Bioinformatic Analysis Of Solanaceae Chloroplast Genomes And Characterization Of An Arabidopsis Protein Disulfide Isomerase In Transgenic Tobacco Chloroplasts

2006

Justin James Grevich

University of Central Florida

Find similar works at: https://stars.library.ucf.edu/etd

University of Central Florida Libraries http://library.ucf.edu

Part of the Microbiology Commons, and the Molecular Biology Commons

STARS Citation

Grevich, Justin James, "Bioinformatic Analysis Of Solanaceae Chloroplast Genomes And Characterization Of An Arabidopsis Protein Disulfide Isomerase In Transgenic Tobacco Chloroplasts" (2006). Electronic Theses and Dissertations. 6102.

https://stars.library.ucf.edu/etd/6102

This Masters Thesis (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of STARS. For more information, please contact lee.dotson@ucf.edu.
BIOINFORMATIC ANALYSIS OF SOLANACEAE CHLOROPLAST GENOMES AND CHARACTERIZATION OF AN ARABIDOPSIS PROTEIN DISULFIDE ISOMERASE IN TRANSGENIC TOBACCO CHLOROPLASTS

by

JUSTIN JAMES GREVICH
B.S. University of Central Florida, 2002

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Department of Molecular Biology and Microbiology in the Burnett College of Biomedical Sciences at the University of Central Florida Orlando, Florida

Spring Term
2006
ABSTRACT

Throughout history, traditional plant breeding has been used to provide resistance to pests, disease and other forms of environmental stress, as well as to increase yield and improve upon quality and processing attributes. Over the last decade, the advancement in sequencing technology and bioinformatic analysis has unleashed a wealth of knowledge about chloroplast genetic organization and evolution. The lack of complete plastid genome sequences is one of the major limitations in advancing plastid genetic engineering to other useful crops. This is due to the fact that plastid genome sequences are essential for the identification of endogenous regulatory sequences and optimal sites for homologous recombination. Analysis of four Solanaceae genomes revealed significant genetic modifications in both coding and non-coding regions. Repeat analysis with Reputer revealed 33 to 45 direct and inverted repeats ≥ 30bp with at least 90% homology. All but five of the 42 repeats shared among all four genomes were located in the exact same genes or intergenic regions, suggesting a functional role. Intergenic analysis found four regions that are 100 percent identical in all four Solanaceae genomes. Such highly conserved intergenic regions are ideal targets for multi-species transformation cassettes.

Protein disulfide isomerases (PDI) are a family of proteins known to function as molecular chaperones and aid in the formation of disulfide bonds during protein folding. They contain at least one thioredoxin domain used for the formation, isomerization, and reduction/oxidation of disulfide bonds. Bioinformatic analysis identified 13 PDI-like (PDIL) proteins found in Arabidopsis that contain at least one thioredoxin domain. In addition to the above-mentioned characteristics, PDIs have been shown to be directly involved in the translational regulation of the psbA mRNA in response to light and could potentially increase the
efficiency of chloroplast engineering in plants. Human serum albumin (HSA) requires 17 disulfide bonds to be properly folded and is an ideal candidate for assessing the disulfide bond formation, protein folding, and other chaperone-like characteristics of PDIL proteins. Therefore, I have coexpressed HSA in order to further characterize an *Arabidopsis* PDIL protein, atPDIL5-4, and in particular, the redox control of the *psbA* 5’UTR. Interestingly, the polyclonal antibody used for identifying the PDIL protein cross-reacted and identified other proteins, but not the transgenic atPDIL5-4. Results of these investigations will be presented.
ACKNOWLEDGMENTS

I would like to thank Dr. Henry Daniell for being a role model who taught me the meaning of dedication and hard work and gave me my first professional experiences as a scientist.

I would like to thank Dr. Antonis Zervos for always being there like a close friend, and for the times when I have needed career and personal advice.

I would like to thank Dr. Roseanne White for being like my mother away from home as well as for giving me the opportunity to take the helm in the QBM lab where I established essential skills in both teaching and management.
TABLE OF CONTENTS

LIST OF FIGURES ....................................................................................................................... ix

LIST OF TABLES ........................................................................................................................ x

LIST OF ACRONYMS/ABBREVIATIONS .................................................................................. xii

INTRODUCTION .......................................................................................................................... 1

  Maternal Inheritance And Gene Containment ................................................................. 3

  Historical Aspects ............................................................................................................. 4

  Plastid Genetic Engineering ............................................................................................ 6

  Heterologous Multigene Operons In Transgenic Chloroplasts ........................................... 8

  Crop Species Stably Transformed Via The Plastid Genome ............................................. 9

  Engineered Agronomic Traits Via Plastid Genome .......................................................... 10

    Cytoplasmic Male Sterility ............................................................................................. 11

  Transgenic Plastids As Pharmaceutical Bioreactors .......................................................... 12

    Human Serum Albumin ............................................................................................... 14

    Human Somatotropin ................................................................................................. 15

    Antimicrobial Peptides ............................................................................................... 15

    Human Interferons .................................................................................................... 16

    Monoclonal Antibodies ............................................................................................. 17

  Transgenic Plastids As Vaccine Bioreactors ................................................................. 18

    Oral Delivery ............................................................................................................. 19

    Industrial Bioreactors ............................................................................................... 20
LIST OF FIGURES

Figure 1: Overview of all final constructs ................................................................. 30
Figure 2: Schematic representing of primers used for PCR confirmation ....................... 33
Figure 3: Genomic map of Tomato and Potato chloroplast genomes .......................... 37
Figure 4: Whole genome and regional output examples from Multipipmaker ............... 38
Figure 5: Histogram representing the number of repeated sequences using Reputer ..... 39
Figure 6: Alignment of ycf2 gene ............................................................................. 41
Figure 7: Alignment of the 5’ portion of the 16S ribosomal RNA ................................ 42
Figure 8: Histogram representing sequence divergence of variable intergenic regions .... 42
Figure 9: Histogram representing the sequence divergence of coding regions ............. 44
Figure 10: Partial digestion of pB-PDI clones with NcoI and XbaI ................................ 47
Figure 11: Restriction digestion analysis of both the full length and shortened psbA 5’UTR .... 48
Figure 12: XbaI digested colonies from pLDg10•PDI•Trps16 transformation .................. 49
Figure 13: 40x amount of DNA loaded with CopyCutter cells .................................... 49
Figure 14: Unusual high molecular weight artifact .................................................... 50
Figure 15: The same artifact in CopyCutter brochure ................................................ 50
Figure 16: Agarose gel showing the digested rolling circle amplification product ........... 51
Figure 17: E. coli westerns with HSA antibody ......................................................... 52
Figure 18: 3P-3M PCR analysis on the initial 30 shoots selected .................................. 53
Figure 19: Preliminary 3P-3M PCR results from 2° plant DNA extracts ..................... 54
Figure 20: 5P-2M PCR confirmation on 2° shoots ...................................................... 54
Figure 21: Plant protein western blot with anti-HSA antibody ................................................... 55

Figure 22: Light dark western blots using anti-atPDI antibody................................................... 56

Figure 23: Southern blot with chloroplast flanking probe. ........................................................ 568

Figure 24: Southern blot of all plant DNA extracts using flanking probe.................................... 57
LIST OF TABLES

Table 1: Various sites of transgene integration into the chloroplast genome................................. 7
Table 2: Expression of agronomic traits.......................................................................................... 11
Table 3: Expression of vaccine antigens and biopharmaceutical proteins ........................................ 18
Table 4: Biomaterials, enzymes and amino acids engineered via the plastid genome .................. 20
Table 5: Tobacco repeat blast hits. .................................................................................................. 40
Table 6: Intergenic spacer regions that are 100% identical............................................................ 43
Table 7: Comparisons of sequence divergence of Solanaceae chloroplast genes ....................... 45
# LIST OF ACRONYMS/ABBREVIATIONS

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDI</td>
<td>Protein Disulfide Isomerase</td>
</tr>
<tr>
<td>PDIL</td>
<td>Protein Disulfide Isomerase Like</td>
</tr>
<tr>
<td>PABP</td>
<td>Poly-A binding protein</td>
</tr>
<tr>
<td>LSC</td>
<td>Large single copy region</td>
</tr>
<tr>
<td>SSC</td>
<td>Small single copy region</td>
</tr>
<tr>
<td>TRX</td>
<td>thioredoxin</td>
</tr>
<tr>
<td>Ct</td>
<td>Chloroplast</td>
</tr>
<tr>
<td>HSA</td>
<td>Human Serum Albumin</td>
</tr>
<tr>
<td>kb</td>
<td>Kilobase</td>
</tr>
<tr>
<td>CTB</td>
<td>Cholera toxin beta subunit</td>
</tr>
<tr>
<td>IR</td>
<td>Inverted repeat</td>
</tr>
<tr>
<td>Bt</td>
<td><em>Bacillus thuringiensis</em></td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated region</td>
</tr>
<tr>
<td>GUS</td>
<td>β-glucuronidase</td>
</tr>
<tr>
<td>INF</td>
<td>Interferon</td>
</tr>
<tr>
<td>EMC</td>
<td>Encephalomyocarditis</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Real-Time Polymerase Chain Reaction</td>
</tr>
<tr>
<td>HSV</td>
<td>Herpes Simplex Virus</td>
</tr>
<tr>
<td>mAbs</td>
<td>Monoclonal Antibodies</td>
</tr>
<tr>
<td>ori</td>
<td>Origin of replication</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>CPV</td>
<td>Canine Parvovirus</td>
</tr>
<tr>
<td>CPL</td>
<td>Chorismate pyruvate lyase</td>
</tr>
<tr>
<td>pHBA</td>
<td>p-hydroxybenzoic acid</td>
</tr>
<tr>
<td>EPSPS</td>
<td>enolpyruvylshikimate-3-phosphate synthase</td>
</tr>
</tbody>
</table>
INTRODUCTION

Plants have been on the earth for over a billion years, ever since the capture of a cyanobacterium by a eukaryotic, mitochondria-possessing cell. Over time the relationship became more intimate and the endosymbiont lost its autonomy and most of its genome (Martin et al., 2002), giving rise to a group of cell organelles collectively referred to as plastids. The majority of more than three thousand genes of the cyanobacterial endosymbiont were either lost or transferred to the nucleus. These organelles have developed a variety of differentiated forms, such as the photosynthetic chloroplasts, starch-storing amyloplasts, and colorful chromoplasts of fruits and flowers. Surprisingly, plastid-like organelles (apicoplasts) have even been found in protozoa. Although the genomes are reduced in complexity, their similarity to chloroplasts makes it difficult to differentiate unicellular animals from unicellular plants (Wilson, 2002). Among higher plants, plastid genomes (plastomes) are highly conserved in both gene content and order (Daniell et al., 2006).

In contrast to plant mitochondria, genome organization and coding capacity in plastids display relatively little interspecific variation in coding regions. The chloroplast genome is a circular, self-replicating molecule that can vary in size from 120 to 220 kb depending upon species, and exists predominantly as a monomer along with other multimeric forms (Fig. 3) (Lilly et al., 2001). Chloroplasts usually consist of 120-130 densely packed genes contained within two inverted repeat regions (IRs), as well as, the large and small single copy regions (LSC and SSC) (Lilly et al., 2001). Chloroplast genes are organized into two categories, photosynthesis-related genes and genetic system genes (ie. rRNAs, tRNAs, ribosomal proteins
and RNA polymerase subunits). Forty to fifty percent of higher plant chloroplast genomes are made up of non-coding spacer regions, introns, and regulatory sequences (Saski et al., 2005). Gene arrangement and expression exhibit a unique mix of both eukaryotic and prokaryotic features. Plastid genes can be organized into operons yet may contain eukaryotic-like introns. Regulation of gene expression can occur during transcription, translation and post-translational processes.

Plastids are attractive candidates for the genetic engineering of plants for many reasons. A typical plant cell contains approximately 100 chloroplasts, each with about 100 identical genomes (Bendich, 1987). Therefore, a single gene may be represented up to 10,000 times within the plant cell. The copy number of genes encoded within the inverted repeats can reach up to 20,000 copies, often contributing to high levels of transgene expression. Transgenic proteins can accumulate to as high as 46% of the plant’s total soluble protein (tsp) (Decosa et al., 2001).

Compared to nuclear engineered systems, plastid engineering lacks epigenic effects, such as gene silencing (Lee et al., 2003; Decosa et al., 2001), positional effects (Daniell et al., 2001a), and has the ease of transgene stacking in operons (DeCosa et al., 2001; Daniell and Dhingra, 2002), all of which allow for a vastly more efficient recombinant protein expression system. Transgenes are integrated via homologous recombination within the plastid genome and facilitated by a RecA-type system (Cerutti et al., 1992), unlike random integration that occurs in nuclear transformations. Plastid transformation vectors use this feature and are designed to contain homologous flanking sequences on either side of the transgenes, allowing for site-specific integration (Daniell 1993, 1997; Kumar and Daniell, 2004; Daniell et al., 2005b). Despite transcripts levels 169 times higher than nuclear transgenic plants and accumulation of foreign proteins of over 46% tsp (DeCosa et al., 2001), there is no gene silencing in chloroplast
transgenic plants (Lee et al., 2003). The accumulation of large amounts of transcripts is a common phenomenon seen in light regulated translation of chloroplast proteins. Plastid engineering significantly increases the containment of transgenes since plastids are maternally inherited in most crops and not spread through pollen (Daniell, 2002; Hagemann, 2004).

Chloroplasts can also provide an ideal place to accumulate proteins or their biosynthetic products that may otherwise be harmful in the cytoplasm (Bogorad, 2000). For example, cholera toxin B subunit (CTB) caused severe reduction in growth when expressed via the nuclear genome. Chloroplast transgenic plants had a normal phenotype despite expression of CTB at levels 500-4000 fold higher than the nuclear transgenic plants (Daniell et al., 2001a; Molina et al., 2004). Similarly, the pharmaceutical preservative, trehalose, is very toxic when accumulated in the cytoplasm, yet nontoxic when compartmentalized within plastids (Lee et al., 2003), and xylanase, an enzyme important in many industrial applications, did not cause cell wall degradation in plastid engineered plants (Leelavathi et al., 2003). Additionally, all chloroplast transgenic lines express similar levels of foreign proteins, within the range of physiological variations (Daniell et al., 2001a).

