PCR-based Synthesis of Codon Optimized *cry*2Aa Gene for Production of Shoot and Fruit Borer (*Leucinodes orbonalis*) Resistant Eggplant (*Solanum melongena* L.) Cultivars

Rahul Gupta

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Masters of Science in Horticulture

Richard E. Veilleux, Chair

Gregory E. Welbaum, Co-chair

Susan Jo Martino-Catt

December 8, 2005

Blacksburg, Virginia

Keywords: *Solanum melongena*, Brinjal shoot and fruit borer, *Bacillus thuringiensis*, GFP, genetic transformation, transgenic plants, codon optimization, codon usage, gene synthesis, synthetic gene

PCR-based Synthesis of Codon Optimized *cry2Aa* Gene for Production of Shoot and Fruit Borer (*Leucinodes orbonalis*) Resistant Eggplant (*Solanum melongena* L.) Cultivars

Rahul Gupta

Abstract

Brinjal shoot and fruit borer (*Leucinodes orbonalis* Guenee) is a major limiting factor in commercial cultivation of eggplant in southeast Asia. Extensive use of pesticides as well as the conventional breeding methods have been ineffective in controlling the borer so there is a need for Integrated Pest Management (IPM) strategies for its control. *Bacillus thuringiensis* (*Bt*) is known to produce a variety of insecticidal crystal proteins toxic to lepidopteran, dipteran and coleopteran pests. The Cry2Aa protein has been found to be more toxic to brinjal shoot and fruit borer than Cry1Ab. My objective was to develop eggplant cultivars that express a codon-optimized *cry*2Aa gene, the sequence of which is based on that of an Indian isolate of *Bt*, with the eventual goal of producing fully resistant cultivars. The *cry*2Aa gene was modified for optimal expression in eggplant using the codon usage frequencies based on solanaceous sequences (eggplant, tomato and pepper). The GC content was increased from 34.3% in the native gene to 41.3% in the optimized gene, thus removing the AT-rich regions that are typical for *Bt cry* genes. Also, other mRNA destabilizing and hairpin forming structure sequences were removed. The gene was synthesized in four different parts with complementary restriction sites. A total of 152 oligonucleotides (oligos) was used to assemble the 1.9 kb gene using dual asymmetric (DA) and overlap extension (OE) PCR techniques. The individual parts were subsequently ligated

using the complementary restriction sites and inserted into vector pCAMBIA 1302. Also, the transformation efficiency of 12 different eggplant cultivars was tested using plasmid pHB2892 to predict utility for transformation with the synthetic *cry*2Aa.

Acknowledgment

First of all, I would like to express my deep gratitude to my advisor, Dr. Richard Veilleux, for his mentorship and support in development of my scientific aptitude, and for his patience during the tough times of this research. He has provided me an excellent example of how to be a great scientist and a good person.

In addition, I would especially like to thank Dr. Greg Welbaum for his suggestions, time, and help. A special thanks to Dr. Susan Martino-Catt for her invaluable advice in the entire course of my research. Also, I am greatly indebted to Dr. Jim Tokuhisa, for being available to me and boosting my morale during the down times of this research.

I would like to thank Earl Petzold and Chengsong Zhao for their technical assistance, Dr. Eric Beers, for providing plasmid pFGC5941, and Dr. John Jelesko in Fralin Biotechnology Center for providing plasmid pCAMBIA 1302.

Funding for this research was supported by the Integrated Pest Management Collaborative Research Support Program (IPM CRSP) (Project no. LAG-G-00-93-00053-00).

Dedication

This thesis is dedicated to my parents N.K Gupta and Sushma Gupta, and my brother, Anuj Gupta (Vicky) for their love and support, without them I could not make it so far.

Table of contents

List of Tables

List of Figures

Chapter 1: Introduction

Eggplant, *Solanum melongena* L., also known as brinjal, aubergine, and Guinea squash, is an economically important vegetable in southern and southeastern Asia. It is widely cultivated in China, India, Japan, Bangladesh, Indonesia, the Philippines, as well as Spain, France and Italy. Eggplant is also a significant crop in the US with an estimated harvest area of 2100 ha (FAOSTAT 2005) and farm gate value of $37 million (USDA 2003). However, each year its commercial cultivation has been seriously hampered by devastating attacks of major lepidopteran pests, particularly the Brinjal shoot and fruit borer (*Leucinodes orbonalis* Guenee, BSFB). This devastation is evident in many parts of world, particularly the drier parts of Asia where BSFB is responsible for crop losses worth billions of dollars. Heavy pesticide usage has led to pest resistant strains, increased the cost of crop production and caused environmental and health hazards from pesticide residues.

