwd	TGTTGTACACAACGAATTCCAGGTAGTAGGTG
DaPLA2 Intein Rev	ATATCGGCTCAGCTTAAGACAGCACAATCACACTAA

Table 3: DrPLA3 primers designed from *D. rotundata* genome

Primer Name	Primer Sequence (5' → 3')
DrPLA3 His(-) Fwd	ACACCATGGGAATTCCAGGTGGGTG
DrPLA3 His(+) Fwd	AGGACATATGGAATTCCAGGTGGGTG
DrPLA3 Rev	GGTGGATCCTTAAGACAGCACAATCACACTGA
DrPLA3 Intein Fwd	TGTTGTACACAACGAATTCCAGGTGGGTG
DrPLA3 Intein Rev	ATATCGGCTCAGCTTAAGACAGCACAATCACACTGA

Table 4: Internal primers for removal of DrPLA intron via fusion PCR

Primer Name	**Primer Sequence (5' → 3')**
DrPLA1 Fusion Rev	ACTTAAACACGAGAGAATCTCCGACTTGG
DrPLA1 Fusion Fwd	AGATTCTCTCGTGTTTAAGTTTGCCGACAAG
DrPLA3 Fusion Rev	TACTTAAACACAAGCTTGTCATGGACTTGG
DrPLA3 Fusion Fwd	GACAAGCTTGTGTTTAAGTACAAGAAGGAAGAGG

Table 5: Mismatch primer to correct SNP in DrPLA3

Primer Name	**Primer Sequence (5' → 3')**
DP3 SNP Rev	TCAGTGGAGCTTCAGGGAAGTGCAATCAGGGTCAAAAGCTCAGTGTGATTGTGCTGTCTT

Table 6: Universal 16s primers

Primer Name	**Primer Sequence (5' → 3')**
16s Fwd	GAGTTTGATCATGGCTCAG
16s Rev	CTACGGCTACCTTGTTACG

B.2 PLA Amino Acid Sequences

GhPLA1

MAAKAFSRSITPLVLLFIFLSFAQG/KEIMVGGKTGAWKIPSSESDSLNKWAEKARFQIGD
SLVWKYDGGKDSVLQVSKEDYTSCNTSNPIAEYKDGNTKVKLEKSGPYFFMSGAKGHC
EQGQKMIVVVMSQKHRYIGISPAPSPVDFEGPAVAPTSGVAGLKAGLLVTVGVLGLF

DrPLA1

26
MASXYWIAFSLFFLITSSQAKEFLVGGAXGAWKVPSSPSESLNKWAEINRFQVGDSLFKF
ADKKDSVLKVSREDYLACRISAPLEEHKDDEKNALVKLDKSGAFYFVSGNQKACEQGE
KLIVVVMSTRHRLAPAPSPVEFDGPAVAPTSSGHKVGMFGGGLVACLVVLFGLML

DrPLA3

MDTVKAFAWLLFMLQIMSCSLAYEFQVGGKDGWILNPHESYSQWSGRNRFQVHDKLV
FKYKKEEDSVLVVSKEDYDKCNVSNPIKKFDDGNSVFEFDKTGPFFFISGASGKCNQGQ
KLSVIVLSLKTKKPSTSPAPSPTSSVPSPPPSSSSSPPPPSSSSSPPSQSPSSAASSPSSSPSSA
ASSPSSSSSPEASPTPSISISPSATTGSPVSSQSPSPSSGASSGVPSSLSPGQSSGATNAASSP
GSSPAGPQSPAPALAATWMSFGWVTLVFCGLFLG

B.3 DrPLA DNA Sequences

DrPLA1

ATGGCATCACWCTACTGGATAGCATTTTCTCTCTTCTTCTTGATAACAAGCTCACAA
GCTAAAGAGTTCTTGGTCGGAGGAGCTGCSGGAGCATGGAAAGTCCCATCATCACCT
TCAGAATCACTAAACAAATGGGCCGAGATCAACCGTTTCCAAGTCGGAGATTCTCTC
GGTATATTTATTTACTCATTCTATATATCTTTGATTATAAAGTATTCGGCAAGATAAG
AGACCCAAGATTTATGTGGTTTATATAATGTCTCTTTATTTATTTWWTTTTTTTATTG
CAGTGTTTAAGTTTGCCGACAAGAAGGATTCAGTACTAAAAGTAAGCAGAGAAGAT
TACTTGGCTTGCCGGATATCGGCGCCGTTGGAGGAGCACAAGGATGATGAGAAGAA
TGCCCTTGTGAAGTTGGATAAGTCTGGTGCTTTTTACTTTGTAAGTGGAAACCAGAA
AGCTTGTGAACAAGGTGAAAAGCTTATTGTGGTGGTGATGTCCACGAGGCATAGGTT
GGCTCCGGCGCCTTCTCCGGTGGAGTTTGATGGACCGGCGGTTGCTCCGACGAGCAG
TGGTCATAAGGTTGGGATGTTTGGAGGTGGTCTTGTTGCTTGTTTGGTTGTGTTGTTT
GGGTTGATGCTTTGA