**Maternal Inheritance And Gene Containment**

In most angiosperm plant species, plastid genes are inherited in a uniparental and strictly maternal fashion (Birky, 1995; Mogensen, 1996; Zhang et al., 2003; Hagemann, 2004). For example, seedlings grown on selective media remain sensitive to antibiotics when pollen from chloroplast transgenic lines containing an antibiotic selectable marker gene is used to pollinate wild-type plants, whereas, pollen from wild-type plants used to pollinate chloroplast transgenic
plants results in progeny that inherit the antibiotic resistance. Although these plants contain pollen with metabolically active plastids, the plastid DNA itself is lost during the process of pollen maturation, and is therefore not transmitted to the next generation (Nagata et al., 1999; Daniell, 1999, 2000). The lack of plastid DNA in pollen thereby minimizes the possibility of spreading transgenes to undesired weeds or crops (Daniell et al., 1998; Scott and Wilkinson, 1999; Daniell, 2002) and reduces the potential toxicity of transgenic pollen to non-target insects (DeCosa et al., 2001).

Employing a failsafe mechanism such as male sterility can significantly enhance gene containment (Ruiz and Daniell, 2005). Using the plastid genome as the site of engineering already provides transgene containment through maternal inheritance, however, when combined with male sterility, plastic engineering provides multiple levels of transgene containment.

**Historical Aspects**

Chloroplast engineering was first considered during the 80’s after discovering how to isolate and convert intact chloroplasts into protoplasts (Daniell and McFadden, 1987; Daniell et al., 2002; Devine and Daniell, 2004). Early studies focused on developing efficient protein expression systems for foreign genes (Daniell and McFadden, 1987). Biolistic devices simplified plastid transformation and made it possible to transform plastids without the need to first isolate them (Klein et al., 1987). The first successful plastid complementation was accomplished using the unicellular green alga, *Chlamydomonas reinhardtii* by Boynton et al. (1988). This experiment replaced the *atpB* gene in mutant, *atpB*-deficient, *Chlamydomonas*. The introduced gene corrected the deletion mutant phenotype and was stably integrated via the
chloroplast genome. The first attempt with a foreign gene, *uidA*, was stably integrated and transcribed, however, no translated product could be found (Blowers *et al*., 1989).

Autonomously replicating chloroplast vectors were found to express foreign genes at higher levels and for extended periods of time compared to vectors without an *ori*. These vectors were used in the first transformation of plastids using cultured cells (Daniell *et al*., 1990) and were later extended to calli, somatic embryos and wheat leaves (Daniell *et al*., 1991). The first selectable marker, the *aadA* gene, was integrated into *Chlamydomonas* conferring spectinomycin and streptinomycin resistance (Goldschmidt-Clermont, 1991), the most common selectable marker currently used in higher plants transformation via the chloroplast genome. Currently, plastid genetic engineering technology has been applied to a variety of crops to confer agronomic traits (Table 1), engineer therapeutic proteins (Table 2), and for many other useful applications such as the production of biomaterials, enzymes, and amino acids (Table 3). A milestone in chloroplast engineering, the expression of *Bacillus thuringiensis* (Bt) Cry2Aa2, resulted in high levels of foreign protein accumulation (46.1% tsp) in transgenic chloroplasts. This was also the first example of the expression of a complete bacterial operon and lack of the transgene product in pollen (DeCosa *et al*., 2001).

Plastid transformation has also aided in the study of plastid biogenesis and function in such areas as: plastid DNA replication origins, RNA editing elements, promoter elements, RNA stability determinants, intron maturases, translation elements, protein import machinery, proteolysis, transgene movement and evolution, and transcription and translation of polycistrons (Daniell *et al*., 2004b).
Plastid Genetic Engineering

As stated earlier, chloroplast proteases are extremely effective and have the ability to drastically reduce the amount of foreign proteins. When under the control of the light regulated psbA untranslated regions (UTRs), there was a three- to ten-fold reduction in HSA protein expression in leaves harvested in the dark (Fernández-San Millán et al., 2003). Because of chloroplast proteases, engineered proteins must often be protected for efficient expression of foreign proteins. The use of the CRY chaperone (encoded by the orf2 gene) to fold the insecticidal protein, Cry2Aa2, into cuboidal crystals protected the foreign proteins from degradation, thereby increasing protein accumulation over 128-fold (from 0.36% to 46.1% tsp, DeCosa et al., 2001). Similarly, when the HSA coding sequence was regulated by the chloroplast psbA 5’ and 3’ UTRs, protein expression increased 500-fold and resulted in the formation of protective inclusion bodies (Fernández-San Millán et al., 2003).

Several studies on transgenic chloroplasts have not found a correlation between increased transcript abundance and translation efficiency. Chloroplast derived RbcS transcripts were measured to be 165-fold and 143-fold more than nuclear control plants when the transgene was regulated by the psbA 5’UTR or the promoterless gene 10 UTR (g10) respectively. Although this was the first example of successful functional rubisco in transgenic chloroplast plants, the g10 transgenic lines performed poorly (Dhingra et al., 2004). The lack of correlation between increased transcript levels and translation efficiency suggests that transcript abundance is of less importance than protein stability in transgenic chloroplasts.

To date, the most preferred site of integration is between the trnl-trnA genes of the rrn operon located within the inverted repeat regions of the chloroplast genome. This site has been
used for 7 out of 9 biopharmaceutical proteins, 5 out of 7 vaccine antigens and 5 out of the 6 agronomic traits engineered via the chloroplast genome so far (Table 4) (Dhingra and Daniell, In Press). The foreign gene expression levels obtained from genes integrated at this site are amongst the highest ever reported. Compared to other sites, the trnI-trnA region appears to be unique and allows for highly efficient transgene integration and expression. The presence of the chloroplast ori in one of the flanking regions (Kunnimalaiyaan et al., 1997; Lugo et al., 2004) may act as an additional selective force facilitating replication of foreign vectors within chloroplasts (Daniell et al., 1990), enhance the probability of transgene integration, and possibly help achieve homoplasy during the first round of selection (Guda et al., 2000). This is further illustrated by the first successful engineering of Rubisco accomplished by integrating the RbcS gene at this site (Dhingra et al., 2004), whereas, all other attempts of rubisco engineering at other plastid integration sites were only partially successful. Various transgenes integrated into the transcriptionally active trnI-trnA spacer region or other read-through or silent intergenic spacer regions are listed in Table 1.

Table 1: Various sites of transgene integration into the chloroplast genome (Grevich, J.J., and Daniell, H., 2005)

<table>
<thead>
<tr>
<th>Site of integration</th>
<th>Transcription Status</th>
<th>Integrated Transgene</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>trnI/trnA</td>
<td>Active</td>
<td>aadA/aroA&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Daniell et al., 1998.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aadA/cytB&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Daniell et al., 2001a.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aadA/cytB-CPf&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Molina et al., 2004.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aadA/gfp-CPV&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>aadA/pag&lt;sup&gt;e&lt;/sup&gt;</td>
<td>Daniell et al., 2004.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aadA/CaF1-Lcrv&lt;sup&gt;f&lt;/sup&gt;</td>
<td>Singleton et al., 2003.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aadA/EG121&lt;sup&gt;g&lt;/sup&gt;</td>
<td>DeGray et al., 2001.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aadA/HGF-1&lt;sup&gt;h&lt;/sup&gt;</td>
<td>Ruiz, 2002.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aadA/INFa&lt;sup&gt;i&lt;/sup&gt;</td>
<td>Torres, 2002.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aadA/INF-a2b&lt;sup&gt;j&lt;/sup&gt;</td>
<td>Falconer, 2003.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aadA/HSA&lt;sup&gt;k&lt;/sup&gt;</td>
<td>Fernandez San-Milan et al., 2003.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aadA/Guy’s 13&lt;sup&gt;l&lt;/sup&gt;</td>
<td>Daniell et al., 2001c.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aadA/Cry2Aa2&lt;sup&gt;m&lt;/sup&gt;</td>
<td>Kota et al., 1999.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aadA/Cry2Aa2-operand&lt;sup&gt;n&lt;/sup&gt;</td>
<td>DeCosa et al., 2001.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aadA/p&lt;sup&gt;n&lt;/sup&gt;</td>
<td>Lee et al., 2003.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aadA/merA-merB&lt;sup&gt;p&lt;/sup&gt;</td>
<td>Ruiz et al., 2003.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aadA/hadh&lt;sup&gt;q&lt;/sup&gt;</td>
<td>Daniell et al., 2000.</td>
</tr>
</tbody>
</table>
Like their bacterial ancestors, the plastid genes of higher plants are mostly organized into operons of both related and unrelated function (Sugita and Sugiura, 1996). Plastid genes are transcribed into polycistronic RNA and may be further processed via alternative splicing, editing or intercistronic processing (Barkan, 1988; Rochaix, 1996; Barkan and Goldschmidt-Clermont, 2000; Monde et al., 2000). In most cases, primary transcripts require some form of processing in order to be functional, representing an additional level of control over plastid gene expression (Marchfelder and Binder, 2004; Barkan, 2004). In the study of five foreign operons, Quesada et al. (2005) discovered that various multigene operons generated polycistrons as the most abundant transcript, which, in most lines, was never processed into monocistrons. The Cry2Aa2 operon, the highest expressed foreign protein reported in the transgenic literature (up to 46.1% of total leaf protein) was also from an unprocessed polycistron. This gives evidence that foreign multigene operons do not need to be processed into monocistrons for efficient translation. In the case of bacterial operons, such as the native Cry2Aa2 operon, secondary structures may play a
critical role in transcript processing. These transcripts exhibited stability even with the removal of the 3’ UTRs, which is necessary for stability of native transcripts (Quesada et al., 2005). Continued research addressing these areas is essential for further advancement in plastid engineering.

**Crop Species Stably Transformed Via The Plastid Genome**

Tobacco is significantly the most commonly used plant for plastid transformation, mostly because of its vast utility as a model organism in plastid engineering, as well as plant science in general. This has led to more transgenes expressed in tobacco genomes than all other plant species combined. Tobacco turns out to be a very beneficial plant for plastid engineering. A single tobacco plant is capable of producing a million seeds, and one acre of tobacco can produce more than 40 metric tons of leaves per year (Cramer et al., 1999). Tobacco has been estimated to be more than 50-times less costly for recombinant protein production when compared to standard large-scale *E. coli* fermentation setups (Kusnadi et al., 1997). Additionally, tobacco has significantly less ability to contaminate other plants, it can produce many proteins that bacteria cannot, and the methods for tobacco plastid transformation have been streamlined, resulting in a very efficient biofactory. Harvesting leaves before flowering offers nearly complete transgene containment and can almost be made even more certain when combined with maternal inheritance and/or engineered male sterility. Even though plastid transformation is efficient in tobacco, this is not always so for many other crops. This is most likely due to the lack of genomic sequences available and the need to optimize species-specific transformation protocols. Despite these disadvantages, stable transformation has been achieved in carrot (Kumar et al.,
2004a), cotton (Kumar et al., 2004), Petunia (Zubko et al., 2004), potato (Sidorov et al., 1999), soybean (Dufourmantel, N., et al., 2004), and tomato (Ruf et al., 2001). In addition to providing an excellent model for edible vaccines, transgenic carrot plants engineered for salt tolerance were able to withstand salt concentrations that only halophytes can tolerate (Kumar et al., 2004a).

**Engineered Agronomic Traits Via Plastid Genome**

Agronomic chemicals are of great economic importance, which is exemplified by the $15 billion spent worldwide on insecticides and fungicides alone (Kiely et al., 2004). Genetic engineering can alleviate much of this annual cost or even eliminate it, providing a much more environmentally friendly solution. Plastids offer an ideal place for engineering these traits for many reasons such as high levels of expression, ease of transformation, and gene containment via maternal inheritance. Many important agronomic traits have already been engineered via the plastid genome such as, herbicide resistance, insect resistance, and tolerance to drought and salt (Table I). Recently, plastids have even been engineered for the remediation of toxic chemicals in the environment, a process known as phytoremediation (Ruiz et al., 2003). Many crops have already been genetically engineered via the nuclear genome, however, these plants have drawbacks inherent to their method of transformation. More than half of the soybean planted worldwide is genetically modified, yet restrictions have to be placed on the location of crops because of the risk of transgenes spreading to wild relatives. Additionally, lower expression levels of Bt toxins engineered via the nuclear genome pose an increased risk of developing Bt resistant pests.
Table 2: Engineering of agronomic traits via the plastid genome (Grevich, J.J., and Daniell, 2005)

<table>
<thead>
<tr>
<th>Agronomic trait</th>
<th>Gene</th>
<th>Site of integration</th>
<th>Promoter</th>
<th>5'/3' UTRs</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insect resistance</td>
<td>Cry1A(c)</td>
<td>trnV/rps12/7</td>
<td>Prrn</td>
<td>rbcL / Trps16</td>
<td>McBride et al., 1995</td>
</tr>
<tr>
<td>Herbicide resistance</td>
<td>aroA (petunia)</td>
<td>rbcL/accD</td>
<td>Prrn</td>
<td>ggagg / TpsbA</td>
<td>Daniell et al., 1998</td>
</tr>
<tr>
<td>Insect resistance</td>
<td>Cry2Aa2</td>
<td>rbcL/accD</td>
<td>Prrn</td>
<td>ggagg (native) / TpsbA</td>
<td>Kota et al., 1999.</td>
</tr>
<tr>
<td>Insect resistance</td>
<td>Cry2Aa2 operon</td>
<td>trnI/trnA</td>
<td>Prrn</td>
<td>native 5'UTRs / TpsbA</td>
<td>De Cosa et al., 2001.</td>
</tr>
<tr>
<td>Drought tolerance</td>
<td>tps</td>
<td>trnI/trnA</td>
<td>Prrn</td>
<td>ggagg / TpsbA</td>
<td>Lee et al., 2003.</td>
</tr>
<tr>
<td>Phytoremediation</td>
<td>merA/merB</td>
<td>trnI/trnA</td>
<td>Prrn</td>
<td>ggagg^a / TpsbA</td>
<td>Ruiz et al., 2003.</td>
</tr>
<tr>
<td>Salt tolerance</td>
<td>badh</td>
<td>trnI/trnA</td>
<td>Prrn-F</td>
<td>ggagg/rps16</td>
<td>Kumar et al., 2004.</td>
</tr>
</tbody>
</table>

^a and b refer to genes and their respective regulatory sequences

Cytoplasmic Male Sterility

Naturally occurring cytoplasmic male sterility has been reported for maize, oilseed rape, rice and Beta beets (Kriete et al., 1996), and has been documented for over 100 years. However, this often desired trait is not available for most crops used in agriculture. Despite the creation of over 175 male-sterile nuclear mutant plants, it is not possible to produce populations of 100% male plants because of the lack of homozygous breeding lines (Kaul, 1988). Male sterility can be conferred via multiple mechanisms, although most affect the tapetum and pollen development.