Bacillus thuringiensis based *cry* genes are a safe, specific and effective approach for pest control (Betz et al. 2000). More than 300 *cry* genes possessing insecticidal properties have been identified in *Bt*. Eggplant has been transformed with a rice optimized *cry1Ab* gene but was unable to confer complete protection against the eggplant borer to the plants (Kumar et al. 1998). However, Cry2Aa was found to be more toxic than Cry1Ab on second instar larvae of BSFB indicating this might be a more effective gene to target BSFB (Rao et al. 1999).

The focus of my thesis is the design and synthesis of an eggplant optimized *cry*2Aa gene and its use for the production of borer resistant eggplant cultivars.

Mode of damage by the borer

The borer causes damage in its larval stage (Figure 1). The life cycle of the borer is shown in Figure 2. The adult moth (Figure 3) lays eggs on leaves, buds and growing fruits (Kumar et al. 1998). The translucent larvae hatch out in 4-5 days and bore into the tender shoots and later into flowers and fruits (Cork et al. 2001). The larvae continue to feed inside the shoot or fruit for 15-20 days, pass through 4-5 larval instars and grow 10-15 mm in size before reaching the pupal stage. At the pupal stage, the larva bores out from the plant at night and pupates in soil where it develops into an adult moth in about 1 wk (Talekar 2002).

The borer is reported to infest the crop in the following ways:

1) Migration of moths from adjacent infested eggplant fields.
2) Use of infested eggplant seedlings during transplantation.
3) Pupae from previous eggplant crop can survive in soil after harvest of the crop and grow into adults in the next season. Thus, crop rotation is required as an integral component of IPM.
4) Eggplant debris may contain the pupae which can develop into adults and cause infestation in the next season.

Symptoms of damage

The appearance of wilted shoots is a peculiar symptom of insect attack because larvae feed inside shoots. Also, the presence of holes filled with frass

(insect droppings or excrement) in shoots, fruits and fruit stalks indicate the presence of insect larvae. This frass also protects the larvae from natural predators as well as applied insecticides (Cork et al. 2001). The attack makes the fruit hollow, inedible and unmarketable (Talekar 2002).

***cry* genes**

Bt is an aerobic, gram-positive entomopathogenic bacterium that produces insecticidal crystal proteins (ICPs) during its sporulation phase (Stobdan et al. 2004). *Bt* was first isolated in Japan in 1902 from diseased silkworms and was named as *Bacillus sotto*. In 1915, a German scientist named Berliner independently reisolated this bacterium from diseased larvae of Mediterranean meal moth (*Ephetia kuhniella*) (Sharma et al. 2004) and named it *Bacillus thuringiensis,* after the German province of Thuringia (Milner 1994). Since then *Bt* has been isolated from a wide variety of habitats including soil, insects, stored grain products and phyllospheres (microenvironment of a leaf) (Kaur 2000).

The ICPs (δ-endotoxins) are highly toxic at low concentrations. Toxicity of ICPs appeared to be specific to insects with little toxic effects to mammals, birds and other forms of life including human beings. The δ-endotoxins were named because of their intracellular location. ICPs have been used as biological pesticides for many years (Tounsi and Jaoua 2003). Due to these insecticidal properties, *Bt has* been widely used as a biopesticide during the past several decades.

The ICPs produce crystalline inclusion bodies during sporulation that are aggregates of large proteins called protoxins and hence are named crystal (*cry*)

genes. The first *cry* gene to be cloned was c*ry1A(a)*, by Schnepf and Whiteley (1981) and since then more than 300 *cry* genes have been identified (Crickmore et al. 1998) with 11 types of *cry*2Aa alone. Each *cry* gene has a unique insecticidal spectrum and is active only against one or a few specific insect species.

Cry genes were initially classified into different major classes based on their structural profile and the insecticidal spectrum of the proteins they encode (Hofte and Whiteley 1989). However, this nomenclature was found to be inconsistent in accommodating new genes with highly homologous sequence with the known genes but different toxin activity. The latest and the most reliable classification is based on toxin specificity. In this system, *cry* genes have been classified into five different subfamilies:

cry1: Lepidoptera

cry2: Lepidoptera and Diptera

cry3: Coleoptera

cry4: Diptera (Hofte and Whiteley, 1989)

cry 5: Lepidoptera and Coleoptera (Tailor et al. 1992)

Another group *cry6* was recognized which was active against nematodes (Feitelson et al. 1992).

Nomenclature of *cry* genes was proposed by Crickmore et al. (1998) on the basis of evolutionary divergence of the genes estimated by phylogenetic tree algorithms. Two software applications CLUSTAL W and PHYLIP were used to align the amino acid sequences and construct a phylogenetic tree of the toxins.

