CHARACTERIZATION OF DNA METHYLTRANSFERASE 1-ASSOCIATED PROTEIN FROM *PHYTOPHTHORA SOJAE*

Alexander E. Howard

A Thesis
Submitted to the Graduate College of Bowling Green State University in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE
August 2017

Committee:
Vipaporn Phuntumart, Advisor
Raymond Larsen
Paul Morris
ABSTRACT

Vipaporn Phuntumart, Advisor

5-Methylcytosine methylation of DNA is an epigenetic mechanism for controlling and regulating expression of genes. A genome analysis of oomycetes showed no definitive DNA Methyltransferase (DNMTs); indicating that DNA methylation might not factor in gene regulation. Yet, DNA methyltransferase associated protein-1 (DMAP-1) was found in the model organism *Phytophthora sojae*. DMAP-1 has been proven to work in conjunction with DNMT in humans. The goal of this study is to understand the role of DMAP-1 in DNA methylation of *P. sojae*, in the absence of known DMNTs. The DMAP-1 gene was isolated and cloned into both *Escherichia coli* (Dam-/Dcm-) and *Saccharomyces cerevisiae* (Ysc1048). DMAP1 was induced and DNA underwent a restriction digestion assay with McrBC, HpaII, MspI, and MspJI. The restriction enzyme analysis failed to yield conclusive results. This suggested that DMAP1 does not methylate DNA or that the method was not sensitive to detect low levels of methylation. Future work must be done to gain an understanding of the gene. First, current methods could be modified to account for codon bias and post translation modification. Second, a more sensitive method could be performed to clarify if DMAP-1 from *P. sojae* has the ability to methylate; such as over expression or knockout of DMAP1 in *P. sojae*. It is also possible that DMAP-1 is not capable of causing 5’ cytosine methylation independently of DMNTs.
ACKNOWLEDGMENTS

To start I would also like to personally thank my advisor Dr. Vipa Phuntumart. Without her I would not have pursued my graduate education and gain important knowledge and skills while working on this project. I greatly appreciated both the support and mentoring that she provided throughout my graduate career. I would like to thank my committee members Dr. Paul Morris and Dr. Raymond Larsen for their aid with equipment and advise whenever the need arose. I would also like to acknowledge the help of all my fellow colleagues in the Phuntumart lab; along with all the undergraduates who helped me over the course of this project. Finally, I would like to thank the USDA for their financial support, and Bowling Green State University for the opportunity to be have a wonderful graduate experience.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER 1: INTRODUCTION</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oomycetes</td>
<td>1</td>
</tr>
<tr>
<td>DNA Methylation &amp; Epigenetics</td>
<td>4</td>
</tr>
<tr>
<td>Detection of 5-Methylcytosine</td>
<td>5</td>
</tr>
<tr>
<td>Bisulfite Sequencing</td>
<td>5</td>
</tr>
<tr>
<td>Restriction Enzyme Analysis</td>
<td>7</td>
</tr>
<tr>
<td>DNA Methyltransferase 1-Associated Protein (DMAP1)</td>
<td>8</td>
</tr>
<tr>
<td>Aims</td>
<td>8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER 2: MATERIALS AND METHODOLOGIES</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aim I - Bioinformatics Analysis</td>
<td>9</td>
</tr>
<tr>
<td>Phylogenetic Tree of DMAP1</td>
<td>9</td>
</tr>
<tr>
<td>Expression Analysis of DMAP1</td>
<td>9</td>
</tr>
<tr>
<td>Aim II – Cloning of DMAP1</td>
<td>9</td>
</tr>
<tr>
<td>Cloning the Synthetic Version of DMAP1 (sDMAP1)</td>
<td>9</td>
</tr>
<tr>
<td>Cloning the Native Version of DMAP1 from <em>P. sojae</em> (nDMAP1)</td>
<td>15</td>
</tr>
<tr>
<td>Aim III – Functional Analysis of DMAP1</td>
<td>18</td>
</tr>
<tr>
<td>Toxicity Assay in Prokaryotes</td>
<td>18</td>
</tr>
<tr>
<td>Toxicity Assay in Eukaryotes</td>
<td>18</td>
</tr>
<tr>
<td>Restriction Digestion Enzymes</td>
<td>19</td>
</tr>
<tr>
<td>Functional Analysis in Prokaryotes</td>
<td>20</td>
</tr>
<tr>
<td>Functional Analysis in Eukaryotes</td>
<td>21</td>
</tr>
</tbody>
</table>
Positive Controls ................................................................. 21

CHAPTER 3: RESULTS .......................................................... 23

Aim I – Bioinformatics Analysis .............................................. 23

Phylogenetic Analysis of DMAP1 .......................................... 23

In silico Expression Analysis of DMAP1 ............................... 26

Aim II – Cloning of DMAP1 .................................................. 28

Cloning the Synthesized Version of DMAP1 (sDMAP1) ......... 28

Cloning the Native Version of DMAP1 from P. sojae (nDMAP1) .... 32

Aim III – Functional Analysis of DMAP1 ................................ 37

Toxicity Assay ................................................................. 37

Functional Analysis with Restriction Enzymes .................... 39

CHAPTER 4: DISCUSSION AND CONCLUSION .......................... 47

Potential Shortcomings in Current Methodology .................. 47

Weakness with Detection Method ........................................ 49

Alternative Methods ......................................................... 50

Conclusions ....................................................................... 51

REFERENCES ................................................................. 52

APPENDIX A: ALIGNMENT OF DMAP1 ................................. 60

APPENDIX B: YEAST DNA EXTRACTION PROTOCOL ............. 61
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>Phytophthora sojae</em> life cycle</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>DNA methylation and gene expression</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>Bisulfite sequencing</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>pDONR221 vector map</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>Cloning method for DMAP1</td>
<td>12</td>
</tr>
<tr>
<td>6</td>
<td>pBAD-DEST 49 vector map</td>
<td>14</td>
</tr>
<tr>
<td>7</td>
<td>pENTR/SD/D-TOPO vector map</td>
<td>16</td>
</tr>
<tr>
<td>8</td>
<td>pYES-DEST 52 vector map</td>
<td>17</td>
</tr>
<tr>
<td>9</td>
<td>DMAP1 domain</td>
<td>23</td>
</tr>
<tr>
<td>10</td>
<td>Phylogenetic tree of DMAP1</td>
<td>25</td>
</tr>
<tr>
<td>11</td>
<td>Expression data of DMAP1 in <em>Phytophthora ramorum</em></td>
<td>26</td>
</tr>
<tr>
<td>12</td>
<td>Expression data of DMAP1 in <em>P. sojae</em></td>
<td>27</td>
</tr>
<tr>
<td>13</td>
<td>RNAseq data of <em>Phytophthora capsici</em> taken from RNArocket</td>
<td>28</td>
</tr>
<tr>
<td>14</td>
<td>Attb-PCR product</td>
<td>29</td>
</tr>
<tr>
<td>15</td>
<td>Colony PCR of sDMAP1 from TOP10 transformants</td>
<td>30</td>
</tr>
<tr>
<td>16</td>
<td>Colony PCR of sDMAP1 in <em>E. coli</em> Dam⁺/Dcm⁻</td>
<td>31</td>
</tr>
<tr>
<td>17</td>
<td>Optimization of PCR for nDMAP1 amplification</td>
<td>33</td>
</tr>
<tr>
<td>18</td>
<td>Colony PCR of nDMAP1 in the TOP10 <em>E. coli</em> cells</td>
<td>34</td>
</tr>
<tr>
<td>19</td>
<td>Colony PCR of nDMAP1 in Dam⁺/Dcm⁻</td>
<td>35</td>
</tr>
<tr>
<td>20</td>
<td>Colony PCR of nDMAP1 in <em>S. cerevisiae</em></td>
<td>36</td>
</tr>
<tr>
<td>21</td>
<td>Toxicity assay…</td>
<td>38</td>
</tr>
<tr>
<td>Page</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Restriction analysis of <em>N. benthamiana</em></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Restriction analysis of positive control</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Digestion assay for sDMAP1 on total DNA from Dam(^{+})/Dcm(^{-})</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Digestion assay for nDMAP1 on total DNA from Dam(^{+})/Dcm(^{-})</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Digestion assay for nDMAP1 on plasmid DNA from Dam(^{+})/Dcm(^{-})</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Digestion assay for nDMAP1 on genomic DNA from <em>S. cerevisiae</em></td>
<td></td>
</tr>
</tbody>
</table>
### LIST OF TABLES

1. Primers for cloning DMAP1 ................................................................. 10
2. PCR protocols for DMAP1 .................................................................. 10
3. Restriction digestion enzymes .......................................................... 19
4. Restriction digestion enzymes .......................................................... 41
CHAPTER 1: INTRODUCTION

Oomycetes

Oomycetes are pathogenic eukaryotic organisms capable of infecting and killing both plant and animal species (Fry and Grunwald, 2010). In humans *Pythium insidiosum* is responsible for the loss of life in tropical and subtropical areas in southwest Asia (Gaastra et al, 2010). While *Phytophthora infestans* has caused the annihilation of entire crop yields, resulting in the Irish potato famine in the 19th century (Hwang et al, 2014). Oomycetes were once broadly considered to be fungi. Recent molecular methods proved that previous classifications were incorrect. They are now identified as being part of the Stramenopila kingdom. This makes them more closely related to diatoms, green plants, and other algae (Fry and Grunwald, 2010). Like fungi oomycetes are capable of producing mycelia during its vegetative state, which lead to misclassification (Stewart, 2011). Genome analysis of oomycetes showed that they have acquired many genes via horizontal gene transfers from fungi and other species (Morris et al., 2009 and Richards & Talbot, 2007). Oomycetes and fungi also have key differences. Oomycete cell walls are composed of cellulose instead of chitin in fungi (Rossman and Palm, 2016). Oomycetes genomes are also typically larger by comparison.