In Petunia hybrida, antisense technology has been used to deplete flavonoid pigments in the anthers, thereby arresting pollen maturation (van der Meer et al., 1992). Similarly, antisense expression was used to inhibit the mitochondrial pyruvate dehydrogenase, causing perturbations in the tapetum that lead to an alternate form of male sterility (Yui et al., 2003). Additionally, the nuclear expression of the barnase gene caused degradation of the tapetum and lack of pollen formation (Kriete et al., 1996). More recently, pollen development was arrested in Arabidopsis by the disruption of AtGPAT1, a gene involved in the initial step of glycerolipid biosynthesis (Zheng et al., 2003).

The possibility of the production of transgenic seeds that spread transgenic traits to non-transgenic plants is a major drawback of the current forms of male sterility systems. Nuclear
encoded male sterility systems are also more vulnerable because of genetic segregation, which will eventually dilute the male sterility trait. In addition, many of these systems incur drawbacks such as the interference with general development or metabolism and are often times restricted to certain species. The overexpression of the \textit{phaA} gene, encoding β-ketothiolase in tobacco chloroplasts addresses some of these concerns (Ruiz and Daniell, 2005). This resulted in transgenic plants with normal phenotype except for the lack of pollen. Further examination using SEM revealed a collapsed morphology of the pollen grains. Additionally, through floral developmental studies, it was found that transgenic lines underwent an accelerated pattern of anther development, resulting in aberrant tissue patterns during maturation. Surprisingly, this phenotype can be reversed by continuous lumination, resulting in viable pollen and subsequent seeds. This research provides a new tool for efficient transgene containment of both nuclear and organelle genomes, while also offering a convenient mechanism for F1 hybrid seed production.

\textbf{Transgenic Plastids As Pharmaceutical Bioreactors}

Genetic engineering has revolutionized the use of therapeutically valuable proteins, many of which could not be made until the technology became widespread. It is now possibly to take genes from practically any other organism and express them in another. As a biofactory, an organism must be capable of producing safe, biologically active material at the lowest cost. Mammalian cells are ideal for the production of human therapeutic proteins because of their ability to generate human proteins, and with the use of stem cells, it is possible to produce a specific patient’s exact protein. However, these cells are costly to culture and have no cost-efficient option for large-scale setup when compared to other bioreactors. \textit{E. coli} grown in large
fermenters is the most common form of a bioreactor used in industrial applications. The use of bacteria also introduces several disadvantages, such as the lack of glycosylation, proper folding, and formation of disulfide bonds. Plant systems are more economical than fermentation systems, transgenic animals, and mammalian cell cultures. Although the nuclear expression of foreign genes is often disappointingly low, several plant-based production platforms have been developed based on protein expression from the nuclear genome and viral vectors in leaf, seed, tuber, and tissue cultured cells (Ma, 2000; Giddings et al., 2000; Stoger et al., 2002).

The technology for harvesting and processing plants and plant products has already been developed and provides a time-tested system that can be used for large-scale recombinant protein production in plants (Daniell et al., 2001c). Plastid engineering typically results in remarkably high-levels of expression and speed in designing a transgenic model when compared to other bioreactors. Plastids have been shown to efficiently express large proteins, as seen by the accumulation of the 133 kDa Cry1A protoxins at levels of 3-5% tsp in leaves (McBride et al., 1995) as well as small, with the twenty amino acid peptide, magainin (DeGray et al., 2001). Plastids have also been used for the expression of several therapeutic proteins including human serum albumin (11.1%, Fernandez San-Milan et al., 2003), somatotropin (7%, Staub et al., 2000), interferon-alpha (19%, Daniell et al., 2004a), interferon-gamma (6%, Leelavathi and Reddy, 2003), and an anti-microbial peptide (21.5%, DeGray et al., 2001) (Table 3). In plastids there is an ideal mix of both prokaryotic and eukaryotic features that allow for the expression of multiple genes in operons and the ability to assemble complex multi-subunit proteins with proper disulfide bonds (Daniell et al., 2004a). Many of these features account for the cost of in vitro processing in other bioreactors, which can account for up to 30% the production cost and 70% of the setup cost (Petridis et al., 1995). Functionality of chloroplast derived vaccine antigens and
therapeutic proteins has been demonstrated by several assays including the macrophage lysis assay (Watson et al., 2004), GM1-ganglioside binding assay (Daniell et al., 2001a; Molina et al., 2004), protection of HeLA cells or human lung carcinoma cells against encephalomyocarditis virus, and many others (Daniell et al., 2004b, Koya et al., 2005). Purification of human proinsulin incorporated novel purification strategies that did not require expensive chromatography (Daniell et al., 2004a). Transgenic chloroplasts can produce functional plant, animal, and human proteins in an environmentally friendly manner. Plant derived biopharmaceuticals are less expensive to produce and store, easy to scale up for mass production, and safer than those derived from animals (Daniell et al., 2001c). Plastid engineering provides a much cheaper alternative for biopharmaceutical production, which may aid in the widespread availability of biopharmaceuticals.

**Human Serum Albumin**

Human serum albumin (HSA) is the most abundant protein found in blood serum and is one of the most widely used intravenous proteins in many human therapies (Peters, 1995). It is most often used for the replacement of blood during incidents of trauma and treatment, as well as in cirrhotic or hepatic illnesses and many other therapies (Peters, 1995). Although over 500 tons of HSA is used annually (Fernández-San Millán et al., 2003), the current demand is hardly met and limited by the availability of human blood extractions. In addition to high costs, HSA from human blood extractions has the potential of transmitting diseases as well as other blood-derived products.

Attempts using nuclear transformation for the expression of HSA have resulted in only 0.2% tsp in tubers and 0.02% tsp in leaves (Farran et al., 2002), far below the levels needed for cost effective production. It was hypothesized that the AT-rich sequence of HSA was
responsible for the poor translation in the eukaryotic cells of transgenic nuclear plants, which is ideal for chloroplast expression. Additionally, HSA requires 17 disulfide bonds in order to be correctly folded, and therefore, protected from protease activity.

The HSA gene was expressed under the control of the psbA regulatory sequences, in the chloroplast genome between the trnA and trnI genes of the chloroplast rrn operon (Fernández-San Millán et al., 2003). HSA was quantified using ELISA and found to be up to 11.1% tsp after 50 hours of continuous illumination. The functionality of the recombinant HSA has yet to be determined due to the complexity of the studies required to measure these properties (Watanabe et al., 2001). Formation of inclusion bodies not only protected HSA from proteolytic degradation but also offered an extraordinary advantage for purification by simple centrifugation.

**Human Somatotropin**

Human somatotropin (hST) is used to treat hypopituitary dwarfism in children, Turner’s syndrome, chronic renal failure, and HIV wasting syndrome. This expression cassette resulted in high levels of hST expression, reaching 7% tsp. Additionally, it was shown to contain proper disulfide bond formation and was identical to the native hST protein.

**Antimicrobial Peptides**

With the increasing amounts of antibiotic-resistant pathogens, effective alternatives to the current selection of antibiotics are needed. MSI-99 is a synthetic lytic peptide that has been expressed via the tobacco chloroplast. Because bacterial membranes are highly conserved, it is unlikely that bacteria will develop resistance to the mechanism of antimicrobial peptides. Cell extracts from transgenic tobacco leaves containing MSI-99 were used to determine its effectiveness towards *Pseudomonas aeruginosa*, a multidrug-resistant bacterium that affects
plants, animals and humans. Transgenic plants produced enough MSI-99 that cell extracts were able to cause 96% inhibition of growth of *P. aeruginosa*.

**Human Interferons**

Human interferon alphas are cytokines of the immune system that are able to cause cells to express anti-viral proteins. They are potent enhancers of the immune system and are used in various clinical treatments. The costs for such treatments are approximately $26,000 a year for hairy cell leukemia and up to $2500 per 2-week treatment for patients with West Nile virus (Devine and Daniell, 2004). These high costs are a direct result of the *E. coli* recombinant expression system used to manufacture the IFNα2b, mostly because of the required *in vitro* processing and purification. Also, when IFNα2b is injected it can bring about severe side affects, 20% of the patients produce anti-IFNα antibodies when IFNα2b combines with the HSA in their blood (Devine and Daniell, 2004). IFNα2b can be orally administered and has been proven to be therapeutically effective when delivered by these means. The expression of IFNα2b in plant plastids does not require *in vitro* processing (Falconer, 2002) and can drastically reduce costs associated with production in addition to providing an alternative means of administration. Recombinant IFNα2b was expressed at levels up to 18.8% tsp and found to be fully functional in its ability to protect HeLa cells against the cytopathic effects of encephalomyocarditis (EMC) virus and was found to be as effective as the commercial product (Daniell *et al.*, 2004a; Daniell *et al.*, 2005a).
Monoclonal Antibodies

Antibodies are a branch of the immune system that protects the body from toxic substances and invading pathogens. Monoclonal antibodies (mAbs) have been developed from hybridomas for many years. The variety of applications for the use of mAbs in a therapeutic setting is consistently growing, however, their production is typically much too expensive to be practical for most applications. Guy’s 13 monoclonal antibody is specific for a surface antigen of *Streptococcus mutans*, a bacterium that most notably causes dental caries. This antibody has been expressed and extracted from tobacco chloroplasts and was found to be correctly folded and identified via western blots, however this model did not express at levels high enough for it to be a cost effective therapeutic (Daniell *et al.*, 2004b). The expression of the large single-chain (lsc) antibody directed against herpes simplex virus glycoprotein D has been achieved in the chloroplast genome of *C. reinhardtii* (Mayfield *et al.*, 2002). The lsc gene was codon-optimized and expressed at 80 times the level of native genes (Franklin *et al.*, 2002). Furthermore, the recombinant protein was found to be stably integrated and shown to be functional in assays against HSV8 proteins. Both studies provide examples of how plastid genomes can be used for the production of monoclonal antibodies in plant-based bioreactors. The raw materials for these systems is over 3000-fold less than of those with mammalian cells (Dove, 2002) and provide a reasonable alternative. Additionally, it has been projected that the mAbs production capacity will eventually fall substantially short of demand, highlighting the need for an alternative production solution (Motmans and Bouche, 2000).

These results demonstrate that plastids have the proper machinery for folding eukaryotic proteins that require disulfide bonds, most likely, through the use of the chloroplast protein
disulfide isomerase. Unlike the most common bacterial bioreactors, plastid bioreactors most often do not incur expensive post-translational in vitro processing that can substantially reduce the costs of biopharmaceutical production.

**Table 3:** Expression of vaccine antigens and biopharmaceutical proteins via the plastid genome (Grevich, J.J., and Daniell, H., 2005).

<table>
<thead>
<tr>
<th>Therapeutic proteins</th>
<th>Gene</th>
<th>Site of integration</th>
<th>Promoter</th>
<th>5'UTRs</th>
<th>% tsp expression</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elastin derived polymer</td>
<td>EG121</td>
<td>trnI/trnA</td>
<td>Prm</td>
<td>T7gene10/TpsbA</td>
<td>none detected</td>
<td>Guda et al., 2000.</td>
</tr>
<tr>
<td>Human somatotropin</td>
<td>hST</td>
<td>trnV/rps12/7</td>
<td>Prm(^a)</td>
<td>T7gene10/TpsbA</td>
<td>7.0 %(^a)</td>
<td>Staub et al., 2000.</td>
</tr>
<tr>
<td>Cholera toxin</td>
<td>CtxB</td>
<td>trnI/trnA</td>
<td>PpsbA(^b)</td>
<td>TpsbA/TpsbA</td>
<td>1.0%(^b)</td>
<td>Daniell et al., 2001a.</td>
</tr>
<tr>
<td>Insulin-like growth factor</td>
<td>IGF-1</td>
<td>trnI/trnA</td>
<td>Prm</td>
<td>T7gene10/TpsbA</td>
<td>33%</td>
<td>Daniell et al., 2005a.</td>
</tr>
<tr>
<td>Interferon alpha 5</td>
<td>INFα5</td>
<td>trnI/trnA</td>
<td>Prm</td>
<td>TpsbA/TpsbA</td>
<td>NT</td>
<td>Torres, 2002.</td>
</tr>
<tr>
<td>Interferon alpha 2b</td>
<td>INFα2B</td>
<td>trnI/trnA</td>
<td>Prm(^a)</td>
<td>TpsbA/TpsbA</td>
<td>19%</td>
<td>Daniell et al., 2004a.</td>
</tr>
<tr>
<td>Human serum albumin</td>
<td>hsa</td>
<td>trnI/trnA</td>
<td>Prm(^a)</td>
<td>TpsbA/TpsbA</td>
<td>0.02%(^a)</td>
<td>Fernandez et al., 2003.</td>
</tr>
<tr>
<td>Monoclonal antibodies</td>
<td>Guy’s 13</td>
<td>trnI/trnA</td>
<td>Prm</td>
<td>TpsbA/TpsbA</td>
<td>NT</td>
<td>Daniell et al., 2004a.</td>
</tr>
<tr>
<td>Anthrax protective antigen</td>
<td>pag</td>
<td>trnI/trnA</td>
<td>Prm</td>
<td>TpsbA/TpsbA</td>
<td>4.5%</td>
<td>Chebola and Daniell, In Press.</td>
</tr>
<tr>
<td>Plague vaccine</td>
<td>CaF1-Lcr</td>
<td>V</td>
<td>trnI/trnA</td>
<td>PpsbA/TpsbA</td>
<td>4.6%</td>
<td>Singleton, 2003.</td>
</tr>
<tr>
<td>CPV VP2</td>
<td>CTB-2L21(^a)</td>
<td>TrnI/trnA</td>
<td>Prm(^a)</td>
<td>TpsbA/TpsbA</td>
<td>31.1%(^a)</td>
<td>Molina et al., 2004.</td>
</tr>
<tr>
<td>CPV VP2</td>
<td>GFP-2L21(^b)</td>
<td>TrnI/trnA</td>
<td>Prm(^b)</td>
<td>TpsbA/TpsbA</td>
<td>22.6%(^b)</td>
<td>Molina et al., 2004.</td>
</tr>
<tr>
<td>Rotavirus VP6</td>
<td>vp6</td>
<td>rbcL/accD</td>
<td>PpsbA(^b)</td>
<td>TpsbA/TpsbA</td>
<td>3%(^b)</td>
<td>Birch-Machin et al., 2004.</td>
</tr>
<tr>
<td>Tetanus toxin</td>
<td>TET C</td>
<td>TrnV/rps12/7</td>
<td>Prm(^a)</td>
<td>T7gene10/TpsbA</td>
<td>0.6%(^a)</td>
<td>Birch-Machin et al., 2004.</td>
</tr>
</tbody>
</table>

\(^a\), \(^b\), and \(^c\) refer to genes and their respective regulatory sequences and % tsp

NT: Not tested
tsp: total soluble protein

---

**Transgenic Plastids As Vaccine Bioreactors**

Subunit vaccines eliminate the risks of vaccines reverting to virulent forms and are also free of trace amounts of toxins and other molecules that can cause undesirable effects. The production of subunit vaccines and subsequent purification can be costly and is a limitation in preventing widespread use. Plastids offer a low-cost approach to producing subunit vaccines with two unique advantages, simple purification and oral delivery. Reducing the cost of vaccines...
is extremely important since many countries cannot afford them because of the high costs associated with production, storage and refrigerated transportation. Plant-derived subunit vaccines are functional and able to withstand pathogen challenge (Mason et al., 1996; Kapusta et al., 1999; Richter et al., 2000; Tacket et al., 2000; Kong et al., 2001; Koya et al., 2005). The use of protective antigens can add alternative methods for introducing subunit vaccines into the body in a manner inherent to each specific carrier (e.g. CTB as a transmucosal carrier to deliver other antigens).