Three vertical boundaries marked this tree at 95 (rightmost), 78 (central) and 45% (leftmost) sequence identity. The name of a toxin depended on location of the node where it entered in this tree relative to the boundaries. If the toxin joined the tree left to the leftmost boundary, it was given a separate primary rank (Arabic number), if the toxin entered between the left and central boundary, it was assigned a separate secondary rank (uppercase letter). The toxin that entered between the central and rightmost boundary was given a separate tertiary rank (lowercase letter) and ones that joined right to the rightmost boundary were given a separate quaternary rank (another Arabic number). The toxins that were independently isolated but had identical sequence identities were also given a separate quaternary rank.

This sequence based nomenclature removed all discrepancies and allowed related toxins to be ranked together. Also, it allowed the easy identification of the role of a toxin because toxins with the same primary rank had the same toxin specificity to the same order of insects and those with different secondary and tertiary ranks had different specificities within an order.

Mode of action of *cry* genes

The ICPs are synthesized in a protoxin form and are generally about 130 kDa in size. However, *cry*2 genes are known to produce 70-71 KDa protoxins (Rukmini et al. 2000). The larvae ingest the toxin while feeding on transgenic shoots, leaves and fruit and upon ingestion, the presence of an alkaline environment inside the insect midgut causes proteolytic conversion of these protoxins into smaller toxic polypeptides of about 65-70 kDa (de Maagd et al.

1999). The toxin then binds to specific receptors present on the microvilli of the epithelial cells in the midgut of specific insects. The toxin then inserts in the membrane forming pores. These pores (ion selective channels) disturb the ionic balance between the midgut lumen and the cell cytoplasm leading to the osmotic lysis and death of midgut epithelial cells eventually killing the insect (Ferre et al. 1995).

Transformation of eggplant

Extensive research has been conducted to develop efficient protocols for eggplant transformation. Guri and Sink (1988) were the first to report *Agrobacterium*-mediated eggplant transformation. Later, different parts of the plant were used in transformation, including cotyledonary leaves from greenhouse grown plants (Rotino and Gleddie 1990), leaves from *in vitro* plants (Kumar et al. 1998), hypocotyls (Chen et al. 1995), cotyledonary explants (Kumar et al. 1998) and roots (Franklin and Sita 2003). The transformation efficiency (number of transformed plants/number of explants cocultivated) depends on the genotype used and explant. For example, none of the explants from greenhouse leaves were found to regenerate following cocultivation with *Agrobacterium* while an efficiency of 7.6% (Rotino and Gleddie 1990) was reported for the same genotype when *in vitro* leaf samples were used. Hanyu et al. (1999) reported an efficiency of 5% using leaf segments. Similarly, a low transformation efficiency of 0.1% was reported with hypocotyl explants (Chen et al. 1995). However, Franklin and Sita (2003) reported a high genotype independent transformation efficiency using root explants. A wide array of genes

including *cry*3A (Jelenkovic et al. 1998), *cry*3B (Arpaia et al. 1997; Billings et al. 1997; Blay and Oakes 1996), *cry*IAb (Jelenkovic et al. 1998; Kumar et al. 1998) and *mtlD* (Prabhavathi et al. 2002) have been introduced into eggplant cultivars to enhance insect resistance, increase disease resistance or abiotic stress tolerance. The first crop transformed to express a *cry* gene was tobacco (Jozani et al. 2005). Since then, *cry* genes have been expressed in crops including transgenic maize, tobacco, cotton, tomato, potato and rice following codon optimization (Iannacone et al. 1997).

Codon optimization

Different species differ dramatically in their preferred genetic codon composition. For enhanced expression of a foreign protein in a plant system, the codon frequency of a gene may be adjusted to the preferred codon usage frequency of the host plant.

Codon optimization is defined as an experimental strategy in which codons within a cloned gene are changed by *in vitro* mutagenesis to the preferred codons, without changing the amino acids of the synthesized protein (Zaid 1999). The difference in codon usage between prokaryotes and eukaryotes has been well documented (Fennoy and Baileyserres 1993) and genome specificity determines the codon usage frequency (Grantham et al. 1980). During the transfer of a gene from a prokaryotic to a eukaryotic genome, the prokaryote-adapted sequence motifs that are inappropriate for expression in plant systems are changed to the preferred sequence of the plant species. Also, mRNA destabilizing and degrading structures are removed (Kang et al. 2004).