DrPLA3

ATGGATACTGTTAAAGCTTTTGCTTGGCTCCTCTTCATGCTCCAAATCATGTCCTGCT
CTTTAGCATATGAATTCCAGGTGGGTGGGAAAGATGGTTGGATTTTGAACCCTCATG
AGAGTTACTCTCAATGGTCTGGCAGGAACCGGTTCCAAGTCCATGACAAGCTTGGTA
TGAATTTCTTTCACAACTTTCTTTTATGAACTATGTATATGTTTTAATCTGTGTTATAT
CAATGCAGTGTTTAAGTACAAGAAGGAAGAGGACTCTGTTTTGGTGGTGAGCAAAG
AGGACTATGACAAGTGCAATGTGAGCAACCCCATCAAGAAGTTTGATGATGGGAAC
TCAGTGTTTGAGTTTGATAAAACTGGTCCTTTCTTCTTCATCAGTGGAGCTTCAGGGA
AGTGCAATCAGGGTCAAAAGCTCAGTGTGATTGTGCTGTCTTTGAAGACTAAGAAAC
CCAGCACATCCCCTGCTCCTTCTCCAACCTCCTCTGTTCCATCTCCACCACCATCATC
ATCATCGTCACCACCACCACCACCATCATCATCATCATCACCACCATCACAATCACC
ATCATCTGCAGCTTCATCTCCATCATCATCACCATCATCTGCAGCTTCATCTCCATCA
TCATCATCATCACCTGAAGCATCTCCTACTCCATCCATATCTATTTCTCCATCTGCAA

CAACGGGCTCCCCTGTTTCATCTCAATCACCATCACCATCTTCTGGAGCTTCATCTGG
GGTGCCTTCTTCATTGTCTCCAGGTCAATCTAGTGGAGCAACAAATGCAGCCAGCTC
TCCTGGAAGTTCACCTGCAGGTCCTCAGTCTCCCGCTCCAGCTTTGGCTGCCACTTGG
ATGTCCTTTGGATGGGTCACCTTGGTCTTTTGTGGCCTCTTCTTAGGCTAA

B.4 Contamination 16s rDNA Sequences

Yam Contamination 1 – *Hyphomicrobium facile*

GTTGTAGGATGGGCCCGCGTAGGATTAGCTAGTTGGTGAGGTAATGGCTCGCCAAG
GCGACGATCCTTAGCTGGTTTGAGAGAACGACCATTCACACTGGGACTGAGACAGG
GGCCAGACTCCTACGGGAGGCAGCAGTGGGGAATGTTGGACAATGGGCGCAAGCCT
GATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTAGGGTTGTAAAGCTCTTTTGCCG
GGGACGATAATGACGGTACCCGGAGAATAAGTCCCGGCTAACTTCGTGCCAGCAGC
CGCGGTAATACGAAGGGGACTAGCGTTGTTCGGAATCACTGGGCGTAAAGCGCACG
TAGGTGGATTTGTAAGTCAGGGGTGAAATCCCGGGGCTCAACCTCGGAACTGCCTTT
GATACTGCAAGTCTTGAGTCCGATAGAGGTGGGTGGAATTCCTAGTGTAGAGGTGA
AATTCGTAGATATTAGGAAGAACACCGGTGGCGAAGGCGGCCCACTGGATCGGTAC
TGACACTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCC
ACGCCGTAAACGATGGATGCTAGCCGTCGGATAGCTTGCTATTCGGTGGCGCAGCTA
ACGCATTAAGCATCCCGCCTGGGGAGTACGGCCGCAAGGTTAAAACTCAAAGGAAT
TGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGACGCAACGCGAAG
AACCTTACCAGCTCTTGACATTCACTGATCGCCTGGAGAGATCCGGGAGTCCCAGCA
ATGGGCAGTGGGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGG
GTTAAGTCCCGCAACGAGCGCAACCCTCGCCATTAGTTGCCATCATTAAGTTGGGCA
CTCTAGTGGGACTGCCGGTGATAAGCCGGAGGAAGGTGGGGATGANGTCAAGTCAT
CATGGCCCTTACGGGCTGGGCTACACACGTGCTACAATGGCGGTGACAATGCGCAG
CCACCTAGCAATAGGGCGCTAATCGCAAAAAGCCGTCTCAGTTCAGATTGAGGTCTG
CAACTCGACCTCATGAAGTCGGAATCGCTAGTAATCGCGCATCAGCATGGCGCGGT
GAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTTGGTCTTCC
CC