**Oral Delivery**

In order to accomplish oral delivery of vaccines with plastids, large quantities of the recombinant protein must be accumulated in the edible parts of plants (Kumar et al. 2004a) and antibiotic resistance genes used for selection must be removed. The oral delivery of vaccine antigens against cholera (Daniell et al., 2001a), tetanus (Tregoning et al., 2003), anthrax (Watson et al., 2004), plague (Daniell et al., 2004a; Daniell et al., 2005a) and canine parvovirus (Molina et al., 2004), have been accomplished using plastid engineering of the chloroplasts (leaves) or non-green plastids (carrots) (Kumar et al., 2004a). The availability of antibiotic free selectable markers (Daniell et al., 2001d) or the ability to excise selectable marker genes as discussed above can alleviate public concern over antibiotic contamination in biopharmaceuticals. Oral delivery may also eliminate the need to purify therapeutic proteins from plant cells and vastly decrease shipping costs associated with products that need cold storage. All this may possibly allow access to modern vaccines and biopharmaceuticals for the vast majority of the world, besides what’s a few seeds to a multi-billion dollar pharmaceutical company.
Industrial Bioreactors

Plastid genetic engineering offers an ideal solution for the cost-effective production of many industrially valuable biomaterials. Many of these applications draw from precursor molecules already found in plants and have resulted in the efficient production of many valuable biomaterials (Table 4).

Table 4: Biomaterials, enzymes and amino acids engineered via the plastid genome (Grevich, J.J. and Daniell, 2005).

<table>
<thead>
<tr>
<th>Desired Biomaterial</th>
<th>Gene</th>
<th>Site of integration</th>
<th>Promoter</th>
<th>5'/3' UTRs</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-hydroxybenzoic acid</td>
<td>adhC</td>
<td>trnV/trnA</td>
<td>Prn</td>
<td>PpsbA/TpsbA</td>
<td>Viitanen et al., 2004.</td>
</tr>
<tr>
<td>Polyhydroxybutyrate</td>
<td>phb operon</td>
<td>trnN/trnR</td>
<td>PpsbA</td>
<td>PpsbA/TpsbA</td>
<td>Lössl et al., 2003.</td>
</tr>
<tr>
<td>Xylanase</td>
<td>xynA</td>
<td>rbcL/accD</td>
<td>PpsbA</td>
<td>PpsbA/TpsbA</td>
<td>Leelavathi et al., 2003.</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>AS42</td>
<td>ndhF/trnL&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Prn</td>
<td>rbcL/rpL32&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Zhang et al., 2001.</td>
</tr>
</tbody>
</table>

Plastid Genomics

A “universal vector” (Daniell et al., 1998), or one containing plastid DNA flanking sequences that transform another species (of unknown genome sequence), was used to successfully transform both potato and tomato plastid genomes with flanking sequences from tobacco (Ruf et al., 2001; Sidorov et al., 1999). However, the efficiency of these transformations was significantly lower when compared to that of tobacco. Only one chloroplast transgenic line per 35 bombarded plates was obtained from potato, and only one transgenic line per 87 bombarded plates was obtained with tomato. Even though the flanking sequences are ~98% homologous, this approach is greatly inefficient when compared to tobacco with an average of 15 transgenic lines yielded from one bombarded plate (Fernandez-San Milan et al., 2003). This same reduction of efficiency was seen when Petunia flanking sequences (also ~98% homologous) were used to transform the tobacco chloroplast genome (DeGray et al., 2001).
Additionally, optimization of transformation protocols may be needed for each species. Tomato plants transformed using standard transformational protocols for tobacco did not produce a single positive transformant or spontaneous mutant (Ruf et al., 2001). The same scientists reported that in order for successful transformation of tomato, one needed to reduce light intensity, increase primary selection time, and reduce the size of leaf pieces used for selection.

Lack of complete chloroplast genome sequence and analysis is one of the factors slowing the extension of plastid engineering to useful crops. Only 23 published crop chloroplast genomes are currently available, although over 200 non-crop genomes have been sequenced or are in progress. Chloroplast genome sequences are essential for the identification of spacer regions and endogenous regulatory sequences used for the integration of transgenes at optimal sites and for ideal expression of transgenes, respectively. The 5’UTR of the psbA mRNA has been shown to be a key site for RNA-protein interactions involved in the translational regulation of this gene (Kim and Mayfield, 1997). The use of this UTR has been shown to increase protein expression by over 500-fold when compared to the use of only a ribosome binding site (Fernández-San Millán et al., 2003).

In part due to the vast amount of recently sequenced genomes, plastid genomics has uncovered a wealth of insight into the genome organization and evolution of plant plastids. Solanaceae (Atropa, Tobacco, Potato and Tomato), legume (Glycin, Lotus, and Medicago), Cotton, and Grape (Vitaceae) genomes were sequenced and thoroughly analyzed using modern bioinformatic tools (Daniell, H., et al., 2006; Saski et al., 2005, Lee, et al., 2006, Jansen, R.K., et al, 2006). A comprehensive genome-wide analysis examined not only gene content and order, but also repeat structures, intergenic spacer regions, sequence divergence, and RNA editing (Daniell, H., et al., 2006). This analysis revealed that gene content and order among the four
Solanaceaea genomes was identical, supporting the notion that chloroplast genomes are highly conserved. With the advancement of bioinformatics, the computational analysis of future plastid sequences will provide data that are even more meaningful, in less time, and with greater ease. This type of analysis should significantly aid the process of expanding plastid genetic engineering to other crops as well as improve the efficiency of crops that have already been successfully transformed.

**Redox Regulation and Protein Disulfide Isomerases In Plants**

Redox regulation of chloroplast proteins has been well studied. It has even been argued that the redox control of some plastid proteins acts as a selective force in preventing redox regulated genes into the nucleus (Allen, 2003). For over 35 years it has been known that thioredoxin (Trx) functions in regulating photosynthesis. Trx has been shown to regulate the redox status of disulfide bonds in proteins that regulate a range of processes such as seed germination, transcription, cell division, radical scavenging and detoxification (Yano, Kuroda, and Buchanan, 2002). More recently, Trx has been shown to also regulate the light activated clp protease (Balmer *et al.*, 2003).

Protein disulfide isomerases (PDIs) are a family of proteins within the thioredoxin superfamily that are known to function as molecular chaperones and aid in the formation of disulfide bonds during protein folding. By definition, they contain at least one Trx domain used for the reduction, oxidation, and isomerization of disulfide bonds (Lumb and Bulleid, 2002). In plants, PDIs were shown to be involved in the folding of storage proteins during biogenesis of protein bodies in the seed endosperm (Shimoni *et al*., 1995; Li and Larkins, 1996). PDI prevents
the aggregation of proglutelin and prolamin by assisting with their segregation in rice seeds (Takemoto et al., 2002). In *Arabidopsis*, some PDI genes are upregulated by the unfolded protein response (Martinez and Chrispeels, 2003). Most recently, it has been proposed that a 65 kDa PDIL associated with the stromal fringe is involved with the biogenesis of transitory starch (Lu and Christopher, 2006). Genomic analysis revealed 13 PDI-like (PDIL) proteins found in *Arabidopsis* that contain at least one Trx domain (Houston et al., 2005), unfortunately such studies have not been done with tobacco. Most advancements in learning about post-transcriptional mechanisms in chloroplast gene expression have stemmed from the *C. reinhardtii* system, yet very little research has been done on these same mechanisms in higher plants, which often times differ from their more primitive counterparts. For instance, the *psbA* 5’utr sequence is highly conserved in *Arabidopsis* when compared to other similar plants, yet not with *C. reinhardtii* (Shen, Danon, and Christopher, 2001).

The most substantial change in chloroplast protein expression occurs during the light induced greening of plastids, aided by an abundant pool of pre-existing mRNAs. (Zou, Z. 2001). Translation of the *psbA* mRNA is light activated (Mattoo et al., 1989) and PDIs have been shown to be directly involved in the translational regulation of the *psbA* mRNA (Kim and Mayfield, 1997; Shen et al., 2001). Translation of the chloroplast *psbA* gene requires the interaction of *trans*-acting factors with the 5’UTRs of these mRNAs. Unlike what is seen in nuclear mRNAs, it is still uncertain whether protein complexes interacting with the 3’UTR in chloroplasts also interact with complexes associated with the 5’UTR (Tarun and Sachs, 1995; Caponigro and Parker, 1996). In *Arabidopsis*, four proteins, RB38, RB47, RB55, and RB60, make up the RNA-protein complex of the *psbA* 5’UTR. This mechanism has been shown to be modulated by a redox-regulated pathway involving other poly(A)-binding proteins (PABPs) (Shen, et al., 2001).
Light enhanced expression of recombinant genes is exhibited when fused to the *psbA* 5’UTR, but not with other UTR regions, suggesting that the translational regulation occurs during the initiation stage and that the *psbA* 5’UTR contains essential regulatory elements involved in that system (Cherukumilli, V.S., 2005; Zou, Z., 2001; Eibl et al., 1999). Experiments in *Chlamydomonas*, suggest that the *psbA* translational regulation is directly mediated by RB47, a PABP known to bind with high affinity to the *psbA* mRNA and complex with three other proteins, RB38, RB55, and RB60 (Danon and Mayfield, 1991; Shen, et al., 2001).

*Chlamydomonas* mutants lacking RB47 cannot translate *psbA* mRNA (Fong, C.L., et al., 1999; Yohn, et al., 1998a; Yohn et al., 1998b). The binding activity of this complex with the *psbA* mRNA directly correlates with the light-enhanced translation under a variety of experimental conditions, both *in vivo* and *in vitro* (Danon and Mayfield, 1991; Hauser et al., 1996; Mayfield et al., 1994; Shen et al., 2001). Mutating the cysteine residues to serine within the RNA recognition motif of RB47 results in high affinity binding of *psbA* mRNA regardless of the redox state (Fong, C.L., et al., 1999). RB38 shows no homology to proteins with known function and the identity of RB55 is still unknown (Yohn et al., 1998a; Shen et al., 2001).

RB60 is an atypical PDI that serves as a redox sensor component of the *psbA* mRNA binding complex (Levitan A., et al., 2005). Interestingly, it is localized in both the chloroplasts and endoplasmic reticulum (ER) even though it contains a C-terminal ER retention signal (KDEL). The first 50-amino acid leader of RB60 is sufficient for targeting to both the ER and chloroplast (Levitan A., et al., 2005). Under steady state conditions within the chloroplast, RB60 occurs in both a soluble form found in the stroma and a form tightly bound to the thylakoid membrane (Trebitsh, T., et al., 2001). RB60 controls the *psbA* mRNA regulatory complex via two separate mechanisms, one that promotes the translation during periods of light and one that
inhibits translation during the dark. In the first, the light environment of the plant cell sends a reductive signal to RB60 via a ferredoxin-thioredoxin system during photosynthesis (Trebitsh, T., 2000). The reduction of RB60 and its related complex protein RB47 increases the translational efficiency of psbA mRNA. In the second mechanism, ADP concentrations attained only in the dark cause an ADP-dependent phosphorylation of RB60 and inactivate the protein complex assembly; therefore, this pathway inhibits psbA mRNA translation in the dark (Danon A., and Mayfield, S.P., 1994a).
MATERIALS AND METHODS

Genome Annotation And Map Construction

Gene annotation was accomplished using the DOGMA application (Dual Organellar GenoMe Annotator; http://bugmaster.jgi-psf.org/dogma/). This program identifies putative protein coding genes by performing BLASTX searches against a custom database of previously annotated chloroplast genomes. DOGMA uses BLASTN searches against the same database when identifying tRNAs and rRNAs. Using the annotated sequence data, chloroplast genome maps were constructed using a mix of GenVision (http://www.dnastar.com/web/r4.php), Microsoft Excel (http://www.microsoft.com/), and Adobe Illustrator (http://www.adobe.com/).

Sequence Content Analysis

Analysis of gene content and order was performed using Multipipmaker (http://pipmaker.bx.psu.edu/pipmaker/) by comparing the newly sequenced potato (DQ347958) and tomato (DQ347959) genomes against two other solanaceae species, Atropa (NC_004561) and tobacco (NC_001879). The gene orders were examined using pair-wise comparisons of the four Solanaceae genomes using Pipmaker (http://pipmaker.bx.psu.edu/pipmaker/).