The transfer of *Bt cry* genes for expression in plants requires extensive sequence modification (Kumar et al. 2005). Successful expression of *Bt* genes in plants typically requires optimization of the *cry* gene to include plant preferred codons, reduction in the overall AT content (preferred in prokaryotic genomes) of the gene by increasing the GC content (preferred in eukaryotic genomes). Also, removal of mRNA destabilizing structures like ATTTA sequences, polyadenylation signal sequences, sites of transcription termination (Misztal et al. 2004) and splicing sequences are essential for proper gene expression in a plant system. Other factors including optimization of bacterial sequences towards plant codon preference sequences using codon usage frequency tables without altering any amino acid composition are required for proper conditioning of the gene for its expression in plants.

A bias towards GC content in *Zea mays* accounted for 40% variation in codon usage (Fennoy and Baileyserres 1993). Similarly, a change in the GC codon frequency from 32 to 60% in a gene encoding GFP from *Aequorea victoria* showed an increase in the GFP expression (Rouwendal et al. 1997) and a 100-fold increase in expression of *cry* genes was reported following optimization for tomato (Perlak et al. 1991).

Earlier transformation work done on eggplant using unoptimized *cry* genes failed to produce adequate gene expression. Transformed eggplant developed by Chen et al. (1995) using the wild type (wt) *Bt cry*3B gene failed to show any insect resistance. This can be due to non-optimization of the gene sequence in addition to other factors including transgene copy number leading to

transcriptional or post-transcriptional gene silencing, position effect of transgene integration site, regulatory sequences including promoter and terminator (Butaye et al. 2005). Also, post-translational factors, such as protein stability, modification and trafficking play crucial role in expression of a heterologous gene (Dai et al. 2005). Similar examples of no or low gene expression without proper codon optimization have been reported in potato (Cheng et al. 1992), tobacco (Barton et al. 1987) and tomato (Fischhoff et al. 1987). This low expression was likely due to a low level of *Bt*-specific mRNA caused by the presence of destabilizing motifs in wt *Bt* genes and non-optimization of *Bt* gene sequences for plant expression. However, when *cry*1Ab was optimized for eggplant using a codon frequency table for rice, expression was also low (Kumar et al. 1998). Use of a monocot (rice) as a reference for optimizing the gene for eggplant expression could be the reason for poor expression. In the present work, the *cry2Aa* gene was optimized using the codon usage frequencies of more related crops (tomato and pepper) (Nakamura et al. 2000) in addition to eggplant to attempt to improve expression of the gene for insect resistance.

PCR-mediated gene synthesis

Proper codon optimization requires considerable changes that can be located throughout the gene sequence of interest. Different methods have been followed to make these codon optimized synthetic genes. Some of these methods include the Fok I method (Mandecki and Bolling 1988), assembly by constructing separate sections of the gene from smaller duplex oligonucleotide fragments, its subcloning in a shuttle vector and ligation of fragments to give a

full length gene (Casimiro et al. 1997) and PCR-based methods (Stemmer et al. 1995). Other methods, such as site-directed mutagenesis have also been used to inculcate only a few changes that are grouped together at certain locations in the wild type gene. In most cases when prokaryotic genes are prepared for expression in eukaryotes, complete gene synthesis is considered a better approach for efficient expression of the gene rather than making localized changes. The FokI method is considered to be an efficient gene synthesis approach with an error frequency of only one per 4000 bp of DNA cloned but it is a lengthy and labor-intensive process. Construction of synthetic genes by ligation of oligos has long been practiced. It involves the pooling of overlapping synthetic phosphorylated oligos and ligating them using a shot-gun approach (Grundstrom et al. 1985). However, this approach can result in large scale mismatches and duplex formation, decreasing the yield significantly. Artificially synthesized gene blocks have been assembled using complementary restriction end sites (Adang et al. 1993). This process requires identification of many unique restriction sites along the gene sequence that may require compromising proper codon optimization. Another popular strategy is PCR-based gene synthesis. Different types of PCR programs like dual asymmetric (DA PCR) (Sandhu et al. 1992), overlap extension (OE PCR) and recursive PCR increase both the product yield and efficiency.

Dual asymmetric (DA) PCR

DA PCR (Sandhu et al. 1992) uses four adjacent oligos with short overlaps between them as primers for the PCR reaction. The amount of the

central two oligos is much lower than the flanking outer primers. The first PCR reaction causes an asymmetric amplification of the two halves of the sequences that overlap each other. These dual asymmetrically amplified fragments have overlaps that may be used to construct a complete gene when followed by an OE PCR to amplify the rest of the sequence.

Overlap extension (OE) PCR

This approach was first described by Higuchi et al. (1988) for site specific mutagenesis of DNA fragments. OE PCR is used to combine the PCR products from different templates by adding an overlapping sequence from one template onto the 5' end of second template and mixing the templates together for amplification by PCR. Thus it allows joining of different templates without requiring compatible restriction sites at the ends.