Yam Contamination 2 – *Bacillus aryabhattai*

TCCTTACGGTTACTCCCACCGACTATCGGGTGTTACAAACTCTCGTGGTGTGACGGG
CGGTGTGTACAAGGCCCGGGAACGTATTCACCGCGGCATGCTGATCCGCGATTACTA
GCGATTCCAGCTTCATGTAGGCGAGTTGCAGCCTACAATCCGAACTGAGAATGGTTT
TATGGGATTGGCTTGACCTCGCGGTCTTGCAGCCCTTTGTACCATCCATTGTAGCACG
TGTGTAGCCCAGGTCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCG
GTTTGTCACCGGCAGTCACCTTAGAGTGCCCAACTAAATGCTGGCAACTAAGATCAA
GGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAA
CCATGCACCACCTGTCACTCTGTCCCCCGAAGGGGAACGCTCTATCTCTAGAGTTGT

28

CAGAGGATGTCAAGACCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATGC
TCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACT
CCCCAGGCGGAGTGCTTAATGCGTTAGCTGCAGCACTAAAGGGCGGAAACCCTCTA
ACACTTAGCACTCATCGTTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTC
CCCACGCTTTCGCGCCTCAGCGTCAGTTACAGACCAAAAAGCCGCCTTCGCCACTGG
TGTTCCTCCACATCTCTACGCATTTCACCGCTACACGTGGAATTCCGCTTTTCTCTTCT
GCACTCAAGTTCCCCAGTTTCCAATGACCCTCCACGGTTGAGCCGTGGGCTTTCACA
TCAGACTTAAGAAACCGCCTGCGCGCGCTTTACGCCCAATAATTCCGGATAACGCTT
GCCACCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCGTGGCTTTCTGGTTAG
GTACCGTCAAGGTACGAGCAGTTACTCTCGTACTTGTTCTTCCCTAACAACAGAGTT
TTACGACCCGAAAGCCTTCATCACTCACGCGGCGTTGCTCCGTCAGACTTTCGTCCA
TTGCGGAAGATTCCCTACTGCTGCCTCCCGTAGGAGTCTGGGCCGTGTCTCAGTCCC
AGTGTGGCCGATCACCCTCTCAGGTCGGCTATGCATCGTTGCCTTGGTGAGCCGTTA
CCTCACCAACTAGCTAATGCACCGCGGGCCCATCTGTAAGTGATAGCCGAAACCATC
TTTCAATCATCTCCCATGAAGGAGAAGATCCTATCCGGTATTAGCTTCGGTTTCCCG
AAGTTAT

Yam Contamination 3 – *Bacillus ginsengihumi*

GGTGTTACAAACTCTCGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATT
CACCGCGGCATGCTGATCCGCGATTACTAGCGATTCCAGCTTCATGCAGGCGAGTTG
CAGCCTGCAATCCGAACTGAGAATGGTTTTATGGGATTGGCTTAACCTCGCGGTCTC
GCAGCCCTTTGTACCATCCATTGTAGCACGTGTGTAGCCCAGGTCATAAGGGGCATG
ATGATTTGACGTCATCCCCACCTTCCTCCGGTTTGTCACCGGCAGTCACCTTAGAGTG
CCCAACTGAATGCTGGCAACTAAGGTCAAGGGTTGCGCTCGTTGCGGGACTTAACCC
AACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCACTCTGTCCCCCG
AAGGGGAAGGCCCTATCTCTAGGGAGGTCAGAGGATGTCAAGACCTGGTAAGGTTC
TTCGCGTTGCTTCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAAT
TCCTTTGAGTTTCAACCTTGCGGTCGTACTCCCCAGGCGGAGTGCTTAATGCGTTAGC
TGCAGCACTAAAGGGCGGAAACCCTCTAACACTTAGCACTCATCGTTTACGGCGTGG
ACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGCGCCTCAGCGTCAGTTA
CAGACCAGAGAGTCGCCTTCGCCACTGGTGTTCCTCCACATCTCTACGCATTTCACC
GCTACACGTGGAATTCCACTCTCCTCTTCTGCACTCAAGTCTTCCAGTTTCCAATGAC
CCTCCACGGTTGAGCCGTGGGCTTTCACATCAGACTTAAAAGACCGCCTGCGCGCGC
TTTACGCCCAATAATTCCGGACAACGCTTGCCACCTACGTATTACCGCGGCTGCTGG
CACGTAGTTAGCCGTGGCTTTCTGGTTAGGTACCGTCAAGGTACCGCCCTATTCGAA
CGATACTTGTTCTTCCCTAACAACAGAGTTTTACGATCCGAAGACCTTCTTCACTCAC
GCGGCGTTGCTCCGTCAGACTTTCGTCCATTGCGGAAGATTCCCTACTGCTGCCTCCC
GTAGGAGTTTGGGCCGTGTCTCAGTCCCAATGTGGCCGATCACCCTCTCAGGTCGGC
TACGCATCGTTGCCTTGGTGAGCCGTTACCTCACCAACTAGCTAATGCGCCGCGGGT
CCATCTGTAAGTGATAGCCGAAGCCATCTTTCAATCTCCCTCCATGCGGAGAGAAAA
GTTATCCAGTATTAGCCCCGGTTTCCCGGAGTTATCCTAGTCTTACAGGCAGGTTACC
CACGTGTTACTCACCCGTCCGCCGCTAACTGATCAAAAGC