Repeat structure of the same four chloroplast genomes was determined first by using REPuter (http://www.genomes.de/) to identify the number, size, and location of both direct and inverted repeats with a minimum size of 30 bp, hamming distance of 3 (i.e., sequence identity of $\geq 90\%$) [repfind –f –p –l 30 –h 3 –best 10000]. Furthermore, repeats from tobacco were blasted
against a custom database containing all four analyzed Solanaceae genomes. Blast hits larger than 30 bp, bit score > 40, and sequence homology of at least 90% were annotated and used to confirm the shared repeats found with REPuter.

DNA sequence divergence was determined using an aligned data set of all the shared genes among the four Solanaceae chloroplast genomes by extracting the sequences from the annotated genomes using either DOGMA or the Chloroplast Genome Database (http://cbio.psu.edu/chloroplast/index.html). This data set was loaded into ClustalX (Higgins, D.G., 1996) for sequence alignment followed by manual adjustments using EditSeq (http://www.dnastar.com/web/r7.php).

**Molecular Evolutionary Analysis**

Molecular evolutionary analyses and estimates of sequence divergence based on the Kimura 2 parameter were performed using MEGA2 (Molecular Evolutionary Genetics Analysis) (Kumar, S., and Gadagkar, S.R., 2001; Kimura, M., 1980). Comparisons of intergenic spacer regions were first compared using Multipipmaker (http://pipmaker.bx.psu.edu/pipmaker/) and verified and further elaborated using a custom suite of applications that iteratively compares pairs of nucleotide sequences and calculates percent identity over a given length of sequence.

**Sequence and DNA Acquisition**

A few potential PDI candidates were identified at the RIKEN Biological Resource Center (http://www.brc.riken.jp/inf/en/) Arabidopsis cDNA bank by using various BLAST searches to determine the availability of higher plant PDI cDNAs by using the *Chlamydomonas* RB60 amino
acid sequence as a query. At the time, searching revealed no definite Arabadopsis chloroplast PDIs or even definite PDIs since this was such new territory. A putative PDI cDNA (as of February 2005 it has been changed to AtPDIL5-4; thioredoxin family protein) was obtained from RIKEN BRC in the plastmid, λFLC-1-B. The HSA gene was obtained from Chlorogen (http://www.chlorogen.com) in the clone, pCTT312, and contained a shortened 5’psbA UTR (promoter elements removed) due to anomalies found when expressing HSA with the full-length UTR. This modified UTR was eventually used for the 5’UTR•PDI constructs as well.

**Preparation of Competent Cells**

Ultra competent cells were prepared using the rubidium chloride method (Kumar *et al.*, 2004) and a lab glycerol stock of *E. coli* XL-1 Blue (Stratagene) cells. Cells were plated on a LB agar plate containing 12.5 µg/ml tetracycline and incubated at 37°C for 16 hours. Single colonies were picked and grown in 5 ml of Psi broth (5g/L Bacto yeast extract, 20g/L Bacto Tryptone, 5g/L magnesium sulfate, pH 7.6) containing 12.5 µg/ml tetracycline and incubated at 37°C for 16 hours in a horizontal shaker at 225 rpm.

Approx. 1 ml of the overnight culture was inoculated in 100 ml of Psi broth and was incubated at 37 °C at 225 rpm until an optical density (OD) 0.48 OD 600 was achieved. The culture was kept on ice for 15 minutes and pelleted by centrifugation at 3000g/5000 rpm for 5 minutes in a sorvall centrifuge. The supernatant was discarded and the pellet was resuspended in 0.4 volume (40 ml) of ice-cold Tfb-I solution. The cells were re-pelleted at 3000g / 5000 rpm for 5 minutes. The supernatant was discarded and the cells were resuspended in (0.04 volume) 4 ml
of Tfb-II solution and immediately iced for 15 minutes. This suspension was divided into 100 µl aliquots, then quick frozen in liquid nitrogen and stored at -80°C.

**Transformation of the Competent E. coli XL1-Blue cells**

The competent cells of one hundred µl aliquots were taken from -80°C and immediately thawed on ice and transferred to a falcon 2059 tube. About one µl (~100 ng) of plasmid DNA was added to the competent cells and was mixed by gentle tapping. The mixture was incubated on ice for 30 minutes with gentle swirling every 10 min. The cells were then heat-shocked at 42°C in a water bath for 45 seconds and then immediately put on ice for two minutes. Approx. 900µl of LB broth was added to cells and were incubated at 37°C for 45 minutes in a horizontal shaker at 225 rpm. The cells were pelleted by spinning at 13,000 rpm for 30 seconds. The eight hundred µl of supernatant was discarded leaving 100µl, followed by resuspending the cells. Two samples, 50µl and 100µl, of the suspension were inoculated on the agar plate with appropriate selection agent and spread with a glass rod.

**Construction of chloroplast transformation vectors**

The pLD constructs based on the universal chloroplast vector pLD (Fig.1) that has been routinely used in this laboratory (Daniell 2001a). The PDIL gene was amplified using PCR with a high fidelity polymerase (Pfu) using the forward primer 5’-CTC CGA CGA CCA TGG TGT CCA CG-3’ and reverse primer 5’-TTG GAA TTT TAG CGG CAA AAT GGG-3’. Primers were designed using MIT’s Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) to engineer NcoI and XbaI restriction sites at the 5’ and 3’ ends of the protein coding sequence,
respectively. The amplified PCR product was cloned into various constructs using the pLD vector. The psbA 5’ UTR is light regulated and known to enhance translation (Fig 1). The g10 UTR is a constitutive expressing viral leader sequence found in the T7 bacteriophage. The constitutive 16S rRNA promoter is upstream of all the constructs and drives the transcription of the aminoglycoside 3’adenylyltransferase (aadA) gene and inserted transgenes. The aadA gene confers spectinomycin resistance and was used as a selectable marker for both bacterial and plant cultures. Both psbA and rps16 3’ UTRs are used to confer transcript stability. CopyCutter EPI400 (http://www.epibio.com/item.asp?ID=435) competent cells were used to overcome the toxic effects of the PDIL gene in E. coli when making the g10•PDI contracts.

**Figure 1**: Overview of all final constructs within the pLD chloroplast transformation vector. **16S**: 3’ end of 16S Ribosomal RNA Gene, part of 5’ chloroplast flanking region, **trnI**: 5’ flanking region, **Prrn**: 16S rRNA promoter, **psbA5’**: psbA 5’UTR-P (without promoter elements, supplied by Chlorogen, St. Louis, MS), **g10**: g10 leader sequence from T4 bacteriophage, **HSA**: Human serum albumin. **PDI**: AtPDIL5-4, **Trps16**: 3’UTR from rps16, **TpsbA3’**: 3’UTR from psbA, **trnA**: 3’ chloroplast flanking region.
Isolation of Plasmid DNA by Alkaline lysis method

A single isolated colony was picked from the desired LB agar plate and inoculated in about 3-5 ml of LB broth, both containing the appropriate antibiotics. The culture was then grown for 12-16 hours at 37 °C in a shaker at 225 rpm. 1.6 ml of the cell culture was put into an eppendorf tube and centrifuged at 13,000 rpm for 1 minute. The supernatant was discarded and the pellet was resuspended in 150 µl of Solution I (50 mM Glucose, 10 mM EDTA, 25 mM Tris, pH- 8) and vortexed to resuspend the cells uniformly. One µl of RNase (100 mg/ml) is then added to each tube and pulse vortexed to degrade any present RNA. 150 µl of Solution II (0.2N NaOH, 10% SDS) was added to each tube and mixed by gently inverting the tube 6-8 times along with barrel rolls. 210 µl of Solution III (60 ml of 5M Potassium Acetate, 11.5 M glacial acetic acid, 28.5 ml sterile dH₂O) was added immediately after adding Solution I and mixed by gently inverting the tube about 6-8 times as described above. A cloudy white precipitate is observed, the mixture is then centrifuged at 13,000 rpm for 10 minutes at 4 °C. The supernatant was transferred into a fresh, eppendorf tube taking care to exclude any white precipitate. Then, 1 ml of ice-cold ethanol (95%) was added to the supernatant and inverted 6-8 times in order to precipitate the plasmid DNA. The supernatant containing the ice-cold ethanol was centrifuged at 13,000 rpm at 4 °C for 10 minutes in order to pellet the plasmid DNA. The supernatant was removed and discarded, taking care not to dislodge the plasmid DNA pellet at the bottom. 500 µl of 70 % room temperature ethanol was added and (without mixing) centrifuged for 5 minutes to remove salts and aid in dissolving the DNA pellet. The ethanol was discarded and the pellet was subsequently dried in the speedvac on medium heat for 3-5 minutes. Dried DNA pellet was resuspended in TE (pH 8.0) or H2O. Concentration and quality plasmid of DNA was measured.
by spectrophotometer using both 1:10 and 1:20 dilutions. DNA was stored at -20 °C. The DNA samples were digested with appropriate enzymes according to manufacturers protocol and run on a 0.8% agarose gel for 40 minutes at 80 volts to confirm the successful isolation of plasmid (Sambrook et al. 1989).

**Plant Bombardment**

Gene delivery by particle bombarded and subsequent selection process was performed as described by Daniell (1997). *Nicotiana tabacum* var Petit Havana leaves were bombarded using a PDS-1000/He (http://www.biorad.com/) biolistic device. After bombardment, the leaves were subjected to two rounds of 500 µg mL⁻¹ spectinomycin on plates RMOP medium and subsequently moved to jars containing Murashige Skoog medium with the same level of antibiotic. After regeneration and selection, plants were transferred to pots and grown in the greenhouse.

**Confirmation of Chloroplast Integration by PCR**

Plant DNA was isolated by using the DNEasy Plant Mini kit (http://www.qiagen.com). The primer pairs 3P (5’-GGA ATT GAA TTC CAT ATG TGT GAG AAC AGA) 3M (5’-AGA ATT GCC TCT AGA CTA TTC TGA AAC-3’) and 5P (5’-ATG TAG AAG TCA CCA TTG TTG TGC-3’) 2M (5’-GAC TGC CCA CCT GAG AGC GGA CA-3’) were used to confirm the integration of the transgenic cassette into the chloroplast DNA and verify the presence of the genes of interest as previously described (Daniell et al. 20001; De Cosa et al., 2001). After positive confirmation of transgenes, plant shoots were transferred to MS0 medium with 500
µg/ml spectinomycin and subsequently put through another round of selection and confirmed via PCR once more.

**Figure 2**: Schematic representing where the primers used for PCR confirmation land and the corresponding sizes of their products. Note*, primer 3P lands within native chloroplast DNA and not within the chloroplast transformation vector.

**Southern blot analysis**

Total plant DNA from different lines of transgenic plants and one wild-type plant was digested overnight with BglII, separated on a 0.7% agarose gel at 50V for 4 hours. The gel was then depurinated by immersing it in 0.25M HCl (depurination solution) for 15 minutes (until the color of the dye became yellow), washed twice in dH₂O for 5 minutes, and then equilibrated in transfer buffer (0.4N NaOH, 1M NaCl, filled to 1 liter with water) for 20 minutes. The DNA was transferred overnight to a nitrocellulose membrane. The next day the membrane was washed with 2X SSC (3M NaCl, 0.3M Na Citrate, H₂O, the pH was adjusted with 1N HCl to 7 and water
was added to 1L) for 5 minutes. Following, the membrane was allowed to dry on a Whatman paper for 5 minutes and then cross-linked using the Bio-Rad GS Gene Cross Linker at setting C3 (150 m joules). pUC-CT vector DNA was digested with *Bam*HI and *Bgl*II to generate a 0.8 kb probe which was used as a flanking probe as it contains the chloroplast flanking sequences for the *trn*I and *trn*A. a 2.0 kb g10-HSA and 650 bp PDIL fragments were used as gene specific probes. After labeling the probe with P$^{32}$, hybridization of the membranes was done using Stratagene QUICK-HYB hybridization solution and protocol (Stratagene, La Jolla, CA). The membrane was washed twice as follows: 50ml of wash solution number 1 (2X SSC and 0.1% SDS) was poured and incubated at room temperature for 15 minutes. The liquid was discarded in the liquid waste container and the step was repeated. A second round of washes was performed twice by pouring 50ml of solution number 2 (0.1X SSC and 0.1% SDS) and incubating it for 15 minutes at 60 °C to increase the stringency. The liquid of these washes were discarded into the radioactive liquid container. The membranes were placed into the plastic wrap, placed in the film cassette and then taken to the dark room. Using the safe light (red light), the X-ray film was placed into the cassette on top of the blot and the intensifier screen was placed on top of the X-ray film. The cassette with the blot and the film was placed into a black bag to protect against light and then incubated overnight at –80 °C. The next day the cassette was taken out from the –80 °C, allowed to thaw, and then moved to the dark room where the film was developed.

**Western Blot Analysis**

Crude extract were made using 3x PBS washes of 1 ml overnight culture in TB for each construct. Approximately 20 µl of crude protein extract was loaded onto a SDS-PAGE gel (4%
Leaf material (100 mg) was ground into a fine powder using liquid nitrogen and resuspended in 200 µL of protein extraction buffer (200 mM Tris-HCl pH 8.0, 100 mM NaCl, 400 mM Sucrose, 14 mM βME, 0.05% Tween-20, 0.1% SDS, 10 mM EDTA, 2 mM PMSF), followed by a centrifugation step at 15.7g for 2 minutes to remove any solids. Leaf extracts were boiled in sample buffer (Bio-Rad) and electrophoresed in SDS-PAGE gel at 30 mA for 2 hrs (15 mA per gel). Proteins were stained with Coomassie Brilliant Blue G-250 or transferred to nitrocellulose for subsequent immunoblotting. The antibodies used and their respective dilutions were the following: anti-HSA (Sigma, St. Louis, MO), dilution 1:3000; anti-atPDI (*Arabidopsis* PDI) (Rose Biotechnology, http://www.rosebiotech.com), dilution 1:1000, anti-rabbit IgG-HRP (Sigma, St. Louis, MO), dilution 1:4000.
RESULTS

Sequence Content Analysis

The tomato and potato chloroplast genomes are 155,460 bp and 155,372 bp (Fig. 3). The inverted repeats regions for the two genomes are 25,613 bp and 25,588 bp, which separate the small single copy (18,361 bp and 18,381 bp) and large single copy regions (85,873 bp and 85,815 bp). Gene content and order of the four genomes is identical (Fig. 4), however, there are significant rearrangements within the certain coding sequences such as the 11 bp addition (ACACGGGAAAC) in the 16S rRNA gene of potato, tomato and Atropa, but not found in tobacco or any other sequenced chloroplast genome (Fig. 7). Notably, there are four substantial deletions within the ycf2 coding region among the four genomes. Two of those deletions are found within a region with repeated sequences, and one of these regions contains a repeated ribosome binding site (GGAGG) (Fig. 6). This modification to the ycf2 gene gives tomato three ribosome binding sites, whereas the other sequenced Solanaceae genomes only contain one.
Figure 3: Genomic map of Tomato and Potato chloroplast genomes with identified repeats. The thick lines indicate the inverted repeats, which separate the genome into small (SSC) and large (LSC) single copy regions. Genes outside the map are transcribed in the clockwise direction and vice versa for genes inside the map. Numbered arrows indicate the position of repeated sequences throughout the Solanacea genomes (See Table 4). *tobacco + tomato **tobacco + atropa ***tobacco only
Figure 4: Whole genome and regional output examples from Multipipmaker (http://pipmaker.bx.psu.edu/pipmaker/). Above, aligned regions appear green and strongly aligned regions are marked in red. Below, “percent identity plot” (PIP), of a regional section within the four Solanaceae genomes.