Young and Dong (2004) combined DA PCR and OE PCR methods to devise a simple and highly efficient two-step approach to gene synthesis. Oligos of 25 bp length with 15 and 10 bp overlaps with adjacent oligos were used for DA PCR. The products from DA PCR were further pooled for OE PCR. The OE PCR product was subsequently amplified by end primers to obtain the final product. Young (2004) used this method for the synthesis of an *Escherichia coli* optimized 470 bp proinsulin gene and three other genes of 1.1 and 1.2 kb size.

Purpose

BSFB (*L. orbonalis*) is a major pest of eggplant and is responsible for crop losses worth billions of dollars. Heavy pesticide use (25-30 sprays per season) is

ineffective for the control of the pest and has led to development of pest resistant strains, increasing the cost of crop production and environmental and health hazards from pesticide residues. *Bt* based *cry* genes are a safe, specific and effective approach for pest control. My research is focused on the use of *Bt* derived *cry*2Aa gene for production of the borer resistant eggplant cultivars.

To achieve this objective various short term objectives were formulated:

1. Codon optimization of the native version of *cry*2Aa gene using the codon usage database of eggplant and related solanaceous species, (tomato and pepper) to enhance expression of the modified version of the gene.
2. PCR based synthesis of the optimized gene and its cloning into an expression cassette.
3. Testing the transformation efficiency of different cultivars with a GFP construct to predict utility for transformation with the modified version of *cry*2Aa.

References

Adang MJ, Brody MS, Cardineau G, Eagan N, Roush RT, Shewmaker CK, Jones A, Oakes JV, McBride KE (1993) The reconstruction and expression of a *Bacillus thuringiensis cry*3A gene in protoplasts and potato plants. Plant Mol Biol 21: 1131-1145.

Arpaia S, Mennella G, Onofaro V, Perri E, Sunseri F, Rotino GL (1997) Production of transgenic eggplant (*Solanum melongena* L.) resistant to Colorado Potato Beetle (*Leptinotarsa decemlineata* Say). Theor Appl Genet 95: 329-334.

Barton KA, Whiteley HR, Yang NS (1987) *Bacillus thuringiensis* delta endotoxin expressed in transgenic *Nicotiana tabacum* provides resistance to lepidopteran insects. Plant Physiol 85: 1103-1109.

Betz FS, Hammond BG, Fuchs RL (2000) Safety and advantages of *Bacillus thuringiensis* protected plants to control insect pests. Regul Toxicol Pharm 32: 156-173.

Billings S, Jelenkovic G, Chin CK, Eberhardt J (1997) The effect of growth regulators and antibiotics on eggplant transformation. J Am Soc Hort Sci 122: 158-162.

Blay E, Oakes JV (1996) *Agrobacterium tumefaciens*-mediated transformation of *Solanum gilo* raddi as influenced by explant type. Plant Cell Rep 15: 582-585.

Butaye KMJ, Cammue BPA, Delaure SL, De Bolle MFC (2005) Approaches to minimize variation of transgene expression in plants. Mol Breed 16: 79-91.

Casimiro DR, Wright PE, Dyson HJ (1997) PCR-based gene synthesis and protein NMR spectroscopy. Structure 5: 1407-1412.

Chen Q, Jelenkovic G, Chin CK, Billings S, Eberhardt J, Goffreda JC, Day P (1995) Transfer and transcriptional expression of coleopteran *cry*3B endotoxin gene of *Bacillus thuringiensis* in eggplant. J Am Soc Hort Sci 120: 921-927.

Cheng J, Bolyard MG, Saxena RC, Sticklen MB (1992) Production of insect resistant potato by genetic-transformation with a delta-endotoxin gene from *Bacillus thuringiensis* var kurstaki. Plant Sci 81: 83-91.

Cork A, Alam SN, Das A, Das CS, Ghosh GC, Farman DI, Hall DR, Maslen NR, Vedham K, Phythian SJ, Rouf FMA, Srinivasan K (2001) Female sex pheromone of brinjal fruit and shoot borer, *Leucinodes orbonalis* blend optimization. J Chem Ecol 27: 1867-1877.

Crickmore N, Zeigler DR, Feitelson J, Schnepf E, Van Rie J, Lereclus D, Baum J, Dean DH (1998) Revision of the nomenclature for the *Bacillus thuringiensis* pesticidal crystal proteins. Microbiol Mol Biol Rev 62: 807-13.