*Phytophthora sojae* is a soil-borne plant pathogen belonging to the genus *Phytophthora*, meaning “Plant Destroyer” (Stewart, 2011; Tyler, 2007). The genome of *P. sojae* spans a total of 132 MB in length and is one of the first oomycete genomes to be published (Tyler et al, 2006). Its primary host is *Glycine max*. It is also reported to be capable of infecting flowering plants from the genus *Lupinus* in the legume family (Beligala, 2016; Tyler, 2007). *P. sojae* is known for causing damping off of seedlings and root rot in mature soybeans plants (Stewart, 2011; Tyler, 2007). It is responsible for the loss of crop yields ranging from 4-100%. It is estimated to cause
1-2 billion dollars in damages every year globally (Sugimoto et al., 2012; Tyler, 2007). *P. sojae* also plays a significant role in recycling matter by feeding off decomposing organisms at the end of its life cycle (Fry and Grunwald, 2010).

Oomycetes can undergo both sexual and asexual reproduction. In sexual reproduction both female and male cells are produced by differentiating into oogonium and antheridium respectively. Once the oogonium is fertilized, it develops a thickened cell wall and becomes an oospore (Figure 1). Oospores can survive in the soil or plant debris due to the protection from their thick cell wall. Oospores can remain dormant in the soil for years until optimal conditions are met for germination (Fry and Grunwald, 2010, Stewart, 2011). In optimal conditions the oospores produce mycelium, which then form sporangia (Stewart, 2011). *P. sojae* sporangia resemble a lemon or egg shape and are found at the end of a hyphae (Stewart, 2011). Sporangium contain multiple zoospores. This is the mobile phase of the organism which helped set it apart from true fungi (Rossman and Palm, 2016). Zoospores have two flagella and are capable of swimming to its host or chemoattractants. In *P. sojae*, they respond to root exudates known as isoflavonones, such as daidzein and genistein (Morris and Ward, 1992). Once they reach the host they shed their flagella. They then encyst and penetrate the epidermis allowing them to infect and grow inside the host. It is capable of being both a biotroph and a necrotroph, feeding off their host even after they are dead. Once they run out of nutrients oospores are formed (Qutob et al, 2002).
Figure 1: *Phytophthora sojae* life cycle. It summarizes the life cycle of the oomycete plant pathogen *P. sojae* infecting *G. max* (soybean). **A)** sporangia are formed at the tip of the mycelium growing from either oospores or chlamydospores. Sporangium contains multiple zoospores. **B)** after zoospores are released they swim towards potential hosts. **C)** infection site of soybeans, found near the root tip. **D)** once a host is found, the zoospore encysts, sheds its flagellum, and a germ tube is then formed, allowing it to invade the epidermis. **E)** formation of sexual oospore, from the fusion of the antheridium and oogonium. **F)** formation of asexual chlamydospore. **G)** infection continues, as host dies. **H)** oospores and chlamydospores survive for years in soil debris. **I)** optimal conditions lead to development of hyphae and the formation of sporangium.
DNA Methylation & Epigenetics

Epigenetics is the study of heritable changes in the expression of genes that do not change the genomic sequence (Jin et al, 2011). DNA methylation is one such example of epigenetics. Methylation occurs when there is a methyl group (-CH3) found on the fifth carbon on the purine base, cytosine (Lim and Maher, 2010). Increased methylation within the CpG islands can result in a decrease of gene expression. CpG islands are regions of GC rich sequences typically located before the start codon of a gene in the promoter region. This occurs by preventing transcription factors from binding to the promoter stopping transcription of the gene (Figure 2, Lim and Maher, 2010). Changes in the levels of DNA methylation can result in a different expression profile for an organism (Lim and Maher, 2010). This can be linked to various forms of cancer and disease (Lim and Maher, 2010).

Figure 2: DNA methylation and gene expression. A) shows how an unmethylated promoter region allows the binding of transcription factors permitting gene expression. B) indicates the presence of methylation on cytosine’s located within the promoter preventing the transcript of the gene.
Methylation plays a role in regulating various genes. In humans, methylation has been shown to play a role in the repression of tumor suppressor genes leading to cancer (Lee et al., 2013). In *Schistosoma mansoni*, a human parasite, adjusting the levels of methylation can interfere with its ability to produce eggs hindering its ability to infect (Lim and Maher, 2010). In *Magnaporthe oryzae*, a fungi that infects rice, methylation aids in repressing the expression of transposable elements (Jeon, 2015). *Helicobacter pylori* infections are known to cause cancer in humans when found outside the stomach and in the small intestines. It can inhibit the expression of tumor suppressor genes by increasing the expression of DNA methyltransferases. This occurs in the lining of gastric mucosa, leading to cancer (Bierne et al., 2012). An enzyme referred to as DNA Methyltransferase (DMNTs) is responsible for causing methylation on the genome.

Oomycetes do not contain any known DMNTs, suggesting that DNA methylation is not present. However, in 2008, Van West showed that 5-aza-2’-deoxycytidine (5-aza) was able to temporarily reverse a silenced gene. Bisulfite sequencing of the same region yielded opposing results. They tested the methylation status of over 50 potential methylation sensitive sites in both the *Inf1* gene and promoter region with bisulfite sequencing. Through bisulfite sequencing and methylation sensitive digestive enzymes they determined that there is no methylation present in *P. infestans* (van West et al., 2008).

**Detection of 5-Methylcytosine**

*Bisulfite Sequencing*

5-Methylcytosine can be detected by bisulfite sequencing. This process has previously been used within the Phuntumart lab. In that study, changes in expression of virulence genes was observed (Spangler, 2012). With this method bisulfite is added to isolated DNA sample. Once
treated, unmethylated cytosines will undergo a transition into uracils. If the cytosine base was modified with the addition of a methyl group, the transition will not occur in the presence of bisulfate. The treated DNA is then used for amplification via polymerase chain reaction (PCR), where the uracil is then turned into a thymine. After PCR, the amplicon can be sequenced and compared against an already published genome (Figure 3). In recent years, it has been used to identify methylation across the whole genome of humans (Libertini et al, 2016). It was also used to identify the methylome (methylation across a genome) on Magnaporthe oryzae (Jeon et al, 2015).

Figure 3: Bisulfite sequencing. The process of bisulfite sequencing: DNA is treated with bisulfite, then the treated DNA was subjected to PCR, and last sequencing to identify the location of cytosine methylation on DNA. Any location that maintains the cytosine base is believed to be methylated. Blue indicates methylated cytosine while red shows the unmethylated base.

One drawback to this method is that it requires a published genome for comparison and primer design. Damage to the DNA template can also occur due to the changes from cytosine to thymine in the non-methylated bases, reducing the effectiveness of this method. Another
shortcoming is that a complete conversion to uracil may not always occur, leading to contradicting results. In addition, its chances of working on an organism with low levels of methylation are not significant. This is due to the fact that it hinges on picking a part of the genome that is methylated (Liu et al, 2004).