Cassava Pink Contamination - *Methylobacterium*

GTGGTCGCCTGCCTCCTTGCGGTTAGCACAGCGCCTTCGGGTAAAACCAACTCCCAT
GGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCGGCATGCTGAT
CCGCGATTACTAGCGATTCCAACTTCATGCACTCGAGTTGCAGAGTGCAATCCGAAC
TGAGATGGCTTTTGGAGATTAGCTCGACCTCGCGGTCTCGCTGCCCACTGTCACCAC
CATTGTAGCACGTGTGTAGCCCAGCCCGTAAGGGCCATGAGGACTTGACGTCATCCC
CACCTTCCTCTCGGCTTATCACCGGCAGTCCCCTTAGAGTGCCCAACTGAATGATGG
CAACTAAGGGCGAGGGTTGCGCTCGTTGCGGGACTTAAACCAACATCTCACGACAC
GAGCTGACGACAGCCATGCAGCACCTGTCACCGGTCCAGCCGAACTGAAGGCCTAA
ATCTCTCTAGGCCGCGACCGGGATGTCAAGGGCTGGTAAGGTTCTGCGCGTTGCTTC
GAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTTA
ATCTTGCGACCGTACTCCCCAGGCGGGAAGCTTAATGCGTTAGCTGCGCCACCGACA
AGTAAACTTGCCGACGGCTAGCTTCCATCGTTTACAGCGTGGACTACCAGGGTATCT
AATCCTGTTTGCTCCCCACGCTTTCGCACCTCAGCGTCAGTACCGGACCAGTGAGCC
GCCTTCGCCACTGGTGTTCCTCCGAATATCTACGAATTTCACCTCTACACTCGGAATT
CCACTCACCTCTTCCGGACTCGAGATTGCCAGTATTAAAGGCAGTTCCAGGGTTGAG
CCCTGGGATTTCACCCCTAACTTAACAATCCGCCTACGTGCGCTTTACGCCCAGTAA
ATCCGAACAACGCTAGCCCCCTTCGTATTACCGCGGCTGCTGGCACGAAGTTAGCCG
GGGCTTCTTCTACGGTTACCGTCATTATCTTCACCGTTGAAAAAACCTTACAACCCTA
GGGCCTTCATCACTCACGCGGCATGGCTGGATCAGGCTTGCGCCCATTGTCCAATAT
TCCCCACTGCTGCCTCCCGTAGGAGTCTGGGCCGTGTCTCAGTCCCAGTGTGGCTGA
TCATCCTCTCAGACCAGCTATGGATCGTCGCCTTGGTAGGCCATTACCCCACCAACT
AGCTAATCCAACGCGGGCTCATCCATCACCGATAAATCTTTCTCCCGAAGGACGTAT
ACGGTATTAGCTCCAGTTTCCCGGAGTTGTTCCGTAGTGATGGGTAGATTCCCACGC
GTTACTCACCCGTCTGCCGCTCCCCTTGCGGGGCGCTCGACTGCA

Sparge Beige Contamination - *Mesorhizobium*

TGCAGTCGAGCGCCCCGCAAGGGGAGCGGCAGACGGGTGAGTAACGCGTGGGAAT
CTACCCATCACTACGGAACAACTCCGGGAAACTGGAGCTAATACCGTATACGTCCTT
CGGGAGAAAGATTTATCGGTGATGGATGAGCCCGCGTTGGATTAGCTAGTTGGTGG
GGTAATGGCCTACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCAC
ACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGG
ACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGT
TGTAAAGCTCTTTCAACGGTGAAGATAATGACGGTAACCGTAGAAGAAGCCCCGGC
TAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGATTTAC
TGGGCGTAAAGCGCACGTAGGCGGATTGTTAAGTTAGGGGTGAAATCCCAGGGCTC
AACCCTGGAACTGCCTTTAATACTGGCAATCTCGAGTCCGGAAGAGGTGAGTGGAA
TTCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAGGAACACCAGTGGCGAAGGC
GGCTCACTGGTCCGGTACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAGGAT
TAGATACCCTGGTAGTCCACGCTGTAAACGATGGAAGCTAGCCGTCGGCAAGTTTAC
TTGTCGGTGGCGCAGCTAACGCATTAAGCTTCCCGCCTGGGGAGTACGGTCGCAAGA
TTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAA
TTCGAAGCAACGCGCAGAACCTTACCAGCCCTTGACATCCCGGTCGCGGCCTAGAG
AGATTTAGGCCTTCAGTTCGGCTGGACCGGTGACAGGTGCTGCATGGCTGTCGTCAG
CTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGCCCTTAGT
TGCCATCATTCAGTTGGGCACTCTAAGGGGACTGCCGGTGATAAGCCGAGAGGAAG