Dai ZY, Hooker BS, Quesenberry RD, Thomas SR (2005) Optimization of *Acidothermus cellulolyticus* endoglucanase (E1) production in transgenic tobacco plants by transcriptional, post-transcription and post-translational modification. Transgenic Res 14: 627-643.

de Maagd RA, Bakker P, Staykov N, Dukiandjiev S, Stiekema W, Bosch D (1999) Identification of *Bacillus thuringiensi*s delta-endotoxin *Cry*1C domain III amino acid residues involved in insect specificity. Appl Environ Microbiol 65: 4369-4374.

FAOSTAT (2005) http://faostat.fao.org/faostat/servlet/XteServlet3?Areas=231&Items=399&Elements=31&Years=2004&Format=Table&Xaxis=Years&Yaxis=Countries&Aggregate=&Calculate=&Domain=SUA&ItemTypes=Production.Crops.Primary&language=EN.

Feitelson JS, Payne J, Kim L (1992) *Bacillus thuringiensis* - Insects and beyond. Bio/Technology 10: 271-275.

Fennoy SL, Baileyserres J (1993) Synonymous Codon Usage in *Zea mays* L nuclear genes is varied by levels of C-ending and G-ending codons. Nucleic Acids Res 21: 5294-5300.

Ferre J, Escriche B, Bel Y, Vanrie J (1995) Biochemistry and genetics of insect resistance to *Bacillus thuringiensis* insecticidal crystal proteins. FEMS Microbiol Lett 132: 1-7.

Fischhoff DA, Bowdish KS, Perlak FJ, Marrone PG, Mccormick SM, Niedermeyer JG, Dean DA, Kusanokretzmer K, Mayer EJ, Rochester DE, Rogers SG, Fraley RT (1987) Insect tolerant transgenic tomato plants. Bio/Technology 5: 807-813.

Franklin G, Sita GL (2003) *Agrobacterium tumefaciens* mediated transformation of eggplant (*Solanum melongena* L.) using root explants. Plant Cell Rep 21: 549-554.

Grantham R, Gautier C, Gouy M (1980) Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. Nucleic Acids Res 8: 1893-1912.

Grundstrom T, Zenke WM, Wintzerith M, Matthes HWD, Staub A, Chambon P (1985) Oligonucleotide-directed mutagenesis by microscale shot-gun gene synthesis. Nucleic Acids Res 13: 3305-3316.

Guri A, Sink KC (1988) *Agrobacterium* transformation of eggplant. J Plant Physiol 133: 52-55.

Hanyu H, Murata A, Park EY, Okabe M, Billings S, Jelenkovic G, Pederson H, Chin C-K (1999) Stability of luciferase gene expression in a long term period in transgenic eggplant, *Solanum melongena*. Plant Biotechnol 16: 403-407.

Higuchi R, Krummel B, Saiki RK (1988) A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. Nucleic Acids Res 16: 7351-67.

Hofte H, Whiteley HR (1989) Insecticidal crystal proteins of *Bacillus thuringiensis*. Microbol Rev 53: 242-255.

Iannacone R, Grieco PD, Cellini F (1997) Specific sequence modifications of a *cry*3B endotoxin gene result in high levels of expression and insect resistance. Plant Mol Biol 34: 485-496.

Jelenkovic G, Billings S, Chen Q, Lashomb J, Hamilton G, Ghidiu G (1998) Transformation of eggplant with synthetic *cry*3A gene produces a high level of resistance to the Colorado potato beetle. J Am Soc Hort Sci 123: 19-25.

Jozani GRS, Komakhin RA, Piruzian ES (2005) Comparative study of the expression of the native, modified, and hybrid *cry*3a genes of *Bacillus thuringiensis* in prokaryotic and eukaryotic cells. Russ J Genet 41: 116-121.

Kang TJ, Loc NH, Jang MO, Yang MS (2004) Modification of the cholera toxin B subunit coding sequence to enhance expression in plants. Mol Breed 13: 143-153.

Kaur S (2000) Molecular approaches towards development of novel *Bacillus thuringiensis* biopesticides. World J Microbiol Biotechnol 16: 781-793.

Kumar KK, Maruthasalam S, Loganathan M, Sudhakar D, Balasubramanian P (2005) An improved *Agrobacterium* mediated transformation protocol for recalcitrant elite indica rice cultivars. Plant Mol Biol Rep 23: 67-73.

Kumar PA, Mandaokar A, Sreenivasu K, Chakrabarti SK, Bisaria S, Sharma SR, Kaur S, Sharma RP (1998) Insect-resistant transgenic brinjal plants. Mol Breed 4: 33-37.

Mandecki W, Bolling TJ (1988) *Fok*i method of gene synthesis. Gene 68: 101-107.

Milner RJ (1994) History of *Bacillus thuringiensis*. Agric Ecosyt Eviron 49: 9-13.