*Restriction Enzyme Analysis*

There are four distinct types of restriction enzymes, type I – IV. Type I restriction enzymes cut at random areas of DNA far from the recognition site and were originally thought to be rare. They offer little experimental use as they form no distinct banding patterns or fragments. The next type of restriction enzyme is divided into three categories: type II, type IIS, and type IIG. Type II are the most common of this type of restriction enzyme and they cut near the recognition sequences. This creates distinct banding patterns and fragments. This class of enzymes are largely unrelated to each other but are increasingly indicted in host parasite interactions. Type IIS differ from type II by only cutting one side of the recognition site. They contain two main domains, a DNA binding domain and a DNA cleavage domain. Type IIG are the largest of the type II restriction enzymes ranging, from 850-1250 amino acids. They cleave outside of the recognition site at either a continuous sequence on one side, or discontinuous sequence on both sides. Type III restriction enzymes are restriction and modification enzymes that cleave outside of the recognition site. They require two sequences in opposite directions with the same recognition site to cleave. This can result in a non-complete digestion. Type IV restriction enzymes can recognize methylated or modified DNA.

Restrictions enzymes have been used to characterize methyltransferases genes and identify methylation in specific promoter regions. M.TneDI, a methyltransferase from *Thermotoga sp.* was characterized with the use of type IV digestive enzyme (Xu et al, 2014).
HpaII/MspI (isoschizomers) have been used to identify methylation in the promoter of the FAS, a gene that may play a role in colorectal cancer in humans (Manoochehri et al, 2016). SmaI and Xmal were used to determine the methylation patterns of healthy human white blood cells (Jelinek et al, 2012).

**DNA Methyltransferase 1-Associated Protein (DMAP1)**

DMAP1 was found in the model organism *P. sojae*. It has an ascension number of EGZ18954.1, and is classified as PHYSO_561151 on FungiDB.org. In humans, DMAP1 is a subunit that forms a complex with DNMT1 and is involved in the repression or activation of transcription. This occurs by interacting directly with the N-terminus of DNA methyltransferase 1. It is also a component of the nucleosome acetyltransferase of H4 complex which works with factors that aid as tumor suppressors (Rountree et al, 2000). While oomycetes lack any known DMNTs they still contain the DMAP1 gene. Homologous of nucleosome acetyltransferase of H4 are also found in oomycetes. The presence of DMAP1 indicates a possible mechanism of DNA methylation on the genome. This is a potential mechanism for controlling gene expression, conceivably even some virulence genes.

**Aims**

Aim I: Identification of DMAP1 from the available genome sequences of oomycetes

Aim II: Isolation of the DMAP1 from *P. sojae*

Aim III: Functional analysis of both synthesized DMAP1 and native DMAP1 from cDNA of *P. sojae*
CHAPTER 2: MATERIALS AND METHODOLOGIES

Aim I - Bioinformatics Analysis

*Phylogenetic Tree of DMAP1*

A bioinformatics approach was used to retrieve all available oomycete DMAP1 sequences from FungiDB.org. To achieve this, a domain search for DMAP1 was performed against all published oomycetes on FungiDB with non-oomycete species obtained from NCBI. Those sequences were then downloaded and aligned using ClustalOmega (Sievers et al, 2011). A phylogenetic tree was constructed using maximum likelihood on MEGA 6 with a 1000 bootstraps (Tamura et al, 2013). The tree was generated to gain a better understanding of the evolutionary track of this gene. The sequence data was also used in the creation of primers for the isolation of the gene from *P. sojae* cDNA.

*Expression Analysis of DMAP1*

Expression analysis of DMAP1 was performed *in silico* using existing RNA-seq data with FungiDB, Galaxy, and RNArocket. Expression data obtained from FungiDB was directly used for this study. Sequence read archive (SRA) from *Phytophthora capsici* (SRA: SPR024305), was identified on ncbi.nlm.nih.gov and uploaded to Galaxy. Those files were then processed through RNA Rocket into a Cufflink file and compared against the reference genome on FungiDB.

Aim II – Cloning of DMAP1

*Cloning the Synthetic Version of DMAP1 (sDMAP1)*

The DMAP1 gene was synthesized by Genewiz (2007), referred to as sDMAP1 hereafter. This version was based on the manual annotation of the *P. sojae* genome without RNAseq
support. The synthesized version of the gene contains the DMAP1 domain however, it is shorter than the native version (Appendix A).

Table 1: Primers for cloning DMAP1. The yellow sequence in the DMAP1 F is the recognition site used to clone into the one directional cloning vector pENTR SD/D-TOPO. The light blue sequence in the DMAP1 attB primers contain the attB recognition sites required for BP Clonase II to insert the desired gene into pDONR221. The green sequence indicates the start codon for DMAP1. M13 primers were also used however, since they are universal primers they were not listed in this table.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5’-3’)</th>
<th>Tm (°C)</th>
<th>GC (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMAP-1 F</td>
<td>CACCATGAGCGACGTGGCG</td>
<td>64</td>
<td>70</td>
</tr>
<tr>
<td>DMAP-1 R</td>
<td>TGTTGTAATGGGATGACATTAGTGCGTAACAAGTTAATGCGG</td>
<td>63</td>
<td>40</td>
</tr>
<tr>
<td>DMAP-1 attB F</td>
<td>GGGGACAAGTTTTGACAAAAAGCAGGCCCTAATGGAAGATA</td>
<td>65</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>TTAAGTATCCGGTACTAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMAP-1 attB R</td>
<td>GGGGACCACCTTTGACAAAAAGCAGCCTGGGCTCACAGGCTTGA</td>
<td>70</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>CTTTCTACCGGTA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: PCR protocols for DMAP1. With the use of Phusion HF DNA polymerase, 72°C and the shorter extension time were used. For Taq2x MM, 68°C was used and the extension time was doubled.

<table>
<thead>
<tr>
<th>PCR Protocol</th>
<th>Primers Used</th>
<th>Temp (°C) - Time (Sec)</th>
<th>Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>sDMAP1</td>
<td>DMAP-1 attB</td>
<td>Initial Denature: 95°C - 30s</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Denature: 95°C - 30s</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Annealing: 59°C - 30s</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Extension: 68/72°C - 45/90s</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Final Extension: 68/72°C - 300s</td>
<td></td>
</tr>
<tr>
<td>nDMAP1</td>
<td>DMAP-1</td>
<td>Initial Denature: 95°C - 30s</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Denature: 95°C - 30s</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Annealing: 60°C - 30s</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Extension: 68/72°C - 45/90s</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Final Extension: 68/72°C - 300s</td>
<td></td>
</tr>
<tr>
<td>Sequencing</td>
<td>M13</td>
<td>Initial Denature: 95°C - 30s</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Denature: 95°C - 30s</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Annealing: 53°C - 30s</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Extension: 68/72°C - 90/180s</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Final Extension: 68/72°C - 300s</td>
<td></td>
</tr>
</tbody>
</table>

An amplicon containing sDMAP1 was created using PCR protocol sDMAP1 (Table 2) and attB primers (Table 1). During the cloning process, Phusion High-Fidelity DNA polymerase from New England Biolabs (NEB, Ipswich, MA) was used, to avoid any mutations during the PCR process. In all cases the PCR was set to 10 uL while primer concentrations were set to
0.5uM for each primer, totaling in 1 uM of primers per reaction. All primers were purchased from Integrated DNA Technologies (IDT), CA. The newly generated PCR fragment contained both the attB recognition sites and sDMAP1. After creating the fragment, gel electrophoresis was performed to confirm the gene was successfully amplified. The PCR product then underwent PCR cleanup. The PCR Purification Kit was performed as described and was purchased from Qiagen, MD. A 100 ng of the AttB-PCR product was incubated with 150 ng of pDONR221 (Figure 4) and BP Clonase II, according to enzymes specifications (Figure 5A). The attB and attP sites reacted with the BP Clonase II enzyme to insert the desired fragment and remove the ccdB from pDONR221. After cloning, 1 uL of the BP reaction was transformed into TOP10 *Escherichia coli* cells. All chemically competent *E. coli* cells were purchased from NEB, MA and all samples underwent heat shock transformation method. All *E. coli* samples were grown in Luria-Bertani media with the appropriate antibiotics at 37°C and shaken at 200 rpm. TOP10 cells were used both to maintain and store the gene, with a glycerol stock stored at -80°C. Colony PCR was performed with universal M13 primers with the sequencing protocol (Table 2) and gel electrophoresis was used to confirm the presence of the gene. The positive samples were grown and plasmid extraction was then performed. All plasmid extractions were completed with plasmid mini prep kit from Zymo, CA without modification. The extracted vectors were then sent out and sequenced using universal M13 primers at DNA Analysis, LLC in Cincinnati, Ohio.
Figure 4: pDONR221 vector map. An entry vector used to clone sDMAP1 (Invitrogen, CA).

Figure 5: Cloning method for DMAP1. 

A) the attB and attP sites (blue) were used to clone sDMAP1 (red) into the entry vector pDONR221, while removing the ccdB toxin (green). 

B) indicates the steps taken in TOPO one directional cloning, this method was used to clone nDMAP1 (red). nDMAP1 was inserted into the entry vector, while removing the toxin gene (green). 