30
GTGGGGATGACGTCAAGTCCTCATGGCCCTTACGGGCTGGGCTACACACGTGCTACA
ATGGTGGTGACAGTGGGCAGCGAGACCGCGAGGTCGAGCTAATCTCCAAAAGCCAT
CTCAGTTCGGATTGCACTCTGCAACTCGAGTGCATGAAGTTGGAATCGCTAGTAATC
GCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCAC
ACCATGGGAGTTGGTTTTACCCGAAGGCGCTGTGCTAACCGCAAGGAGGCAGGCGA
CC

Sparge Yellow Contamination - *Mesorhizobium*

GTGGTCGCCTGCCTCCTTGCGGTTAGCACAGCGCCTTCGGGTAAAACCAACTCCCAT
GGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCGGCATGCTGAT
CCGCGATTACTAGCGATTCCAACTTCATGCACTCGAGTTGCAGAGTGCAATCCGAAC
TGAGATGGCTTTTGGAGATTAGCTCGACCTCGCGGTCTCGCTGCCCACTGTCACCAC
CATTGTAGCACGTGTGTAGCCCAGCCCGTAAGGGCCATGAGGACTTGACGTCATCCC
CACCTTCCTCTCGGCTTATCACCGGCAGTCCCCTTAGAGTGCCCAACTGAATGATGG
CAACTAAGGGCGAGGGTTGCGCTCGTTGCGGGACTTAAACCAACATCTCACGACAC
GAGCTGACGACAGCCATGCAGCACCTGTCACCGGTCCAGCCGAACTGAAGGCCTAA
ATCTCTCTAGGCCGCGACCGGGATGTCAAGGGCTGGTAAGGTTCTGCGCGTTGCTTC
GAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTTA
ATCTTGCGACCGTACTCCCCAGGCGGGAAGCTTAATGCGTTAGCTGCGCCACCGACA
AGTAAACTTGCCGACGGCTAGCTTCCATCGTTTACAGCGTGGACTACCAGGGTATCT
AATCCTGTTTGCTCCCCACGCTTTCGCACCTCAGCGTCAGTACCGGACCAGTGAGCC
GCCTTCGCCACTGGTGTTCCTCCGAATATCTACGAATTTCACCTCTACACTCGGAATT
CCACTCACCTCTTCCGGACTCGAGATTGCCAGTATTAAAGGCAGTTCCAGGGTTGAG
CCCTGGGATTTCACCCCTAACTTAACAATCCGCCTACGTGCGCTTTACGCCCAGTAA
ATCCGAACAACGCTAGCCCCCTTCGTATTACCGCGGCTGCTGGCACGAAGTTAGCCG
GGGCTTCTTCTACGGTTACCGTCATTATCTTCACCGTTGAAAAAACCTTACAACCCTA
GGGCCTTCATCACTCACGCGGCATGGCTGGATCAGGCTTGCGCCCATTGTCCAATAT
TCCCCACTGCTGCCTCCCGTAGGAGTCTGGGCCGTGTCTCAGTCCCAGTGTGGCTGA
TCATCCTCTCAGACCAGCTATGGATCGTCGCCTTGGTAGGCCATTACCCCACCAACT
AGCTAATCCAACGCGGGCTCATCCATCACCGATAAATCTTTCTCCCGAAGGACGTAT
ACGGTATTAGCTCCAGTTTCCCGGAGTTGTTCCGTAGTGATGGGTAGATTCCCACGC
GTTACTCACCCGTCTGCCGCTCCCCTTGCGGGGCGCTCGACTGCA

B.5 Signal Peptide and mCherry Sequences

B.5.1 Amino Acid Sequences

LePLA Signal Peptide

MAVFSRNVFVMAILFFSLLSFTEA

XGlu-hy Signal Peptide

MGIIKGVLFSIVLINLSLVVFCG

mCherry

MAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSP
QFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYK
VKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVK
TTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDE

B.5.2 DNA Sequences

LePLA Signal Peptide

ATGGCTGTTTTTTCAAGAAATGTGTTTGTAATGGCAATTCTGTTTTTCAGTCTATTAA
GCTTCACAGAAGCA

XGlu-hy Signal Peptide

ATGGGTATCATAAAAGGAGTTTTATTTAGTATTGTTTTGATTAATTTGTCACTTGTTG
TATTTTGTGGG

mCherry

ATGGCCATCATCAAGGAGTTCATGCGGTTCAAGGTGCACATGGAAGGCAGCGTGAA
CGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAAGGCAGGCCCTACGAGGGCACC
CAGACCGCCAAGCTGAAGGTGACCAAGGGCGGACCTCTGCCCTTCGCCTGGGACAT
CCTGAGCCCCCAGTTCATGTACGGCAGCAAGGCCTACGTGAAGCACCCCGCCGACA
TCCCCGACTACCTGAAGCTGTCCTTCCCCGAGGGATTCAAATGGGAGCGGGTGATGA
ACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACAGCAGCCTGCAGGACGGC
GAGTTCATCTACAAGGTGAAACTGCGGGGCACCAACTTCCCCAGCGACGGCCCCGT
GATGCAGAAAAAGACCATGGGCTGGGAGGCCAGCAGCGAGCGGATGTACCCCGAG
GATGGCGCCCTGAAGGGCGAGATCAAGCAGCGGCTGAAGCTGAAGGATGGCGGCC
ACTACGACGCCGAGGTGAAAACCACCTACAAGGCCAAGAAACCCGTGCAGCTGCCT
GGCGCCTACAACGTGAACATCAAGCTGGACATCACCAGCCACAACGAGGACTACAC
CATCGTGGAGCAGTACGAGCGGGCCGAGGGCAGACACAGCACCGGCGGCATGGAC
GAG