**Repeat Analysis**

By excluding all genomic inverted repeats from all repeat analysis, REPuter found 33 to 45 direct and inverted repeats 30 bp or longer with a sequence identity of at least 90% (Fig. 5). The majority of these repeats are between 30-40 bp in length and no longer than 57 bp, which were found only in tomato. Both tobacco and potato share a 50 and 56 bp repeat, whereas *Atropa* does not have a single repeat in the 50+ bp size range.

BlastN comparisons of the tobacco repeats against the chloroplast genomes of *Atropa*, potato and tomato identified 42 repeats that show a sequence identity ≥ 90% with sequences ≥ 30 bp and a bit score greater than 40 (Table 5, Fig. 3). 37 of the 42 repeats are found in all four Solanaceae chloroplast genomes, all of which are located in the same genes or intergenic regions.
**Figure 5**: Histogram representing the number of repeated sequences of at least 30 bp with a sequence identity of at least 90% in the four Solanacea genomes using Reputer.
Table 5: Tobacco repeat blast hits ≥ 30 bp, a sequence identity ≥ 90%, and bit-score ≥ 40. N = *Nicotiana* (tobacco), A = *Atropa*, P = potato, T = tomato; IGS = intergenic spacer.

<table>
<thead>
<tr>
<th>Repeat #</th>
<th>Size(bp)</th>
<th># of hits</th>
<th>Location</th>
<th>Genomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>2</td>
<td>IGS(1bp) - <em>trnS-GCC</em></td>
<td>NAPT</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>1</td>
<td>IGS - <em>(psbC - <em>trnS-UGA</em>)</em></td>
<td>NT</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>1</td>
<td>IGS(1bp) - <em>trnS-UGA</em></td>
<td>NAPT</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>1</td>
<td>Intron - <em>(ycf3 exon 2 - ycf3 exon 3)</em></td>
<td>NAPT</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>2</td>
<td><em>trnS-GCU</em> - IGS(1bp), <em>trnS-GGA</em> - IGS(1bp)</td>
<td>NAPT</td>
</tr>
<tr>
<td>6</td>
<td>30</td>
<td>1</td>
<td>Intron - <em>(clpP exon 2 - <em>clpP</em> exon 3)</em></td>
<td>NA</td>
</tr>
<tr>
<td>7</td>
<td>30</td>
<td>2</td>
<td>Ycf2</td>
<td>NAPT</td>
</tr>
<tr>
<td>8</td>
<td>30</td>
<td>2</td>
<td>Ycf2</td>
<td>NA</td>
</tr>
<tr>
<td>9</td>
<td>30</td>
<td>2</td>
<td>IGS - <em>(<em>bps12</em> 3'end - <em>trnV-GAC</em>)</em></td>
<td>NAPT</td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>2</td>
<td>IGS - <em>(trnV-GAC - <em>bps12</em> 3'end)</em></td>
<td>NAPT</td>
</tr>
<tr>
<td>11</td>
<td>30</td>
<td>2</td>
<td>Ycf2</td>
<td>NAPT</td>
</tr>
<tr>
<td>12</td>
<td>30</td>
<td>2</td>
<td>Ycf2</td>
<td>NAPT</td>
</tr>
<tr>
<td>13</td>
<td>30</td>
<td>2</td>
<td>Ycf2</td>
<td>NA</td>
</tr>
<tr>
<td>14</td>
<td>30</td>
<td>2</td>
<td>Ycf2</td>
<td>NAPT</td>
</tr>
<tr>
<td>15</td>
<td>31</td>
<td>2</td>
<td>IGS(2bp) - <em>trnS-GCU</em>, IGS(1bp) - <em>trnS-GGA</em></td>
<td>NAPT</td>
</tr>
<tr>
<td>16</td>
<td>31</td>
<td>1</td>
<td><em>trnG-GCC</em> - IGS(4bp)</td>
<td>NAPT</td>
</tr>
<tr>
<td>17</td>
<td>31</td>
<td>1</td>
<td>IGS(2bp) - <em>trnS-UGA</em></td>
<td>NAPT</td>
</tr>
<tr>
<td>18</td>
<td>31</td>
<td>1</td>
<td><em>trnG-GCC</em> - IGS(3bp)</td>
<td>NAPT</td>
</tr>
<tr>
<td>19</td>
<td>31</td>
<td>1</td>
<td>Intron - <em>(rpl16 exon 1 - rpl16 exon 2)</em></td>
<td>NAPT</td>
</tr>
<tr>
<td>20</td>
<td>31</td>
<td>3</td>
<td>IGS - <em>(bps12 3'end - <em>trnV-GAC</em>)</em> x2, Intron - <em>(ndhA exon 1 - ndhA exon 2)</em></td>
<td>NAPT</td>
</tr>
<tr>
<td>21</td>
<td>32</td>
<td>2</td>
<td>IGS - <em>(trnH-GUG - <em>psbA</em>)</em></td>
<td>N</td>
</tr>
<tr>
<td>22</td>
<td>34</td>
<td>4</td>
<td>IGS - <em>(rrn4.5 - rrn5)</em></td>
<td>NAPT</td>
</tr>
<tr>
<td>23</td>
<td>34</td>
<td>4</td>
<td>IGS - <em>(rrn4.5 - rrn5)</em></td>
<td>NAPT</td>
</tr>
<tr>
<td>24</td>
<td>34</td>
<td>4</td>
<td>IGS - <em>(rrn4.5 - rrn5)</em></td>
<td>NAPT</td>
</tr>
<tr>
<td>25</td>
<td>34</td>
<td>4</td>
<td>IGS - <em>(rrn4.5 - rrn5)</em></td>
<td>NAPT</td>
</tr>
<tr>
<td>26</td>
<td>35</td>
<td>4</td>
<td>IGS - <em>(ycf15 - <em>trnL-CAA</em>)</em></td>
<td>NAPT1</td>
</tr>
<tr>
<td>27</td>
<td>35</td>
<td>4</td>
<td>IGS - <em>(ycf15 - <em>trnL-CAA</em>)</em></td>
<td>NAPT2</td>
</tr>
<tr>
<td>28</td>
<td>37</td>
<td>4</td>
<td>Ycf2</td>
<td>NAPT</td>
</tr>
<tr>
<td>29</td>
<td>37</td>
<td>4</td>
<td>Ycf2</td>
<td>NAPT</td>
</tr>
<tr>
<td>30</td>
<td>37</td>
<td>4</td>
<td>Ycf2</td>
<td>NAPT</td>
</tr>
<tr>
<td>31</td>
<td>37</td>
<td>4</td>
<td>Ycf2</td>
<td>NAPT</td>
</tr>
<tr>
<td>32</td>
<td>39</td>
<td>4</td>
<td>Intron - <em>(ycf3 exon 2 - ycf3 exon 3)</em>, IGS - <em>(bps12 3'end - <em>trnV-GAC</em>)</em> x2,</td>
<td>NAPT</td>
</tr>
<tr>
<td>33</td>
<td>39</td>
<td>4</td>
<td>Intron - <em>(ndhA exon 1 - ndhA exon 2)</em></td>
<td>NAPT</td>
</tr>
<tr>
<td>34</td>
<td>39</td>
<td>4</td>
<td>Intron - <em>(ndhA exon 1 - ndhA exon 2)</em></td>
<td>NAPT</td>
</tr>
<tr>
<td>35</td>
<td>39</td>
<td>4</td>
<td>Intron - <em>(ndhA exon 1 - ndhA exon 2)</em></td>
<td>NAPT</td>
</tr>
<tr>
<td>36</td>
<td>41</td>
<td>4</td>
<td>Intron - <em>(ycf3 exon 2 - ycf3 exon 3)</em>, IGS - <em>(bps12 3'end - <em>trnV-GAC</em>)</em> x2,</td>
<td>NAPT</td>
</tr>
<tr>
<td>37</td>
<td>41</td>
<td>4</td>
<td>Intron - <em>(ndhA exon 1 - ndhA exon 2)</em></td>
<td>NAPT</td>
</tr>
<tr>
<td>38</td>
<td>41</td>
<td>4</td>
<td>Intron - <em>(ycf3 exon 2 - ycf3 exon 3)</em>, IGS - <em>(bps12 3'end - <em>trnV-GAC</em>)</em> x2,</td>
<td>NAPT</td>
</tr>
<tr>
<td>39</td>
<td>48</td>
<td>2</td>
<td>IGS(47bp) - <em>psbN</em> (1bp)</td>
<td>NAPT3</td>
</tr>
<tr>
<td>40</td>
<td>50</td>
<td>2</td>
<td><em>psaB</em>, <em>psaA</em></td>
<td>NAPT</td>
</tr>
<tr>
<td>41</td>
<td>50</td>
<td>2</td>
<td><em>psaB</em>, <em>psaA</em></td>
<td>NAPT</td>
</tr>
<tr>
<td>42</td>
<td>56</td>
<td>2</td>
<td>Intron - <em>(petD exon 1 - petD exon 2)</em></td>
<td>NAPT4</td>
</tr>
</tbody>
</table>

*Blat hit is 4 bp less in tomato,*  
*Blat hit is 4 bp less in tomato,*  
*Blat hit is 17 bp less in potato,*  
*Blat hit is 2 bp less in tomato*
Intergenic Spacer Region Sequence Divergence

Among intergenic spacer regions >8 bp, only four spacer regions (rps11 - rpl36, rps7 - rps12 3’ end, trnI-GAU - trnA-UGC, ycf2 - ycf15) have 100% sequence identity between all four genomes, and three of these regions are located within the inverted repeats (Fig. 8, Table 6). Tomato and potato are most similar with 21 intergenic spacer regions having 100% sequence identity. Only 8 regions have 100% sequence identity between tomato and Atropa, tobacco and potato, Atropa and potato, 9 regions between tobacco and tomato and 10 regions between tobacco and Atropa. Interestingly, one of the intergenic spacer regions that is 100% similar between Atropa and potato (trnI-CAU - ycf2) has only 66-69% sequence identity among the other Solanaceae species examined. Similarly, ycf4 - cemA has only 27% identity between tobacco and Atropa, potato and tomato, whereas it has greater than 90% identity between other Solanaceae species examined. These types of observations are most likely a result of the deletions or insertions in the intergenic spacer regions between trnQ - rps16, trnE - trnT, trnK - rps16, trnT - ycf15, trnS - trnG, ycf2 - trnI, ycf4 - cemA, ycf15 - trnL.

Figure 6: Alignment of ycf2 gene of all four Solanaceae genomes with highlighted insertion, deletion, and substitution events.
**Figure 7:** Alignment of the 5' portion of the 16S ribosomal RNA subunit showing a 9 bp insertion found only in *Atropa*, potato, and tomato. Insertion, deletions and substitutions have been highlighted.

**Figure 8:** Histogram representing the sequence divergence of the 21 most variable intergenic regions using pairwise comparisons. *, **, *** indicate the tier 1, tier 2, and tier 3 regions reported by Shaw et al. The plotted values were converted from a percent identity to sequence divergence on a scale from 0 to 1 (Also see Table 6).
Table 6: Intergenic spacer regions that are 100% identical to at least one other member of the sequenced Solanaceae genomes. Names of genomes compared are abbreviated: Pot for potato, Tom for tomato, Atr for Atropa, and Tob for tobacco