C) depicts how the destination vector was generated with the use of attR and attL sites (orange).
After confirming the gene was present and in frame, the entry vector was re-extracted from *E. coli*. It then underwent a LR reaction, to clone the gene into the destination vector pBAD-DEST49 (Figure 5C). A 100 ng of pDONR221 was incubated with 300 ng of pDEST49 with LR Clonase as specified in the manual. LR Clonase enzyme recognizes the attL and attR sites and effectively swaps the two, resulting in the addition of the desired gene and the removal of the toxin gene in the destination vector. Both the entry vector and destination vectors contain different antibiotic selection markers. The destination vector (1 uL of the LR reaction) was then subsequently transformed into TOP10 *E. coli* cells. To reaffirm the presence of the gene colony PCR was performed using protocol sDMAP1 (Table 2) and attB primers (Table 1) with Taq 2X Master Mix (NEB MA) and confirmed through gel electrophoresis. Positive transformants then had their plasmid extracted and transformed (100 ng of plasmid DNA) into the Dcm⁻/Dam⁻ mutant strain of *E. coli*. Another round of colony PCR (Table 2) with attB primers (Table 1) was used to again reconfirm the presence of the gene. The colonies containing the gene were grown in liquid culture to produce a glycerol stock stored at -80°C. Due to the high mutation rate of this strain each experimental run was performed using fresh samples taken from this glycerol stock so subculturing was avoided.

The *E. coli* mutant Dam⁺/Dcm⁻ was chosen for this experiment due to its lack of any methyltransferase genes. As a result, it had a lower transformation efficiency. This strain is also nuclease deficient due to the lack of methylated DNA. As a result; it was given a new method of resistance to the T1 phage by removing the ferric hydroxamate receptor. Due to the lack of methylation the strain is known to have a slower growth rate when compared to other strains. It is also known to have a higher mutation rate, due to the loss of the DNA mismatch repair pathway.
Figure 6: pBAD-DEST 49 vector map. It is an expression vector used in *E. coli* (Invitrogen, CA).

The plasmid, pBAD-DEST 49 (Figure 6), is 6.1 kb, it contains the selective marker for ampicillin (Amp) resistance and has an inducible system containing the araBAD operon. It is regulated by the addition or omission of L-arabinose. The araC gene causes the negative regulation of the operon along with the DMAP1 gene further downstream by binding to the araI₁ and araO₂ sites. When arabinose is added, it binds to araC protein allowing the operon to be accessed by transcription machinery. Arabinose also acts as an activator, as it forms a complex between araC and arabinose. This complex can then bind to the promoter and allow for RNA polymerase to bind and transcribe the gene. pBAD-DEST 49 also contains the ccdB gene, which serves as a negative control marker to select for the positive transformants. If the mutant strain has successfully taken in the plasmid without the insert, it will create the ccdB toxin which will lead to cell death. The vector must be maintained in a Ptrc-ccdA strain that can create the ccdA gene to negate the ccdB toxin. The attR₁ and attR₂ sites allow for easy cloning of a gene into other cloning vectors that use the same gateway system and is based off a viral sequence that allows for the insertion of DNA at specific sites. There is an *rrn*B site contained after the poly-histidine sequence near the C-terminus to ensure a strong transcriptional stop site. The poly-
histidine tag sequence on the C-terminus is present to allow protein purification on a metal chelating resin after inducing the pBAD operon. Near the N-terminus, there is the HP-thioredoxin which is used to purify the crude extract. Thioredoxin also helped create a soluble protein product.

*Cloning the Native Version of DMAP1 from P. sojae (nDMAP1)*

The native version of DMAP1 (nDMAP1) was isolated from *P. sojae* cDNA. *P. sojae* (strain P6497) was grown on 15% V8 media with agar containing a layer of cellophane. After five days of growth mycelium was removed from the cellophane and used for RNA extraction. The extraction process was carried out on 100 mg of mycelium with Qiagen RNease Mini Kit (Qiagen, MD), according to product specifications. Liquid nitrogen with a mortar and pestle was used to homogenize the sample prior to RNA extraction. The quality and integrity of the RNA sample was ascertained using photo spectrometry (Nanodrop 2000) and gel electrophoresis, respectively. After RNA extraction, 1 ug of total RNA was used to generate cDNA with Oligo(dT) primers and SuperScript III Reverse Transcriptase (Thermo Fisher, MA), according to the manufacturer specifications. All equipment used during the creation of cDNA originated from the SuperScript™ III Reverse Transcriptase kit. PCR with nDMAP1 protocol (Table 1) was used with DMAP1 primers (Table 1) to amplify DMAP1 from cDNA. This was used to generate a fragment that contains CACC at the start of the sequence. That sequence is recognized by TOPO (Topoisomerase I) and inserted into the Gateway entry vector pENTR/SD/D-TOPO (Figure 7), while removing the ccdB toxin gene (Figure 5B). In every PCR used for cloning, Phusion High-Fidelity DNA polymerase from NEB, MA was used. Gel electrophoresis was then used to confirm positive amplification of the gene. After confirmation, PCR clean-up was performed with Qiagen PCR Purification Kit according to the products specifications.
Figure 7: pENTR/SD/D-TOPO vector map. Shows the entry vector used to isolate and clone the native version of DMAP1 from the cDNA of *P. sojae* (Invitrogen, CA).

TOPO one directional cloning was used to clone nDMAP1 into an entry vector. A 100 ng of the PCR product was cloned into the Gateway entry vector pENTR/SD/D-TOPO (Figure 7) according to the product specifications (Figure 5B). Following cloning, transformation (1 µL of the TOPO PCR cloning reaction) was performed on TOP10 competent cells. Then colony PCR (protocol Sequencing Table 2, with universal M13 primers) was used to determine if cloning was successful. The colonies containing a positive hit were then grown in liquid media for plasmid extraction, as previously stated. The plasmids were then sequenced at DNA Analysis, LCC in Cincinnati Ohio, to determine if it was both the correct gene and in frame.

The entry vector containing the in-frame gene was re-isolated. An LR reaction then took place as previously stated to transfer the gene into the destination vector (Figure 5C). It was transferred into two destination vectors; pBAD-DEST 49 (*E. coli*) and pYES-DEST 52 (*Saccharomyces cerevisiae*). Once cloned into the destination vectors, they were then transformed into *E. coli* TOP10 cells. The plasmids were again re-extracted and transformed into *E. coli* and *S. cerevisiae* mutants. In *S. cerevisiae*, a standard protocol for electroporation was
used for transformation (“Yeast Transformation”, n.d.). This method was chosen over chemical
transformation due to its higher transformation efficiency. Colony PCR (protocol cDNA, Table 2
with DMAP1 primers, Table 1) and gel electrophoresis were then performed on the
transformants to confirm the presence of DMAP1 in the mutant samples.

pYES-DEST 52 (Figure 8) is 7.6kb in size and shares many of the same features as
pBAD-DEST 49 except that it is designed for use with \textit{S. cerevisiae} instead of \textit{E. coli}. It contains
URA3 gene as a selection marker to confirm positive transformants and uses the galactose
promoter to induce the gene of interest instead of L-arabinose (Invitrogen, 2010). This plasmid
was used with \textit{S. cerevisiae} YSC1048 that was derived from the parental strain BY4741 and is a
MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 mutant. The purpose of using a yeast strain was twofold.
First, it is a unicellular eukaryote model organism that should lead to the correct conformation of
the DMAP1 gene. This allowed for post translation modification (PTM) of the protein, while a
prokaryotic system might lead to incorrect folding of the protein. Second, it serves as a good
model for gain of function of methylation gene due to the lack of any definitive methylation on
its genome (Jost, 1993).

![Figure 8: pYES-DEST52 vector map. Expression vector used in \textit{S. cerevisiae} (Invitrogen, CA).](image.png)
Aim III – Functional Analysis of DMAP1

Toxicity Assay in Prokaryotes

Before investigating the function of DMAP1, the toxicity of both sDMAP1 and nDMAP1 was determined in *E. coli*. To accomplish this, the gene was induced according to pBAD-DEST49 protocol (Invitrogen, CA). Growth was monitored via OD600 with a photospectometer once every hour. The Dam−/Dcm− mutant containing pUC19 was also measured, as to monitor the normal growth rate of the mutant. Dam−/Dcm− mutant containing pUC19 also served as an in vitro control. The inducing agent L-arabinose (Sigma-Aldrich) was made into a 20% stock, filter sterilized through a 0.22um filter, and stored at -20°C until used. It was added once the OD600 reached ~0.5, when the cells were in exponential growth phase. L-arabinose was set to final concentration of 0.2% in the culture. All samples were compared to a negative control where the inducible agent was omitted.