Misztal LH, Mostowska A, Skibinska M, Bajsa J, Musial WG, Jarmolowski A (2004) Expression of modified *Cry*1Ac gene of *Bacillus thuringiensis* in transgenic tobacco plants. Mol Biotechnol 26: 17-26.

Nakamura Y, Gojobori T, Ikemura T (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucleic Acids Res 28: 292.

Ooi P Ecoport - The consilience engine. http://ecoport.org/.

Perlak FJ, Fuchs RL, Dean DA, McPherson SL, Fischhoff DA (1991) Modification of the coding sequence enhances plant expression of insect control protein genes. Proc Natl Acad Sci USA 88: 3324-3328.

Prabhavathi V, Yadav JS, Kumar PA, Rajam MV (2002) Abiotic stress tolerance in transgenic eggplant (*Solanum melongena* L.) by introduction of bacterial mannitol phosphodehydrogenase gene. Mol Breed 9: 137-147.

Rao NGV, Majumdar A, Mandaokar AD, Nimbalkar SA, Kumar PA (1999) Susceptibility of brinjal shoot and fruit borer to the delta-endotoxins of *Bacillus thuringiensis*. Curr Sci 77: 336-337.

Rotino GL, Gleddie S (1990) Transformation of eggplant (*Solanum melongena* L) using a binary *Agrobacterium tumefaciens* vector. Plant Cell Rep 9: 26-29.

Rouwendal GJA, Mendes O, Wolbert EJH, deBoer AD (1997) Enhanced expression in tobacco of the gene encoding green fluorescent protein by modification of its codon usage. Plant Mol Biol 33: 989-999.

Rukmini V, Reddy CY, Venkateswerlu G (2000) *Bacillus thuringiensis* crystal delta-endotoxin: Role of proteases in the conversion of protoxin to toxin. Biochimie 82: 109-116.

Sandhu GS, Aleff RA, Kline BC (1992) Dual asymmetric PCR - one-step construction of synthetic genes. Biotechniques 12: 14-16.

Schnepf HE, Whiteley HR (1981) Cloning and expression of the *Bacillus thuringiensis* crystal protein gene in *Escherichia coli*. Proc Nat Acad Sci USA 78: 2893-2897.

Sharma HC, Sharma KK, Crouch JH (2004) Genetic transformation of crops for insect resistance: Potential and limitations. Crit Rev Plant Sci 23: 47-72.

Stemmer WPC, Crameri A, Ha KD, Brennan TM, Heyneker HL (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. Gene 164: 49-53.

Stobdan T, Kaur S, Singh A (2004) Cloning and nucleotide sequence of a novel *cry* gene from *Bacillus thuringiensis*. Biotechnol Lett 26: 1153-1156.

Tailor R, Tippett J, Gibb G, Pells S, Pike D, Jordan L, Ely S (1992) Identification and characterization of a novel *Bacillus thuringiensis* delta-endotoxin entomocidal to coleopteran and lepidopteran larvae. Mol Microbiol 6: 1211-1217.

Talekar NS (2002) Controlling eggplant fruit and shoot borer: A simple, safe and economical approach In: AVRDC International Cooperators' Guide. Asian Vegetable Research and Development Center, Shanhua, Taiwan, pp 1-4.

Tounsi S, Jaoua S (2003) Characterization of a novel *cry2Aa*-type gene from *Bacillus thuringiensis* subsp. *kurstaki*. Biotechnol Lett 25: 1219-23.

USDA (2003) Vegetables 2002 summary, Agricultural Statistics Board, National Agricultural Statistics Service. p20.

Young L, Dong QH (2004) Two-step total gene synthesis method. Nucleic Acids Res 32.

Zaid A, Hughes H.G, Porceddu E, Nicholas F (1999) Glossary of biotechnology and genetic engineering. FAO 7.

Figure 1. Full grown larva of the eggplant shoot and fruit borer (*L. orbonalis*). Full grown larva grows between 10-15 mm and represents the most damaging stage of the borer (Ooi).

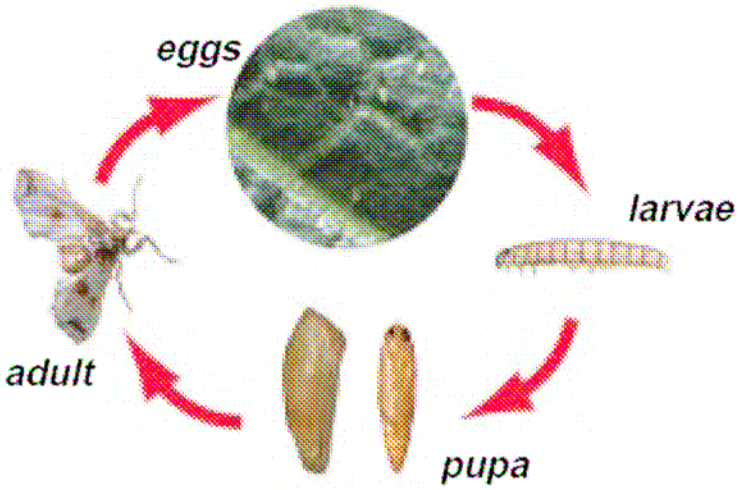

Figure 2. . Life cycle of eggplant shoot and fruit borer (*L. orbonalis*) (Talekar 2002). One life cycle is completed in 25-32 days. Multiple generations have been reported on tropical crops