mCherry - Codon Optimized for Tomato

ATGGCCATCATCAAGGAGTTCATGCGGTTCAAGGTGCACATGGAAGGCAGCGTGAA
CGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAAGGCAGGCCCTACGAGGGCACC

CAGACCGCCAAGCTGAAGGTGACCAAGGGCGGACCTCTGCCCTTCGCCTGGGACAT
CCTGAGCCCCCAGTTCATGTACGGCAGCAAGGCCTACGTGAAGCACCCCGCCGACA
TCCCCGACTACCTGAAGCTGTCCTTCCCCGAGGGATTCAAATGGGAGCGGGTGATGA
ACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACAGCAGCCTGCAGGACGGC
GAGTTCATCTACAAGGTGAAACTGCGGGGCACCAACTTCCCCAGCGACGGCCCCGT
GATGCAGAAAAAGACCATGGGCTGGGAGGCCAGCAGCGAGCGGATGTACCCCGAG
GATGGCGCCCTGAAGGGCGAGATCAAGCAGCGGCTGAAGCTGAAGGATGGCGGCC
ACTACGACGCCGAGGTGAAAACCACCTACAAGGCCAAGAAACCCGTGCAGCTGCCC
AAGCTGGACATCACCAGCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCG
GGCCGAGGGCAGACACAGCACCGGCGGCATGGACGAGTAA

B.6 Alcohol Inducible Vector Sequences

AlcA Promoter

ATGCATGCGGAACCGCACGAGGATGCATGCGGAACCGCACGAGGATGCATGCGGA
ACCGCACGAGGATGCATGCGGAACCGCACGAGGATGCATGCGGAACCGCACGAGG
TCGATAGTTGTGATAGTTCCCACTTGTCCGTCCGCATCGGCATCCGCAGCTCGGGAT
AGTTCCGACCTAGGATTGGATGCATGCGGAACCGCACGAGGGCGGGGCGGAAATTG
ACACACCACTCCTCTCCACGCACCGTTCAAGAGGTACGCGTATAGAGCCGTATAGA
GCAGAGACGGAGCACTTTCTGGTACTGTCCGCACGGGATGTCCGCACGGAGAGCCA
CAAACGAGCGGGGCCCCGTACGTGCTCTCCTACCCCAGGATCGCATCCCCGCATAGC
TGAACATCTATATAAGGAAGTTCATTTCATTTGGAGAGGAC

AlcR Transcription Factor

ATGGCAGATACGCGCCGACGCCAGAATCATAGCTGCGATCCCTGTCGCAAGGGCAA
GCGACGCTGTGATGCCCCGGTAGGTTGCCGATATCGGCTCCCCAGCGTGTGCACTGA
CAGTCGCTGAGATGTAACACAGGAAAATAGAAACGAGGCCAATGAAAACGGCTGG
GTTTCGTGTTCAAATTGCAAGCGTTGGAACAAGGATTGTACCTTCAATTGGCTCTCA
TCCCAACGCTCCAAGGCAAAAGGGGCTGCACCTAGAGCGAGAACAAAGAAAGCCA
GGACCGCAACAACCACCAGTGAACCATCAACTTCAGCTGCAACAATCCCTACACCG
GAAAGTGACAATCACGATGCGCCTCCAGTCATAAACTCTCACGACGCGCTCCCGAG
CTGGACTCAGGGGCTACTCTCCCACCCCGGCGACCTTTTCGATTTCAGCCACTCTGCT
ATTCCCGCAAATGCAGAAGATGCGGCCAACGTGCAGTCAGACGCACCTTTTCCGTG
GGATCTAGCCATCCCCGGTGATTTCAGCATGGGCCAACAGCTCGAGAAACCTCTCAG
TCCGCTCAGTTTTCAAGCAGTCCTTCTTCCGCCCCATAGCCCGAACACGGATGACCT
CATTCGCGAGCTGGAAGAGCAGACTACGGATCCGGACTCGGTTACCGATACTAATA
GTGTACAACAGGTCGCTCAAGATGGATCGCTATGGTCTGATCGGCAGTCGCCGCTAC
TGCCTGAGAACAGTCTGTGCATGGCCTCAGACAGCACAGCACGGCGATATGCCCGT
TCCACAATGACGAAGAATCTGATGCGAATCTACCACGATAGTATGGAGAATGCACT
GTCCTGCTGGCTGACAGAGCACAATTGTCCATACTCCGACCAGATCAGCTACCTGCC
GCCCAAGCAGCGGGCGGAATGGGGCCCGAACTGGTCAAACAGGATGTGCATCCGGG