Toxicity Assay in Eukaryotes

In *S. cerevisiae*, only the nDMAP1 in pYES-DEST 52 was analyzed. To monitor its normal growth rate, yeast without a plasmid was also grown in YPD (Yeast extract, Peptone, Dextrose) and OD600 was measured. The induced samples containing pBAD-DEST52 were grown in SC-ura (minimal media for *S. cerevisiae*) with 2% galactose (Induced) or 2% glucose (Not-Induced), according the plasmids prescribed protocol (Invitrogen, 2010). Growth of the samples were measured every four hours for 24 hours using the OD600 value. After confirming that the gene was not toxic to the cell, a functional analysis with restriction enzymes was performed on extracted DNA.
**Restriction Digestion Enzymes**

After determining the toxicity of the protein, a methylation sensitive restriction digestion assay was carried out on DNA from all samples. All DNA was isolated from non-induced samples serving as the negative control, and induced cultures creating the test group. To carry out the digestion assay multiple restriction enzymes were used: McrBC, MspJI, HpaII, and MspI.

Table 3: Restriction digestion enzymes. Covers the enzymes used in this study, along with their consensus sequences and their ability to digest the sequence. Light blue indicates the presence of a modified cytosine base. Red shows the cytosine base without methylation. Green indicates the scenarios where the enzyme will cut while red shows when it will not digest.

Table 3: Restriction digestion enzymes.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Consensus Sequence</th>
<th>Ability to Cut</th>
<th>Key</th>
</tr>
</thead>
<tbody>
<tr>
<td>McrBC</td>
<td>RC&lt;sup&gt;+&lt;/sup&gt;-30-3000bp-RC&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Will Cut</td>
<td>R = A,G</td>
</tr>
<tr>
<td></td>
<td>RC&lt;sup&gt;-&lt;/sup&gt;+30-3000bp-RC&lt;sup&gt;-&lt;/sup&gt;</td>
<td>Blocked</td>
<td>N = A,T,G,C</td>
</tr>
<tr>
<td>MspJI</td>
<td>CNNY</td>
<td>Will Cut</td>
<td>Y = A,T</td>
</tr>
<tr>
<td></td>
<td>CNNY</td>
<td>Blocked</td>
<td></td>
</tr>
<tr>
<td>MspI</td>
<td>CCGG</td>
<td>Will Cut</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CCGG</td>
<td>Will Cut</td>
<td></td>
</tr>
<tr>
<td>HpaII</td>
<td>CCGG</td>
<td>Blocked</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CCGG</td>
<td>Will Cut</td>
<td></td>
</tr>
</tbody>
</table>

McrBC is an endonuclease restriction enzyme that can identify CpG dinucleotides, and can only cut when methyl-cytosine is present (Table 3). Its recognition site is a purine followed by a methylated cytosine, a gap of 40-3000 bases and another purine and methylated cytosine (Sutherland, 1992). It can detect methylation up to 3-kilobases apart and is competent even if methylation is very minute. It cleaves within approximately 30 bases from the methylated base that was recognized and has not been known to create distinct ends. Due to this in-distinction, there can be an overlap of sequences after digestion and sequencing at the end of each fragment. McrBC is produced from two separate plasmids one containing McrB and another one carrying
McrC (Sutherland, 1992). MspI and HpaII are isoschizomers and are both capable of recognizing the CCGG motif (Table 3). When the first cytosine in the motif is methylated, they are not capable of cutting; however, MspI is able to cut when there is methylation on the second cytosine. The HpaII gene originates from *Haemophilus parainfluenzae* while the MspI gene comes from *Moraxella* species, both are produced from an *E. coli* strain. MspJI (Table 3) recognizes and only cuts in the presence of a modified cytosine base; it can recognize both C5-methylation (5-mC) and C5-hydroxymethylation (5-hmC). MspJI originated from *Mycobacterium* species. HF-EcoRI (not shown in Table 3) was also used as a means of linearizing plasmid DNA. All enzymes and associated buffers were purchased from NEB, MA.

*Functional Analysis in Prokaryotes*

The digestion restriction assay was performed on both genomic and plasmid DNA originating from the Dam⁺/Dcm⁻ *E. coli* mutant. Both versions of DMAP1 were studied in this *E. coli* mutant. The samples were grown as previously specified and DNA was extracted from those samples. A double digestion was carried out on only plasmid DNA; one for linearizing the DNA, the other to test for the presence of methylation. All *E. coli* plasmid DNA was isolated using Zymo plasmid mini prep kit, according to the manufacturers specifications. Total DNA from *E. coli* was extracted with a basic phenol/chloroform protocol (He, 2011). Every restriction digestion assay was performed on 1 ug of DNA and was implemented according to the specifications of each enzyme. CutSmart buffer from NEB (MA) was used for each reaction; with the addition of GTP and BSA to McrBC, and Enzyme Activator Solution in MspJI. Each reaction was carried out at 37°C for one hour, followed by deactivation at 65°C for 15 mins. Gel electrophoresis with was performed with a 1 kb ladder (NEB, MA) and the image was captured with Bio-Docs EZ-Gel imager and image lab software.
Functional Analysis in Eukaryotes

Only the nDMAP1 was studied in the eukaryotic model *S. cerevisiae*. Genomic DNA from *S. cerevisiae*, was extracted with a modified phenol/chloroform DNA isolation method; with the use of lithium acetate (LiAc) as the lysing agent (Protocol in Appendix B). Genomic DNA was used to then carry out the digestion assay. Every restriction digestion assay was performed on 1 ug of DNA and was set up according to the specifications of each enzyme, as previously described. Gel electrophoresis was performed with a 1 kb ladder (NEB, MA) and the image was captured with Bio-Docs EZ-Gel imager and image lab software.

Positive Controls

An *in vitro* positive control was created with genomic DNA taken from Dam'/Dcm' containing the pUC19 plasmid treated with HpaII methyltransferase. HpaII methyltransferase can methylate the internal cytosine in the CCGG recognition sequence. Once it modified the DNA the digestive enzymes used in this study should interact with the DNA; except MerBC which doesn’t contain the CC recognition sequence. A total of 5 ug of *E. coli* DNA was treated with HpaII methyltransferase according to the specifications of the enzyme. Then 1 ug was aliquoted for each of the of the four restriction enzymes used in the assay, digestions were performed as previously described. Leaving 1 ug to as a control to assess the integrity of the undigested DNA. An image was then captured with Bio-Docs EZ-Gel imager and image lab software after undergoing gel electrophoresis.

*Nicotiana benthamiana* was used as the *in vivo* positive control. It was chosen because it contains thirteen different DNA methyltransferase members (Kumar et al, 2016). Cytosine methylation in plant genomes typically occur as CG, CHG, and CHH; with more than one third
of it occurring at non-CG sites (Vanyushin and Ashapkin, 2011). DNA from *N. benthamiana* was isolated with CTAB and phenol/chloroform protocol, and the restriction digestion assay was performed on its genomic DNA, as previously described.
CHAPTER 3: RESULTS

Aim I – Bioinformatics Analysis

Phylogenetic Analysis of DMAP1

An *in-silico* analysis was performed to identify DMAP1 in all published oomycete genomes. These results indicated that both sDMAP1 and nDMAP1 contain the DMAP1 domain, as previously described (Figure 9). However, nDMAP1 also contains a SANT/Myb-like domain that sDMAP1 lacks (Figure 9). This domain is a tandem repeat of three alpha helices forming a helix turn helix motif, resembling a Myb DNA-binding domain (EMBL-EBI, 2017).

![Figure 9: DMAP1 domain. Shows the domains in sDMAP1 and nDMAP1. The domain search was generated from PFAM (http://pfam.xfam.org/) with an amino acid search. The amino acid sequence for both versions of the gene were used and the alignment (shown from green/good alignment to red/poor alignment) of DMAP1 domain, which is present in both versions. The nDMAP1 version also contains a potential SANT domain which resembles a DNA binding Myb transcription factor.](image-url)
A phylogenetic tree was generated from sequence data taken from FungiDB. The generated tree shows that DMAP1 groups closest to plant, diatoms, and algae. Based off the phylogenetic tree the DMAP1 protein is conserved in all published oomycete genomes found on FungiDB.org (Figure 10).
Figure 10: A phylogenetic tree of DMAP1. All sequences were taken from FungDB.org. The tree was constructed using maximum likely hood with 1000 bootstraps with MEGA6 software (Tamura et al, 2013). All values under 60 were omitted. *P. sojae* was indicated with a red arrow.
In silico Expression Analysis of DMAP1

DMAP1 in Phytophthora ramorum shows distinct levels of expression in all tested stages of its life cycle (Figure 11). It illustrated that a chlamydospore has the highest levels of expression with the zoospore stage having the lowest. DMAP1 in P. sojae shows elevated levels of expression during the mycelium stage (Figure 12). While the lowest levels occurred during late infections (three-day post infection of soybean hypocotyls). Phytophthora capsici expression data (Figure 13) was generated from Galaxy-RNArocket (as of 12/2015). It indicated that highest levels of expression occurred during the mycelium stage. Additionally, it suggested that the prediction of mRNA gene model may not be accurate, as the expressed sequences do not match the gene model. In short, P. capsici expression data mirrors that of P. sojae and P. ramorum.