<table>
<thead>
<tr>
<th>Intergenic ID</th>
<th>Tob vs Atr</th>
<th>Tob vs Pot</th>
<th>Tob vs Tom</th>
<th>Atr vs Pot</th>
<th>Tom vs Pot</th>
<th>Tom vs Atr</th>
</tr>
</thead>
<tbody>
<tr>
<td>rps11:rpfl36</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>rps12_3end:rpfl7</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>trnA-UGC:trnL-GAU</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>ycf15:ycf2</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>trnV-GAC:rrn16</td>
<td>100</td>
<td>98</td>
<td>98</td>
<td>98</td>
<td>100</td>
<td>98</td>
</tr>
<tr>
<td>rnb4.5:rrn5</td>
<td>100</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>100</td>
<td>97</td>
</tr>
<tr>
<td>psbl:psblH</td>
<td>96</td>
<td>96</td>
<td>96</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>trnA-UGC:rrn23</td>
<td>96</td>
<td>100</td>
<td>96</td>
<td>100</td>
<td>100</td>
<td>96</td>
</tr>
<tr>
<td>trnM-CAU:rpfl14</td>
<td>100</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>100</td>
<td>97</td>
</tr>
<tr>
<td>trnN-GUU:ycf1</td>
<td>100</td>
<td>96</td>
<td>100</td>
<td>96</td>
<td>96</td>
<td>100</td>
</tr>
<tr>
<td>ycf1:trnN-GUU</td>
<td>100</td>
<td>96</td>
<td>100</td>
<td>96</td>
<td>96</td>
<td>100</td>
</tr>
<tr>
<td>trn23:trnA-UGC</td>
<td>96</td>
<td>100</td>
<td>95</td>
<td>100</td>
<td>96</td>
<td>100</td>
</tr>
<tr>
<td>psbl:psblH</td>
<td>95</td>
<td>95</td>
<td>95</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>rpl23:trnL-CAU</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>100</td>
<td>97</td>
</tr>
<tr>
<td>rnb4.5:rrn23</td>
<td>100</td>
<td>95</td>
<td>95</td>
<td>95</td>
<td>100</td>
<td>95</td>
</tr>
<tr>
<td>rps8:rpfl14</td>
<td>94</td>
<td>95</td>
<td>95</td>
<td>95</td>
<td>100</td>
<td>95</td>
</tr>
<tr>
<td>trnL-UAG:ccsA</td>
<td>95</td>
<td>94</td>
<td>94</td>
<td>95</td>
<td>100</td>
<td>95</td>
</tr>
<tr>
<td>trnD-GUC:trnY-GUA</td>
<td>94</td>
<td>94</td>
<td>94</td>
<td>94</td>
<td>100</td>
<td>94</td>
</tr>
<tr>
<td>ndhJ:ndhA</td>
<td>92</td>
<td>93</td>
<td>93</td>
<td>95</td>
<td>100</td>
<td>95</td>
</tr>
<tr>
<td>ndhD:psac</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td>94</td>
<td>100</td>
<td>94</td>
</tr>
<tr>
<td>rpoA:rpfl1</td>
<td>89</td>
<td>100</td>
<td>100</td>
<td>89</td>
<td>100</td>
<td>89</td>
</tr>
<tr>
<td>psbrH:petB</td>
<td>95</td>
<td>92</td>
<td>92</td>
<td>92</td>
<td>100</td>
<td>92</td>
</tr>
<tr>
<td>rpoC2:rpoxC1</td>
<td>95</td>
<td>92</td>
<td>92</td>
<td>91</td>
<td>100</td>
<td>93</td>
</tr>
<tr>
<td>rps14:psaB</td>
<td>95</td>
<td>91</td>
<td>91</td>
<td>91</td>
<td>100</td>
<td>92</td>
</tr>
<tr>
<td>trnL-CAU:ycf2</td>
<td>69</td>
<td>69</td>
<td>81</td>
<td>100</td>
<td>66</td>
<td>66</td>
</tr>
</tbody>
</table>

Coding Sequence Divergence

The chloroplast genes were categorized into 11 functional groups for the sequence divergence analysis within the coding regions (Fig. 9; Table 7). Using the Kimura 2-parameter method (Kimura, M., 1980), this analysis calculates the proportion of nucleotide sites that differ within each of the coding sequences examined. As expected, the two most closely related genomes (tomato and potato) exhibited the least amount of sequence divergence among all classes of genes. Among the functional groups, the cemA, clpP, matK, and ccsA were most
divergent while the RNA, photosynthesis, and ATP synthase genes were the most conserved. This data enforces the phylogenetic relationships described in previously published articles (Bohs, L., and Olmstead, R.G., 1997; Olmstead, R.G., et al., 1999; Spooner, D.M., et al., 1993).

**Figure 9:** Histogram representing the sequence divergence of coding regions for 11 different functional groups (Table 7). The plotted values were converted from a percent identity to sequence divergence on a scale from 0 to 1.
Table 7: Comparisons of sequence divergence of Solanaceae chloroplast genes among the 11 different functional groups. Standard errors are in parentheses and pairwise distances calculated using the Kimura 2-parameter model (Kimura, M., 1980). Names of genomes compared are abbreviated: Pot for potato, Tom for tomato, Atr for *Atropa*, and Tob for tobacco.

<table>
<thead>
<tr>
<th>Gene group</th>
<th>Length (bp)</th>
<th>Number of genes</th>
<th>Pot vs Tom</th>
<th>Pot vs Atr</th>
<th>Pot vs Tob</th>
<th>Tom vs Atr</th>
<th>Tom vs Tob</th>
<th>Atr vs Tob</th>
</tr>
</thead>
<tbody>
<tr>
<td>NADH</td>
<td>12102</td>
<td>11</td>
<td>0.005</td>
<td>0.015</td>
<td>0.012</td>
<td>0.017</td>
<td>0.014</td>
<td>0.013</td>
</tr>
<tr>
<td>Photosynthesis</td>
<td>14081</td>
<td>26</td>
<td>0.002</td>
<td>0.008</td>
<td>0.009</td>
<td>0.009</td>
<td>0.011</td>
<td>0.008</td>
</tr>
<tr>
<td>Ribosomal Protein</td>
<td>10207</td>
<td>22</td>
<td>0.003</td>
<td>0.010</td>
<td>0.010</td>
<td>0.010</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>RNA polymerase</td>
<td>10473</td>
<td>4</td>
<td>0.004</td>
<td>0.014</td>
<td>0.014</td>
<td>0.016</td>
<td>0.016</td>
<td>0.009</td>
</tr>
<tr>
<td>matK maturase ccsA-cytochrome synthesis</td>
<td>1530</td>
<td>1</td>
<td>0.011</td>
<td>0.025</td>
<td>0.022</td>
<td>0.031</td>
<td>0.029</td>
<td>0.017</td>
</tr>
<tr>
<td>cemA-envelope membrane protein</td>
<td>942</td>
<td>1</td>
<td>0.011</td>
<td>0.027</td>
<td>0.027</td>
<td>0.034</td>
<td>0.034</td>
<td>0.023</td>
</tr>
<tr>
<td>clpP-Protease</td>
<td>690</td>
<td>1</td>
<td>0.009</td>
<td>0.102</td>
<td>0.101</td>
<td>0.102</td>
<td>0.104</td>
<td>0.010</td>
</tr>
<tr>
<td>ATP synthase genes</td>
<td>621</td>
<td>1</td>
<td>0.033</td>
<td>0.090</td>
<td>0.099</td>
<td>0.109</td>
<td>0.117</td>
<td>0.026</td>
</tr>
<tr>
<td>tRNAs</td>
<td>4968</td>
<td>6</td>
<td>0.000</td>
<td>0.015</td>
<td>0.014</td>
<td>0.015</td>
<td>0.015</td>
<td>0.015</td>
</tr>
<tr>
<td>rRNAs</td>
<td>2751</td>
<td>27</td>
<td>0.000</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Vector Construction

The first cloning step after the PCR amplification with PDI had resulted in only six colonies after plating all of the transformation reaction (100, 200, and 700 µl onto LB Amp/Tet plates). Doing this with any other construct had yielded an average of ~10, ~50, and ~400 colonies respectively. After restriction digestion analysis, one of the colonies looked promising, but subsequent digestions did not work, suggesting that it was most likely a mutant. After repeating the cloning step multiple times and getting the same results, a different *psbA* 5' UTR (pB-BADH) containing vector was used. Again, the transformation yielded only a few colonies from the entire reaction, all of which proved to false positives.
Based on the theory that perhaps the \textit{NcoI} enzyme was not cutting on the 5’ end of the PCR product (due to not having enough overhanging base pairs), the PDI PCR product was blunt-end cloned into pBluescript KS+ via the \textit{EcoRV} restriction site. On two different occasions when this experiment was performed, there was a drastic reduction in the number of positive colonies when compared to a similar blunt ended transformational control. Out of the 12 white colonies picked from these plates, only one colony was successfully transformed into the vector, however, not in the desired orientation. Upon picking 20 more colonies, only one more was positive and again, not in the desired orientation. In retrospect, I assume the proximity of the inherent promoters within the pBluescript vector played an active role in selecting for the undesired orientation of the insert. The successful blunt end cloning of PDI into pBKS+, albeit in the undesired orientation, proved that the \textit{NcoI} site was engineered correctly and present within PDI. Using the pB-PDI clone for my insert DNA, I tried my original cloning strategy using the first PDI fragment via \textit{NcoI} – \textit{XbaI}. Even after multiple transformations employing various vector backbones (pUC, pZERO-2, and pBKS+), various \textit{psbA} 5’ UTRs, varying the ligation conditions, and checking over 60 colonies, the same trends as before were seen and all colonies picked were mutants. It was interesting to note that colonies were only seen when the PDI insert were present, however, all of the resultant colonies were mutants. These results were confirmed when this cloning step was attempted by a senior lab member known for his cloning expertise.

Due to these anomalies, the experiments were controlled more tightly allowing me to conclude that PDI was having toxic affects on the cell. There were about 4-5x fewer colonies when comparing the PDI transformations to the positive controls (digested vector [with same enzymes] + ligase). Additionally, picked colonies were positive 100% of the time on the control plates. In another experiment, there were only five colonies on the PDI plate and over 100 on the
pBg10-**AphA6-Trps16** control plate. Again, all five of the PDI clones came back negative, and all five of the tested controls were positive.

At this time, a new strategy was developed employing partial digestions of PDI (Fig. 10) allowing for the cloning of the entire PDI fragment in one step, as well as the use of a **psbA** 5’UTR without the promoter elements (**psbA** 5’UTR-P) (supplied by Chlorogen, St. Louis, MS). There were only 6 colonies from the pUC5’UTR•PDI transformation, whereas every other transformation resulted in at least 50 colonies.

**Figure 10**: Partial digestion of pB-PDI clones with *Nco*I and *Xba*I. Circled portion gel extracted for subsequent ligation.

Upon restriction digestion analysis, all six pUC5’UTR • PDI colonies were mutants and all 10 tested pB5’UTR-P • PDI were positive (Fig. 11). All controls came out as expected, however, it again is important to point out that there was still a reduction in the number of colonies even with a promoterless **psbA** 5’ UTR.
Subsequent cloning steps with PDI and the g10 leader sequence resulted in reduced transformational efficiency as seen previously, with only two positive colonies created after multiple transformations. Again, the same previous anomalies seen with the full length 5’UTR were seen when trying to clone the g10•PDI•Trps16 fragment into the final pLD vectors. This was most likely due to the activity of the 16S ribosomal RNA promoter found upstream of the MSC in the pLD vector. Evidence of this was seen when during the non-directional cloning step while creating the pLDg10•PDI•Trps16 vector. The resulting plates after transformation had 2 distinct populations of colonies, one large and one small, in a ratio of about 3:2 respectively. Upon restriction analysis, the majority of the smaller colonies appeared much fainter and had the correct insert, albeit in the opposite orientation. The larger colonies appeared to be self-ligated vectors (Fig. 12). These results were repeatedly seen in multiple types of competent cells, including Top 10, XL1 Blue, and DH5 alpha.
Figure 12: XbaI digested colonies from pLDg10 • PDI • Trps16 transformation. Lanes 1-7: smaller colonies, Lanes 9-15: larger colonies.

Figure 13: 40x amount of loaded DNA used in order to reveal positive transformants with CopyCutter cells.

To overcome the toxicity problems, competent cells ideal for cloning toxic proteins (CopyCutter EPI400, Epicentre Biotechnologies, Madison, WI) were used. These cells reduced the copy number of the plasmid until induced. The use of the copy cutter cells resulted in 12 out of 14 positive colonies, however, at a drastically reduced concentration after plasmid extraction.
In order to see the positive clones, 40x (Fig. 13) the normal amount of DNA was loaded on the gel (note the amount of RNA and higher molecular weight artifacts found when using the CopyCutter cells). It was interesting to note that any clones containing the directional PDI clone showed a much larger amount of a high molecular weight artifact (Fig. 14), also seen in the Epicentre brochure (Fig. 15).

**Figure 14**: Example showing the unusual high molecular weight artifact only within those clones containing the correct insert.

**Figure 15**: Example showing the same unusual high molecular weight artifact in the brochure for the CopyCutter cells ([http://www.epibio.com/item.asp?ID=435](http://www.epibio.com/item.asp?ID=435)). U: Uninduced, I: Induced.

This higher molecular weight artifact distorted the DNA concentration values when using a spectrophotometer, so instead, relative concentration was determined by comparing the bands against a ladder of known quantities. Any construct with g10 and PDI resulted in
extremely low concentrations, usually around 10-20 ng/µL. Many strategies were attempted in order to concentrate the extracted DNA from the CopyCutter cells, however, no method was successful in achieving near the level of DNA required for efficient and successful plant bombardment (~ 1 mg/ml). Therefore, an alternative method for amplifying toxic DNA was developed using rolling circle amplification (RCA). This method successfully amplified the gel-extracted plasmid from the previous experiments and the product was able to be digested using restriction enzymes (Fig. 16).

**Figure 16:** Agarose gel showing the digested rolling circle amplification product as well as the high molecular weight undigested product formed by the reaction. **Lanes 1-4:** Serial dilutions of RCA product digested with XbaI. **Lanes 6-8:** Undigested RCA product (0.2 µl, 0.4 µl, and 0.8 µL). *Note it is such a large complex that the majority remains within the well.*
Figure 17: *E. coli* westerns with HSA antibody. **Lane 1**: pLD5’UTRPDI, **Lane 2**: pLD5’UTRPDI• g10HSA, **Lane 3**: pLDg10HSA, **Lane 4**: pLD5’UTRHSA. Note the drastic reduction of HSA when coexpressed with PDI.

**Bacterial Westerns**

Before bombardment, *E. coli* clones were checked for correct protein expression via western blots. During these experiments PDI clones exhibited additional characteristics of toxicity. When comparing relative protein concentrations of HSA there was a drastic reduction in any clones that coexpressed PDI (Fig. 17). Additionally, there was a noticeable difference in the amount of dead cells when comparing the liquid cultures of those containing PDI. Because of cloning difficulties, only four of the six constructs were bombarded. Any clone containing PDI with the g10 leader sequence proved too problematic to gather enough clean and concentrated DNA for bombardment.
Figure 18: 3P-3M PCR analysis on the initial 30 shoots selected. (Positive control: CTB-GFP, Negative control: wild-type). Preliminary 3P-3M PCR analysis revealed no positive shoots for 5’UTR HSA (1-10), 1 positive g10 HSA (11-17), 2 positive shoots for HSA/PDI (18-26), and 2 positive shoots for 5’UTR PDI (27-30).