33

TGTGCCGGCTAGATCGCGTATCTACCTCATTACGCGGGCGCGCCCTGAGTGCGGAAG
AGGACAAAGCCGCAGCCCGAGCCCTGCATCTGGCGATCGTAGCTTTTGCGTCGCAAT
GGACGCAGCATGCGCAGAGGGGGGCTGGGCTAAATGTTCCTGCAGACATAGCCGCC
GATGAGAGGTCCATCCGGAGGAACGCCTGGAATGAAGCACGCCATGCCTTGCAGCA
CACGACAGGGATTCCATCATTCCGGGTTATATTTGCGAATATCATCTTTTCTCTCACG
CAGAGTGTGCTGGATGATGATGAGCAGCACGGTATGGGTGCACGTCTAGACAAGCT
ACTCGAAAATGACGGTGCGCCCGTGTTCCTGGAAACCGCGAACCGTCAGCTTTATAC
ATTCCGACATAAGTTTGCACGAATGCAACGCCGCGGTAAGGCTTTCAACAGGCTCCC
GGGAGGATCTGTCGCATCGACATTCGCCGGTATTTTCGAGACACCGACGCCGTCGTC
TGAAAGCCCACAGCTTGACCCGGTTGTGGCCAGTGAGGAGCATCGCAGTACATTAA
GCCTTATGTTCTGGCTAGGGATCATGTTCGATACACTAAGCGCTGCAATGTACCAGC
GACCACTCGTGGTGTCAGATGAGGATAGCCAGATATCATCGGCATCTCCACCAAGG
CGCGGCGCTGAAACGCCGATCAACCTAGACTGCTGGGAGCCCCCGAGACAGGTCCC
GAGCAATCAAGAAAAGAGCGACGTATGGGGCGACCTCTTCCTCCGCACCTCGGACT
CTCTCCCAGATCACGAATCCCACACACAAATCTCTCAGCCAGCGGCTCGATGGCCCT
GCACCTACGAACAGGCCGCCGCCGCTCTCTCCTCTGCAACGCCCGTCAAAGTCCTCC
TCTACCGCCGCGTCACGCAGCTCCAAACCCTCCTCTATCGCGGCGCCAGCCCTGCCC
GCCTTGAAGCGGCCATCCAGAGAACGCTCTACGTTTATAATCACTGGACAGCGAAGT
ACCAACCATTTATGCAGGACTGCGTTGCTAACCACGAGCTCCTCCCTTCGCGCATCC
AGTCTTGGTACGTCATTCTAGACGGTCACTGGCATCTAGCCGCGATGTTGCTAGCGG
ACGTTTTGGAGAGCATCGACCGCGATTCGTACTCTGATATCAACCACATCGACCTTG
TAACAAAGCTAAGGCTCGATAATGCACTAGCAGTTAGTGCCCTTGCGCGCTCTTCAC
TCCGAGGCCAGGAGCTGGACCCGGGCAAAGCATCTCCGATGTATCGCCATTTCCATG
ATTCTCTGACCGAGGTGGCATTCCTGGTAGAACCGTGGACCGTCGTTCTTATTCACTC
GTTTGCCAAAGCTGCGTATATCTTGCTGGACTGTTTAGATCTGGACGGCCAAGGAAA
TGCACTAGCGGGGTACCTGCAGCTGCGGCAAAATTGCAACTACTGCATTCGGGCGCT
GCAATTTCTGGGCAGGAAGTCGGATATGGCGGCGCTGGTTGCGAAGGATTTAGAGA
GAGGTTTGAATGGGAAAGTTGACAGCTTTTTGTAG

<u>TEV 5’ UTR</u>

AATTCTCAACACAACATATACAAAACAAACGAATCTCAAGCAATCAAGCATTCTACT
TCTATTGCAGCAATTTAAATCATTTCTTTTAAAGCAAAAGCAATTTTCTGAAAATTTT
CACCATTTACGAACGATAGCC

Appendix C

Methods Details

ZYM-5052 Autoinduction Media (adapted from Studier 2005)

1.045x ZY	957 mL
50x M	20 mL
50x 5052	20 mL
M1 MgSO4	2 mL
1000x Metals	1 mL

Stock Solutions:

ZY 1.045x
- 10g Tryptone
- 5g yeast extract
- 957mL water

M 50x
- 88.73g Na2HPO4 (final conc.=1.25M, MW=141.96 g/M)
- 85.06g KH2PO4 (final conc.=1.25M, MW=136.09 g/M)
- 66.86g NH4Cl (final conc.=2.5M, MW=53.49 g/M)
- 17.76g Na2SO4 (final conc=0.25M, MW=142.04 g/M)
- water to 500mL
- filter sterilize