Figure 11: Expression data of DMAP1 in Phytophthora ramorum (FungiDB.org). The Y-axis indicates the expression level of the DMAP1 gene in fragments per kilobase of exon model per million mapped sequences. The x-axis shows that different life cycles and growth media can affect the rate of expression of DMAP1. It indicated that it had the highest levels of expression during the dormant stage of its life cycle and the lowest during its motile stage (Stajich et al., 2012).
Figure 12: Expression data of DMAP1 in *P. sojae* (FungiDB.org). The Y-axis indicated the expression level of the DMAP1 gene in fragments per kilobase of exon model per million mapped sequences. The x-axis indicated that changing the life cycle of the organism influenced the rate of expression of DMAP1. Infection data was taken 3 days’ post infection from soybean hypocotyls. The graph illustrated that DMAP1 has the highest levels of expression during vegetative growth (Stajich et al, 2012).
Figure 13: RNAseq data of *Phytophthora capsici* taken from RNArocket (2015). *P. capsici* during three different life cycles compared against the reference genome taken from FungiDB. During the mycelium stage (Myc), it showed the highest levels of expression and the lowest while being a cyst (Cys). Zoospores (Zoo) had moderate expression levels between the two. It shows that changing the life cycle again effects the level of expression of the DMAP1 gene. It also showed which regions of the gene were used to create the finished transcript.

**Aim II – Cloning of DMAP1**

*Cloning the Synthesized Version of DMAP1 (sDMAP1)*

A synthetic version of the gene, sDMAP1 was ultimately cloned into the destination expression vector pBAD-DEST49. To achieve that, the attB fragment (Figure 14) was initially cloned into pDONR221. It was then subsequently transformed into *E. coli* TOP10 cells. Colony PCR of the transformants with M13 primers confirmed the presence of the gene with the desired ~1.5 kb band (Figure 15). The entry vector from samples G, H, and I were isolated and
sequenced. Sequencing results indicated that sample H contained the correct gene and was in frame. The pDONR221 plasmid containing the gene from sample H underwent an LR reaction and was cloned into the destination vector. Once in pBAD-DEST49 it was ultimately transformed into Dam\(^-\)/Dcm\(^-\) cells; where colony PCR (Figure 16) was used to confirm the presence of the gene. Sample H was randomly selected and used to carry out the restriction digesting assay.

![Figure 14: Attb-PCR product. This fragment was used for cloning sDMAP1 into pDONR221. The Attb-PCR band was generated with PCR protocol sDMAP1 (Table 2) with Attb primers (Table 1). It is consistent with the expected band size of ~1.3 kb due to the addition of the attB sites. The negative control underwent the same reaction without the presence of DNA (indicated as N). The ladder was a 1 kb ladder purchased from NEB, MA.](image-url)
Figure 15: Colony PCR of sDMAP1 from TOP10 transformants (M13 primers with Sequencing protocol from Table 2 were used). This showed the results of nine random colonies that underwent transformation with the pDONR221 entry vector containing DMAP1. Samples G, H, and I were isolated and sent out for sequencing as they matched with the expected band size of ~1.5 kb. A 1 kb ladder (NEB, MA) was used and labeled as Ladder. A negative control (shown as N) underwent the same reaction with the absence of *E. coli* cells.
Figure 16: Colony PCR of sDMAP1 in *E. coli* Dam⁺/Dcm⁻. Ten random Dam⁺/Dcm⁻ transformants were selected and screened, to ascertain if the mutant contained DMAP1 gene in the pBAD-DEST49 plasmid. Attb primers (Table 1) and the sDMAP1 protocol (Table 2) were used. A 1 kb ladder (NEB, MA) was used (indicated as Ladder), A negative control was used in the absence of DNA samples (labeled as N). Sample B, D, H, and J appeared to have the desired band size of ~1.3 kb.
Cloning the Native version of DMAP1 from P. sojae (nDMAP1)

The cDNA created from a P. sojae mRNA was used to isolate nDMAP1. To achieve this, PCR was used to amplify it with DMAP1 primers. The resulting PCR products showed multiple bands, indicating that non-specific binding occurred (Figure 17). Since the expected size band (~1.5 kb) was present, it was used in the cloning procedure. Colony PCR with universal M13 primers were used to screen the transformants. Twenty random colonies were screened in total, producing five different band sizes (Figure 18). One colony from each of the band sizes were sequenced. A sequence result from sample O indicated that it contained the full length of nDMAP1 and was in frame. Sample O then was used to clone into two destination vectors pBAD-DEST52 for E. coli and pYES-DEST52 for S. cerevisiae. The two destination vectors containing the cDNA version of DMAP1 were then ultimately transformed into their respective mutants. A final round of colony PCR was used, confirming the presence of the nDMAP1 in both E. coli (Figure 19) and S. cerevisiae (Figure 20). For E. coli, sample H was selected to be used in the digestion assay while sample G was chosen for S. cerevisiae.
Figure 17: Optimization of PCR for nDMAP1 amplification. This PCR reaction used protocol nDMAP1 (Table 2) with a temperature gradient from 55-59°C was used for the optimization of the DMAP1 primers (Table 1). Multiple band sizes indicate that nonspecific binding occurred with the DMAP1 primers and the cDNA template from *P. sojae*. The expected band size was around ~1.5 kb. The negative control (indicated by N) where it underwent the same reaction in the absence of cDNA.
Figure 18: Colony PCR of nDMAP1 in the TOP10 *E. coli* cells. Twenty random colonies were selected and screened with PCR with protocol Sequencing (Table 2) and M13 primers. Five different band sizes appear on this gel. Samples C ~1-1.5 kb, E & O at ~2 kb, H at ~1.5 kb, I, J, and S at ~1 kb, and finally sample N at ~750bp. The negative control (indicated as N in the top portion of the gel) which underwent the same reaction in the absence of a colony. Blue lines generated from with Bio-Docs EZ-Gel imager and image lab software were removed and replaced with black to improve visibility.
Figure 19: Colony PCR of nDMAP1 in Dam\textsuperscript{-}/Dcm\textsuperscript{-}. Ten random colonies were selected and screened with DMAP1 primers (Table 1) and protocol nDMAP1 (Table 2). Samples A, D, F, G, and H contain the cDNA version of DMAP1 as they have a ~1.5 kb band. A 1 kb ladder from NEB, MA was used (labeled as Ladder). A negative control (labeled as N) was generated by the same reaction without the presence of culture.
Figure 20: Colony PCR of nDMAP1 in *S. cerevisiae*. Seven random colonies were tested for nDMAP1 in the Ysc1048 mutant. The negative control (indicated by N) underwent the same reaction in the absence of a colony. A one kb ladder from NEB, MA was used (labeled as Ladder).
Aim III – Functional Analysis of DMAP1

Toxicity Assay

All samples were grown to test if either version of DMAP1 was toxic to their respective host cell. In *E. coli*, it was grown with and without L-arabinose (previously described) and then growth was measured by OD600 from 4-6 hours’ post induction, once every hour. Both versions of DMAP1 proved to be non-toxic to *E. coli* (Figure 21); as indicated by positive slopes. Only the nDMAP1 version was transformed into *S. cerevisiae*, it was grown for 24 hours in glucose (Non-Induced) and galactose (Induced), both showing normal growth patterns (Figure 21). This suggest that both versions of DMAP1 are not toxic to their respective host cells. Thus, permitting their use in the restriction digestion assay.
Figure 21: Toxicity assay. This was carried out for both versions of DMAP1 in *E. coli* and *S. cerevisiae*. 

**A**) sDMAP1 in Dam^{-}/Dcm^{-} was compared against a negative control (pUC19 in red) to observe normal growth rates. The non-induced sample gave a comparison for when sDMAP1 was induced. Based off the increasing growth curve sDMAP1 is nonfatal to *E. coli*. 

**B**) nDMAP1 in Dam^{-}/Dcm^{-} was compared against a negative control (pUC19 in red) to observe normal growth rates. The non-induced sample (yellow) gave a comparison for when nDMAP1 was induced. Based off the increasing growth curve, nDMAP1 is nonfatal to *E. coli*. 