**PCR Confirmation**

Three to four weeks after bombardment the tobacco leaves generated ~5-10 shoots per a plate in all constructs except the rolling circle amplified g10•HSA/5’UTR•PDI, which had none. The lack of shoots with the RCA amplified construct is most likely a result of the globular nature of the reaction product. When the DNA product is one large mass rather than many little “bullets”, as in plasmid amplified DNA, the bombardment may have a drastically reduced chance of successfully recombining with chloroplast DNA. The shoots can either be chloroplast transgenic or nuclear transgenic clones, or mutant shoots, most likely a result of a modified 16S ribosomal subunit. Successful chloroplast transformation was verified using the 3P-3M and 5P-2M primer pairs (Fig. 2 and Fig. 18). Oddly, none of the 5’UTR HSA shoots were positive with 3P-3M, indicating that every shoot from that bombardment was most likely a mutant or nuclear transgenic clone.
Positive clones were passed on through a second round of selection. Again clones were checked for chloroplast integration via PCR using the 3P-3M and 5P-2M primer pairs (Fig. 19). Interestingly, all secondary clones were positive with 3P-3M, however, both PDI constructs came back negative with 5P-2M, even after many attempts (Fig. 20).

**Figure 19**: Preliminary 3P-3M PCR results from 2° plant DNA extracts. 1: g10HSA #1 (1° plant recheck), 2: g10HSA #1a, 3: g10HSA #1b, 4: HSA/PDI #1a, 5: HSA/PDI #5a, 6: 5’UTR PDI #3a, +: positive control, -: negative control.

**Figure 20**: 5P-2M PCR confirmation on 2° shoots. Lane 1-6: g10HSA #1a-f, Lanes 7-12: 5’UTR PDI #1a, 2a, 2c, 2d, 3a, 3b. Interestingly, no positive 5P-2M confirmation was seen for any of the PDI constructs.
Plant Western Blots

Western blots were used to verify protein expression in plants. Using an anti-HSA antibody (Sigma, St. Louis, MO), crude plant protein extracts were probed for the presence of HSA. HSA was found in both the g10•HSA and 5’UTR HSA plants, and perhaps faintly in the g10•HSA/5’UTR•PDI (Fig. 21). The protein was extracted from a very young plant, therefore low levels of transgenic protein expression are not uncommon.

Unfortunately, there was no specific AtPDIL5-4 antibody at the time of the study. The anti-atPDI antibody from Rose Biotechnology was raised against an Arabidopsis ER localized PDI and exhibited a high amount of cross-reactivity towards both *E. coli* and plant crude extracts. Under varied conditions, the theorized 53 kDa AtPDIL5-4 was not seen in western blots using the anti-atPDI antibody no matter how long the blot was exposed. However, the antibody did identify proteins found in transgenic lines not found in the wild-type.

To investigate this situation further, light-dark western blots were performed upon the theory that the antibody was cross-reacting with other PDI-like proteins, or possibly those involved in its regulatory pathway due to the presence of common thioredoxin (Trx) domains (Fig. 22). Western blot analysis with this antibody revealed proteins of various sizes (40, 44, 66,
68.5, and 230 kDa), with two proteins found unique to the dark. A unique band of ~44 kDa was detected in all transgenic lines of light-harvested plant protein extracts but not in the wild-type. Dark-harvested extracts exhibited 40, 44, 66, and 68.5 kDa bands found in transgenic lines but not in the wild-type.

Figure 22: Light dark western blots using anti-atPDI antibody (Rose Biotechnology, http://www.rosebiotech.com) reacting with proteins of approximately 40, 44, 66\(^1\), 68.5, and 230 kDa in size, ~20µg. 1: WT P. Havana 2: g10•HSA/5’UTR•PDI 3B g10•HSA 4: 5’UTR•HSA\(^2\) 
\(^1\) Only seen in dark-harvested leaves. \(^2\) Only light-harvested leaves available.

Figure 23: Southern blot with chloroplast flanking probe. 1: g10•HSA 2: g10•HSA/5’UTR•PDI 3: 5’UTR•PDI 4: WT.

Southern Blots

Preliminary Southern blots using a tobacco flanking probe resulted in a homoplasmic G10•HSA clone and heteroplasmic 5’UTR•PDI clone. The HSA/PDI clone seemed to be a
mutant since the transgenic DNA was undetectable in the southern blot, yet it grew in selective medium (Fig. 23). When further analyzing all PCR positive plants, all clones containing the PDI were heteroplasmic or mutants, whereas, all G10•HSA clones had only a faint wild-type band (Fig. 24).

Figure 24: Southern blot of all G10•HSA/5’UTR•PDI, 5’UTR•PDI, and G10•HSA plant DNA extracts using tobacco flanking probe.

DISCUSSION

The Importance of Plastid Sequence Information

Although various sites (trnfM/trnG, rbcL/accD, trnV/3’-rps 12, and 16S rRNA/orf 70B ) have been used to transform both potato and tomato, none of these have 100% sequence homology between the transformed genomes and tobacco flanking sequences. The lack of 100% sequence homology is very likely the major factor contributing to the reduction of efficiency in
both potato and tomato when compared to tobacco plastid transformation (Grevich and Daniell, 2005). The use of a “universal vector” is possible, but typically resulted reduced transformation frequency. For example, the transformation of *Arabidopsis*, potato and tomato chloroplast genomes have been transformed using heterologous flanking sequences (~98% homologous), but at a low frequency. In contrast, efficient transformation of tobacco, carrot, cotton, and soybean plastids with completely homologous flanking regions was achieved with high efficiency, up to 15 transgenic lines per bombarded plate in tobacco (Fernandez-San Milan *et al*., 2003). Also, there is a drastic reduction in efficiency when petunia flanking sequences (also ~98% homologous) are used to transform tobacco chloroplasts (DeGray *et al*., 2001), further supporting the theory that the lack of complete homology in flanking sequences can significantly reduce the efficiency of transformations.

Only a few crop plastid genomes have been sequenced despite the low cost and ability to sequence entire plastid genomes in a single day. Therefore, additional plastids need to be sequenced in order to facilitate chloroplast transformation in other crop species. The use of a region between tobacco and tomato or potato with 100% sequence identity (Table 6) might have enhanced recombination efficiency and thereby increased the efficiency of plastid transformation in these species. With the complete genome sequence one can pick the ideal intergenic spacer regions for transgene integration using a species-specific vector.

In addition to homologous flanking sequences, other factors may affect transformational efficiency. Monocots have proven to be much more difficult for plastid transformation compared to dicots. Better understanding of somatic embryogenesis and selection conditions that promote homoplasmy is essential for stable plastid transformation in monocots. Additionally, recent studies have highlighted the importance of using somatic embryos or
embryogenic callus for plastid transformation. Many crops, including cereals, legumes, oil
crops, cash crops, vegetable crops, fruits and nuts, beverages, and timber trees, etc. are all
regenerated in vitro via somatic embryogenesis (Grevich and Daniell, 2005). Plastid sequences
for several monocots have been available for years, including rice, wheat, and maize, yet none of
these species have been successfully transformed using plastid engineering. Therefore, plastid
genome sequences alone may not be adequate for successful plastid transformation in monocots.
In some monocots, a better understanding of DNA delivery, selection, regeneration, and
conditions that promote homoplasmy may be just as essential for successful transformation as
choosing ideal flanking sequences.

Evolutionary Analysis

Repeat Analysis

The repeat analysis identified 42 repeats shared among various members of the four
genomes (Table 1, Fig. 1). These repeats are located within genes, introns, and intergenic spacer
regions (IGS). It is interesting to note that 37 of those 42 repeats are shared among all four
genomes. Furthermore, the shared repeats are located in the exact same position. The high level
of conservation of the repeat’s location between genomes suggests that these elements may play
a functional role. Traditionally, it was thought to be uncommon for chloroplast genomes to
contain many repeats other than the large inverted repeats (Palmer, J.D., 1991). However, recent
analysis on both Solanacea and legume genomes suggest otherwise (Daniell et al., 2005; Saksi et al., 2005). Additional studies are needed to elaborate on the possible functional role of repeat elements within chloroplasts.

Coding and Non-Coding Sequence Divergence

Intergenic spacer regions (IGS) are the most widely used chloroplast markers for phylogenetic analysis with plants of lower taxonomic levels (Raubeson, L.A. and Jansen, R.K., 2005; Kelchner, S.A., 2000; Shaw et al., 2005). These markers are used because IGS regions have more variability and therefore provide more information. However, this dogma of the systematic utility of using these regions for such phylogenetic studies has been recently challenged (Kelchner, S.A., 2000). Shaw et al. compared the phlyogenetic utility of 21 noncoding chloroplast DNA regions and ranked them into three tiers based on their phylogenic utility (with tier one being the most useful for finding potentially informative characters). Furthermore, A genome-wide comparison of sequence divergence in the four Solanaceae genomes (Table 6 & 7, Fig. 6 & 7) exhibited a wide range of divergence in different regions. In this analysis, both coding and non-coding regions examined agree with the general contention that IGS regions are more variable as well as the usefulness of the relative rankings identified by Shaw et al. The genome-wide comparisons identified several IGS regions that have higher sequence divergence than even the most variable tier 1 regions. Even though the IGS regions are thought to be the most variable regions within the chloroplast genome, there are four IGS regions with 100% sequence identity between all four genomes and 21 among tomato and potato. These regions could provide useful for generating multi-species chloroplast transformational vectors. Therefore, these comparisons provide new and valuable information for the plant systematics.
community about further phylogenetic utility of chloroplast IGS regions as well as providing valuable information for plastid bioengineers.

**Vector Construction And Gene Bombardment Woes**

The toxicity of the PDI to *E. coli* vastly increased the difficulty and amount of effort required to successfully build a chloroplast transformation vector with this gene. I would imagine that other PDI like proteins may cause similar problems when cloned. Therefore, I have proposed the use of CopyCutter cells to potentially help overcome these effects. Additionally, if the resulting plasmid isolation is of a low yield, rolling circle amplification (RCA) may be used to successfully amplify the weakly concentrated plasmid. Although this method was shown to successfully amplify and digest a bacterial plasmid, further testing needs to be done to see if it can be used to successfully bombard a plant. Because RCA amplifies its products in one large complex, it may be a hit or miss event. However, in the case of a direct hit, at least 1000-fold more DNA would come in contact with that region of the chloroplast and perhaps help in achieving strong homoplasmy from the first round.

It is interesting to note the strong reactions via 3P-3M from the heteroplasmic plants, indicating that the relative strength of the PCR reaction is not directly correlated to the amount of transgenic homoplasy. Perhaps the over expression of a *psbA* regulator actually reduces the amount of D1 protein being produced since the amount of reducing equivalents remains the same, but the pool of regulators dependent upon those equivalents is now significantly increased. This could perhaps be an explanation why all clones with PDI came back heteroplasmic, whereas G10•HSA was strongly homoplasmic. Additionally, there were never any positive shoots found
from the 5’UTR-HSA bombardment, indicating that perhaps the bombardment was not done under optimal conditions and emphasizing the importance of doing multiple bombardments.

**Light/Dark Analysis**

The anti-atPDI antibody did not identify the AtPDIL5-4 protein, but did cross-react significantly with a few other proteins from crude extracts. Western blots were performed on protein extracts from both light and dark harvested leaves under the assumption that the antibody may identify other PDILs due to the presence of thioredoxin domains (trx), a common feature of both PDILs and many of their associated proteins. Analysis from both light and dark extracted proteins revealed unique expression profiles when comparing both transgenic/wild-type and light/dark crude protein extracts. A 44 kDa band was identified in all transgenic clones examined but not in the wild-type, 40 and 230 kDa bands were found in all light-harvest protein extracts. Dark-harvested transgenic extracts revealed approximately 66 and 68.5 kDa bands in addition to the 40 kDa band not found in wild-type samples. Unfortunately, none of these bands were of the size of the predicted tobacco proteins involved in *psbA* light regulation (38, 47, 55, and 60 kDa), however, these blots may perhaps reinforce future tobacco research on redox-regulation.

**Future Studies**

As was stated earlier, the bioinformatics analysis of chloroplasts is still in its infancy. Programs need to be streamlined and redesigned with a more appealing and intuitive interface than what is currently present, making it easier for less-than-savvy computer users to manage.
For example, many bioinformatics programs run from the command line and require previous Unix and programming knowledge. Also, in addition to streamlining the entire data mining process, related applications should be combined into suites, thereby eliminating the possible error created when the user must transfer or convert data from one application to another.

Although the *C. reinhardtii* system has provided much of the foundation for our understanding of the light regulated translation of *psbA* mRNA, the tobacco system has been under-utilized and could offer more insight into the functions of these mechanisms within higher plants. The availability of the complete and annotated genome sequence along with the high efficiency of recombinant expression using chloroplast engineering make it an ideal system to potentially define the individual functions of all PDIs and proteins involved in its various pathways. A similar phylogenetic analysis of the tobacco PDIs could be easily done as with *Arabidopsis* (Houston, N.L., et al. 2005), thereby paving the way for future wet lab studies in tobacco. *AtPDIL*-1-4, the RB60 homolog in *Arabidopsis* would be my number one choice for assessing the disulfide bond formation capabilities of higher plant PDIs and light regulation of *psbA* mRNAs. This is the only known Arabidopsis PDI (*AtPDIL*) protein with a homolog in any of the working models of light regulation in any plant system. Additionally, it is the only known *AtPDIL* that is definitely localized to the chloroplast. The over expression or deregulation (by mutating the redox control domain) of RB60 may also be useful for optimizing the expression of foreign proteins under the control of the *psbA* 5’UTR. Moreover, isolating the RB47 gene and reengineering a feedback insensitive version (by mutating the cysteine residues to serine within its RNA recognition motifs) should allow for consistent translation of all *psbA* mRNAs regardless of light status.
REFERENCES


Carrer, H. and Maliga, P. 1995. Targeted insertion of foreign genes into the tobacco plastid


19:71-74.


antimicrobial peptide via the chloroplast genome to control phytopathogenic bacteria and


Devine, A. and Daniell, H. 2004. Chloroplast genetic engineering for enhanced agronomic traits


Moller, S., Eds., Blackwell Publishers, United Kingdom.

Dhingra, A. and Daniell, H. In Press. Chloroplast Genetic Engineering via Organogenesis or

Somatic Embryogenesis. In: Arabidopsis Protocols II. Sanchez-Serrano, J.J.and Salina, J,

Eds., Humana Press, New Jersey.


RbcS gene restores SSU levels and photosynthesis in nuclear antisense RbcS plants. Proc.


plastid psbA, rbcL, rpl32 UTR elements by chloroplast transformation: tobacco plastid

gene expression is controlled by modulation of transcript levels and translation efficiency.


Falconer, R. 2002. Expression of interferon alpha 2b in transgenic chloroplasts of a low-nicotine


canine parvovirus in dogs. *J. Virology*. **68:**4506-4513.