5052 50x
- 125g glycerol (25% w/v)
- 12.5g glucose (2.5% w/v)
- 50g alpha-lactose monohydrate (10% w/v)
- filter sterilize

Metals 1000x
- 4.06g FeCl3 (final conc.=50mM, MW=162.20 g/M)
- 247.5mL of water
- 2.5mL of concentrated HCL (12.5N)

50mL ZnSO4 from 1M autoclaved stock
- 10mL CaCl2 from 1M autoclaved stock
- 5mL MnCl2 from 1M autoclaved stock
- 1mL CoCl2 from 1M autoclaved stock
- 1mL CuCl2 from 1M autoclaved stock

1mL NiCl2 from 1M autoclaved stock
1mL Na2MoO4 from 1M autoclaved stock
1mL Na2SeO3 from 1M autoclaved stock
1mL H3Bo from 1M autoclaved stock

add 208mL of sterile water

Lysis Buffer

Tris	5 mM
NaH2PO4	50 mM
NaOH	100 m

BIBLIOGRAPHY

1. Poon S, Heath RL, Clarke AE. 2012. A chimeric arabinogalactan protein promotes somatic embryogenesis in cotton cell culture. Plant Physiol 160:684–95.
2. Egertsdotter U, von Arnold S. 1995. Importance of Arabinogalactan proteins for the development of somatic embryos of Norway spruce (Picea abies). Physiol Plant 93:334–345.
3. Pena-Francesch A, Florez S, Jung H, Sebastian A, Albert I, Curtis W, Demirel MC. 2014. Materials fabrication from native and recombinant thermoplastic squid proteins. Adv Funct Mater 24.
4. Wu W-Y, Mee C, Califano F, Banki R, Wood DW. 2006. Recombinant protein purification by self-cleaving aggregation tag. Nat Protoc 1:2257–62.
5. Green MR, Sambrook J. 2012. Molecular Cloning: A Laboratory Manual, 4th ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
6. Lee J, Lee HJ, Shin MK, Ryu WS. 2004. Versatile PCR-mediated insertion or deletion mutagenesis. Biotechniques 36:398–400.
7. Studier FW. 2005. Protein production by auto-induction in high-density shaking cultures. Protein Expr Purif 41:207–234.
8. Lai TS. 2016. Ehancement of Somatic Embryogenesis in Theobroma cacao Using Phytocyanin-like-1 arabinogalactin (PLA) Protein Domains. The Pennsylvania State University.
9. Narina SS, Buyyarapu R, Kottapalli KR, Sartie AM, Ali MI, Robert A, Hodeba MJD,

Sayre BL, Scheffler BE. 2011. Generation and analysis of expressed sequence tags (ESTs) for marker development in yam (Dioscorea alata L.). BMC Genomics 12:100.

10. Wu W-Y, Mee C, Califano F, Banki R, Wood DW. 2006. Recombinant protein purification by self-cleaving aggregation tag. Nat Protoc 1:2257–62.

11. van der Linde K. 1999. Improved bacteriological surveillance of haemodialysis fluids: a comparison between Tryptic soy agar and Reasoner's 2A media. Nephrol Dial Transplant 14:2433–2437.

12. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 1991. 16S Ribosomal DNA Amplification for Phylogenetic Study. J Bacteriol 173:697–703.

13. Lidstrom ME, Chistoserdova L. 2002. Plants in the pink: Cytokinin production by Methylobacterium. J Bacteriol 184:1818.

14. Buckner JS, Freeman TP, Ruud RL, Chu CC, Henneberry TJ. 2002. Characterization and functions of the whitefly egg pedicel. Arch Insect Biochem Physiol 49:22–33.

15. Zhang X, Mason H. 2006. Bean yellow dwarf virus replicons for high-level transgene expression in transgenic plants and cell cultures. Biotechnol Bioeng 93:271–279.

ACADEMIC VITA

Academic Vita of Eva Mei Shouse

Email: evameishouse@gmail.com

Education: Schreyer Honors College, The Pennsylvania State University

Graduation: August 2018

Degree: Bachelor of Science in Microbiology

Honors: Chemical Engineering

Thesis Title: Heterologous expression of phytocyanin-like arabinogalactan protein from *Diascorea rotundata* as a potential treatment to promote

Thesis Supervisor: Dr. Wayne Curtis

Work Experience

July – December 2012; August – December 2015; May 2017 - present

Undergraduate Researcher

The Pennsylvania State University

Dr. Wayne Curtis

June – December 2013

Lab Assistant

The Pennsylvania State University

Dr. Chobi Debroy

July – December 2013

Undergraduate Researcher

The Pennsylvania State University

Dr. Chun Liu

Grants Received: Braddock Scholarship, Academic Excellence Scholarship

Community Service Involvement: Volunteer at Discovery Space (April – June 2018)

Language Proficiency: English, Chinese