**C**) nDMAP1 in Ysc1048, *S. cerevisiae* was grown in complex media (YPD) without plasmid (Ysc1048 in YPD, shown in red) for growth comparison. A non-induced sample (yellow) showed the normal growth rate in minimal media without the expression of DMAP1. DMAP1 was induced (blue) and a similar increasing slope when compared to control suggests that the gene is non-toxic to yeast and that it can be used in future assay.
Functional Analysis with Restriction Enzymes

*N. benthamiana* genomic DNA contained methylated cytosine and was used as a positive control (Vanyushin and Ashapkin, 2011). Results indicated that this method was capable of detecting 5-cytosine methylation (Figure 22). McrBC detected methylation as it was capable of completely digesting the genomic DNA, as indicated by a smear and an absence of a genomic band. HpaII and MspI show a different pattern demonstrating that the internal cytosine is methylated and that most of the organism DNA is methylated at sites outside of the CCGG sequence. MspJI showed the greatest amount of digestion proving the presence of methylation.

To develop a positive control, extracted genomic DNA from Dam\(^{-}\)/Dcm\(^{-}\) *E. coli* mutant containing pUC19 was methylated with HpaII methyltransferase and then used in the restriction digestion assay (Figure 23). *In vitro* methylation was performed by treating DNA extracted from the Dam\(^{-}\)/Dcm\(^{-}\) *E. coli* mutant containing pUC19 with HpaII methyltransferase. The newly methylated DNA was then subjected to restriction enzymes. The sample treated with HpaII methyltransferase and McrBC remained intact. This occurred because of the difference in their recognition sites, HpaII methyltransferase only methylate’s CCGG sites which McrBC cannot recognize. MspI and HpaII showed different bands indicating that the methyltransferase protein protected DNA from HpaII, blocking its ability to digest. MspJI showed a partial digestion, which confered with both consensus sequences. HpaII methyltransferase methylates the internal cytosine in the CC\(^{m}\)GG sequence. MspJI cuts C\(^{m}\)NNY, meaning digestion occurred anytime HpaII methyltransferase methylated a CC\(^{m}\)GGY sequence.

All test groups with both version of DMAP1 in both *E. coli* and *S. cerevisiae* yielded inconclusive results (Figure 24-27). The sDMAP1 version in Dam\(^{-}\)/Dcm\(^{-}\) *E. coli* mutant failed to prove that DMAP1 can methylate DNA (Figure 24). By comparing both the negative control
(Non-Induced samples) to the test group (Induced) we can observe that methylation is not present. DNA subjected to both methylation sensitive enzymes (McrBC & MspJI) appeared intact, signifying the absence of methylation. The DNA exposed to both HpaII and MspI contained similar banding patterns. Showing that the internal cytosine in the CCGG recognition site lacked methylation. The nDMAP1 gene in the Dam'/Dcm' E. coli yielded results that are similar to sDMAP1 (Figure 25). Induced plasmid DNA from nDMAP1 (Figure 26) also mirrored that of the genomic DNA. The linear vs nonlinear (Negative) bands contained a supercoiled plasmid which appeared as a wider band. Due to higher concentrations of DNA, multiple supercoiled bands show up as one individual band as they overlapped each other, appearing as one. If the samples contained a lower concentration it would have appeared as multiple bands. Identical banding patterns also were present in S. cerevisiae with nDMAP1 (Figure 27).

All methylation sensitive enzymes (McrBC & MspJI) in all cases failed to detect methylation as the DNA appears to be largely intact in all test samples. In all experimental groups, MspI and HpaII showed a similar banding pattern also demonstrating a lack of methylation on the internal cytosine. Confirmation of the presence of DNA cytosine methylation with the use of McrBC, HpaII, and MspII the gel results indicated that the AC, GC, and CC may not contain methylation. The recognition sites CNNY (from MspJI) covered the remaining TC sites. There is also the possibility that there may have been low levels of DNA methylation within the host that this method cannot detect.
Table 4: Restriction digestion enzymes. Enzymes used to carry out the functional analysis of DMAP1 (NEB, MA).

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Consensus Sequence</th>
<th>Ability to Cut</th>
<th>Key</th>
</tr>
</thead>
<tbody>
<tr>
<td>McrBC</td>
<td>R**(G/A)<strong>G</strong>(G/A)<strong>, R</strong>(G/A)<strong>G</strong>(G/A)**</td>
<td>Will Cut</td>
<td>( R = A,G )</td>
</tr>
<tr>
<td></td>
<td>R<strong>A</strong>G**(G/A)**</td>
<td>Blocked</td>
<td>( N = A,T,G,C )</td>
</tr>
<tr>
<td>MspII</td>
<td>C<strong>N</strong>NY</td>
<td>Will Cut</td>
<td>( Y = A,T )</td>
</tr>
<tr>
<td></td>
<td>C<strong>N</strong>NY</td>
<td>Blocked</td>
<td></td>
</tr>
<tr>
<td>Mspl</td>
<td>C<strong>C</strong>GG</td>
<td>Will Cut</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C<strong>C</strong>GG</td>
<td>Will Cut</td>
<td></td>
</tr>
<tr>
<td>HpaII</td>
<td>C<strong>C</strong>GG</td>
<td>Blocked</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C<strong>C</strong>GG</td>
<td>Will Cut</td>
<td></td>
</tr>
</tbody>
</table>

Figure 22: Restriction analysis of *N. benthamiana*. This was performed on gDNA. The no enzyme lane shows the genomic DNA that underwent the same process without the addition of a digestive enzyme. McrBC, HpaII, Mspl, and MspJI show the results of the digestion with their respective enzymes. A 1 kb ladder originating from NEB, MA was used.
Figure 23: Restriction analysis of positive control. This analysis was performed on the total DNA from the *E. coli* mutant Dam−/Dcm− with pUC19 plasmid. The lane labeled as no enzyme lane underwent all reactions without the presences of any enzyme. The extracted genomic DNA was then subjected to HpaII methyltransferase enzyme which methylates CCmGG (indicated in green). DNA treated with HpaII Methyltransferase was subsequently labeled as M+. The treated DNA (M+) was then exposed to McrBC, HpaII, MspI, and MspJI. A 1 kb ladder originating from NEB, MA was used.
Figure 24: Digestion assay for sDMAP1 on total DNA from Dam-/-Dcm-. The non-induced samples originate from the mutant grown in the absence of arabinose (negative control). The induced data set shows DNA coming from those that had the synthetic version of DMAP1 induced by arabinose. The no enzyme lanes in both cases undergo the same reaction with the absence of digestive enzymes. The rest are represented by the presence of their respectively labeled enzymes. A 1 kb ladder was used from NEB, MA.
Figure 25: Digestion assay for nDMAP1 on total DNA from Dam^-/Dcm^- pBAD-Dest49 nDMAP1. The Non-Induced set originates from the mutant grown in the absence of Arabinose (negative control). While the Induced bands originates from genomic DNA isolated from E. coli that had the DMAP1 gene induced by addition of arabinose. No enzyme lanes in both cases underwent the same reaction without being exposed to any digestive enzyme. The rest are represented by the presence of their respectively labeled enzymes.
Figure 26: Digestion assay for nDMAP1 on plasmid DNA from Dam\(^{-}/\)Dcm\(^{-}\). Non-induced set originates from the mutant grown in the absence of Arabinose. The induced band shows genomic DNA coming from those that had the DMAP1 gene induced by arabinose. The bands shown under no enzyme in both cases undergo the same reaction in the absence of digestive enzymes. Linear band represents a linear plasmid being that was by EcoRI-HF. The remaining bands are represented by the presence of their respectively labeled enzymes (McrBC, MspI, HpaII, MspII) and the addition of EcoRI-HF. A 1 kb DNA marker was used (NEB, MA).
Figure 27: Digestion assay for nDMAP1 on genomic DNA from *S. cerevisiae*. Restriction. The reaction contains genomic DNA extracted from the *S. cerevisiae*. Non-induced set originates from the mutant grown in the glucose, inhibiting DMAP1. Induced band shows genomic DNA coming from samples grown in galactose. The no enzyme lane in both cases undergo the same reaction in the absence of enzymes. The rest are represented by the presence of their respectively labeled enzymes (McrBC, MspI, HpaII, MspJI). DNA maker is a 1 kb ladder purchased from NEB, MA.
CHAPTER 4: DISCUSSION AND CONCLUSION

To date, no gene model for DNA methylation has been identified in oomycetes. In-silico analysis using both pFAM and the phylogenetic tree, indicating that DMAP1 is conserved in all published oomycete genomes found on FungiDB. Existing expression data also showed that the gene is present and conserved in the genomes of the studied oomycetes. This suggested that DMAP1 also serves a purpose however, the function of DMAP1 is still unknown in oomycetes. Based off expression data, DMAP1 can be considered essential to the organism. Although its exact function in the absences of DMNT is still unknown. This study yielded inconclusive results, as restriction enzyme analysis failed to identify if the gene is capable of causing methylation independently. The failure of this assay could have occurred due to a myriad of reasons during its transition from transcript to protein.

Potential Shortcomings in Current Methodology

In both prokaryotic and eukaryotic species, gene expression hinges on how well various steps are carried out in sequence; from transcription to post translational modification (PTM). The final product can be affected at any step along its path to creating the functional form of DMAP1 protein. Codon bias is one such example that can occur during translation. This factor arose because of to the degeneracy of the genetic code. Organisms prefer one type of codon over another, changing the availability of that specific anticodon/tRNA. Some tRNA are present in higher levels as compared to others that are present at low levels. Codon bias can directly affect the levels of expression of a foreign gene if they are not codon optimized for the organism in use (Heitzer et al, 2007). Codon bias can also lead to translation errors which can create a nonfunctional protein (Heitzer et al, 2007). Expression of the gene might be lowered or even absent in the presence of insufficient translation (Fakruddin et al, 2012). In this study, DMAP1
was not codon optimized for expression in either *E. coli* or *S. cerevisiae*. Both versions of DMAP1 could contain codons that are rarely used within either *E. coli* or *S. cerevisiae*, limiting the rate of translation, and perhaps even halting it (Robinson et al, 1984). To address this problem, a codon optimized sequence of DMAP1 would be generated and then placed in the appropriate mutant.

Another potential factor that could lead to a nonfunctional protein is the absence of a specific subunit or interactive partners. If a protein complex is the desired finished product, then the omission of its partner would be detrimental. When only one subunit is expressed the complex will not form and likely not function as intended. The exclusion of its unknown partner in oomycetes may have led to a nonfunctional protein; since this subunit may not be present in either *E. coli* or *S. cerevisiae*. One method to determine if it is a part of a complex is to do co-immunoprecipitation; extract DMAP1 and then identify its partner with protein sequencing.

In prokaryotic species, recombinant protein expression from eukaryotes genes can result in the expression of proteins that are either unfolded or misfolded (Francis and Page, 2010). The most common causes of protein misfolding in *E. coli* are premature finish, correct conformation failure, or environmental stress. *E. coli* favors de novo protein folding to refold improperly folded proteins. Misfolding occurs because some proteins require longer times or other chaperone proteins to form the correctly folded protein. While shorter proteins with smaller domains lack this problem larger multi-domain proteins end up needing assistance to reach the correct conformation. Without this correct conformation, the fate of the protein is either degradation or the creation of inclusion body (Baneyx and Mujacic, 2004).

Inclusion bodies are commonly formed in bacteria from large proteins lacking PTM. Inclusion bodies accumulate in the cytoplasm and are eventually exploited for production of
other proteins that are typically easy to refold due to their resistance to proteolysis (Baneyx and Mujacic, 2004). One possible method to avoid this, is to induce at lower temperatures.

*E. coli* also lacks chaperone proteins that are required to generate some recombinant proteins. Molecular chaperones are constitutively expressed ubiquitous features of eukaryotic cells, that function as a quality control for protein synthesis. Chaperonins are classified as stress or heat-shock proteins since they are upregulated in stress events. They do not have strict target motifs since their target motifs of short stretches of hydrophobic amino acids flanked by basic residue and lacking acidic residue (Baneyx and Mujacic, 2004).

Post translation modification (PTM) is another potential factor for DMAP1 in *E. coli*. Prokaryotes lack the ability to perform some types of PTM (Grangeasse et al, 2017). It is commonly needed to activate or change the newly made protein (EMBL, 2017). In eukaryotes, some proteins require glycosylation and phosphorylation after translation. Prokaryotes are not capable of performing these types of PTM. Disulfide bonding is another PTM can also be essential to correct protein formation. Extra disulfide bonds can lead to protein misfolding. *E. coli* lacks disulfide bonding in the cytoplasm, because of the negative redox potential and reducing environment therein (Rosano and Ceccarelli, 2017). To address this problem DMAP1 was expressed in a eukaryotic system however, the results were still inconclusive.

**Weakness with Detection Method**

There is a possibility that the experiment was successful however, restriction enzymes were not sensitive enough to detect low levels of methylation. The lowest concentration one can typically visualize on a gel, for one band, is around 10 ng of DNA. It is possible that DMAP1
only methylates at very low levels, beyond the detection of gel electrophoresis. Creating a situation where we cannot detect it with the naked eye through gel electrophoresis.

The use of restriction enzymes hinges on the presence of specific recognition sequences. Without that recognition site the enzymes cannot cleave. For example; with McrBC it would not cut if there is a pyrimidine followed by a methylated cytosine, as the enzyme only recognizes a purine followed by a methylated cytosine. While the use of HpaII/MspI does not work if it does not contain CCGG motif. In conclusion, it is possible that DMAP1 methylated a site that is not recognized by the enzymes used in this study.

**Alternative Methods**

Different methods could still be used to characterize DMAP1. For example, an *in vitro* analysis of the gene could be performed with S-Adenosyl methionine (SAM). To achieve this, a western blot should be performed to confirm the presence of the protein. DMAP1 would then be isolated with a 6xHis tag and then used in an *in vitro* analysis of with the addition of substrate. Those samples could then be used in another restriction digestion assay. In another example, one could grow an induced sample (where DMAP1 is expressed) in the presence of a radioactively labeled carbon on the methyl group of SAM. In this experiment, the gene would be induced, then extract the DNA, and then tested for the radioactive labeled methyl group addition to the DNA. One could also make use of high performance liquid chromatography which can be used to profile an entire genome (Armstrong et al, 2010). Over expression and knocking out the gene in *P. sojae* could also aid in understanding the function of the gene.
Conclusion

Currently there is no gene model for DNA methylation in oomycetes, due to the lack of any known DMNTs. Yet, pFAM and phylogenetic analyses showed that DMAP1 was present and conserved in the eighteen published oomycete genomes. DMAP1 was isolated from *P. sojae* and cloned into both *E. coli* and *S. cerevisiae*. The transformants containing DMAP1 proved to be non-toxic to its host cell. The restriction enzyme digestion assay used in this study failed to detect DNA methylation in both *E. coli* mutant and *S. cerevisiae*. 
REFERENCES


Protein Expression and Purification Core Facility - Cloning - Choice of Expression Systems.

(n.d) EMBL - European Molecular Biology Laboratory.

SANT domain. (n.d) EMBL - European Molecular Biology Laboratory.


OhioLINK Electronic Theses and Dissertations Center. 10 Jun 2017.


doi:10.1270/jsbbs.61.511 [doi]


Molecular Plant Pathology, 8(1), 1-8. doi:10.1111/j.1364-3703.2006.00373.x [doi]


APPENDIX A: ALIGNMENT OF DMAP1
APPENDIX B: YEAST DNA EXTRACTION PROTOCOL

1. Spin down 2mL of overnight culture at 5000 RPM for 5mins and decant supernatant.

2. Resuspend in 800uL of lysis buffer (200mM LiAc, 50mM EDTA, 10mM Tris-HCl (pH 7.8), and 0.5% SDS)

1. Freeze sample in EtOH dry ice mixture until frozen (~1min)

2. Thaw in 85°C water bath until thawed (~30-45sec)

   a. Repeat steps 3-4 three more times

3. Add 2uL of RNase (10mg/mL) and incubate at 37°C for 60 mins

4. Add 800uL of Pheonol:Chloroform:Isoamyl and briefly vortex to form an emulsion. Then spin at 13,000 RPM for 10 mins.

   a. Transfer the upper aqueous layer into a new tube without disturbing the lower phase.

5. Place a 1:1 ratio of Chloroform:Isoamyl and briefly vortex to form an emulsion. Spin down at 13,000 RPM for 10 mins.

   a. Transfer the upper aqueous layer into a new tube without disturbing the lower phase.

6. Put 10uL of 3M sodium acetate then add 1mL of ice cold EtOH and place in -20°C for at least an hour (can be stored overnight as well).

7. Spin down at 13,000 RPM at 4°C for 5 mins.

8. Decant all EtOH and wash with 70% EtOH.

9. Spin down at 13,000 RPM for 5 mins and decant all liquid.

10. Dry at 37°C for ~15 mins (until there is no more EtOH)

11. Suspend in 50uL of dH2O or TE buffer.