Figure 3. Female (left) and male (right) adult moths of *L. orbonalis*. Females are slightly larger and have a blunt abdomen (Ooi).

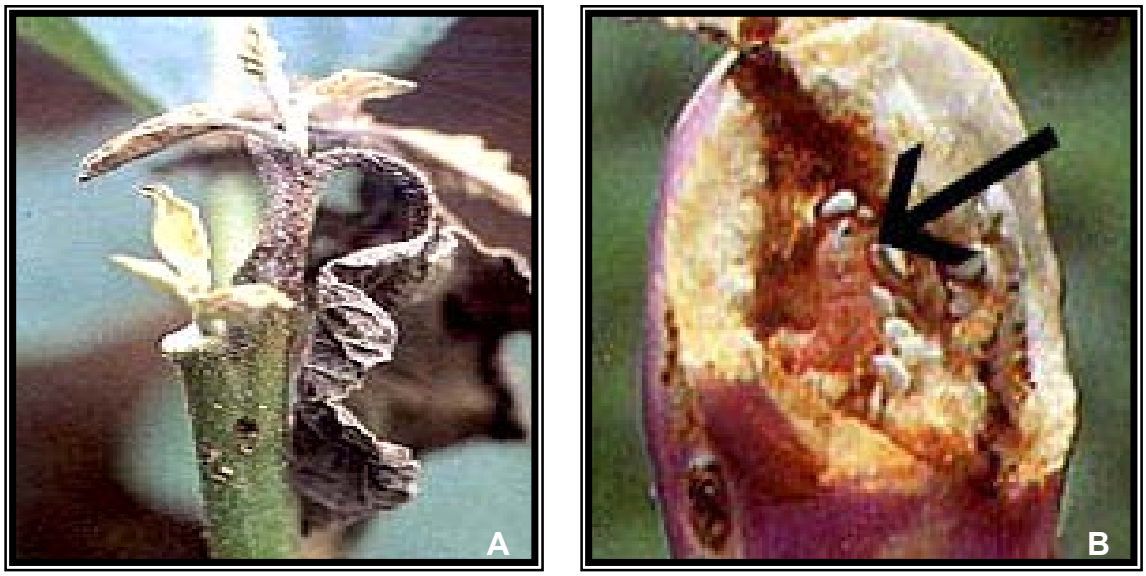

Figure 4. Symptoms of attack: A) Wilted shoot; B) Infested fruit which is rendered inedible (Talekar 2002).

Chapter 2: Transformation of eggplant

Introduction

Solanum melongena L. (eggplant) is an important vegetable crop in Asia, Africa, Europe and North America (Kashyap et al. 2003) and is native to the Indian sub-continent. It is particularly important in the Indian subcontinent where a major proportion of the population is vegetarian. However, production of the crop in this area is seriously hampered by a major lepidopteran pest, *Leucinodes orbonalis* Guenee (Brinjal shoot and fruit borer, BSFB). Agrochemicals have been ineffective in controlling the borer. Conventional breeding methods are slow and ineffective due to deleterious gene linkages of wild relatives with desirable traits (Franklin and Sita 2003). Thus alternate Integrated Pest Management (IPM) strategies might prove to be effective in improving eggplant cultivation.

Genetic engineering involving the transfer and expression of *Bacillus thuringiensis* derived *cry* genes into plants has resulted in improved insect resistance in several crops including tobacco (Barton et al. 1987), tomato (Fischhoff et al. 1987), cotton (Perlak et al. 1990), maize (Koziel et al. 1993), potato (Adang et al. 1993) and soybean (Macrae et al. 2005). Application of similar technology in eggplant might also be useful for improvement in its commercial cultivation by introduction of novel genes into the elite background in a single event. Such genetic engineering approaches can help to overcome the constraints of traditional breeding methods and the need to identify and introduce genes from wild relatives of the crop.

A prerequisite for efficient transfer of genes into any heterologous system is the availability of a suitable transformation system. *Agrobacterium* has been widely used for development of efficient transformation method for transferring genetic information into more than a 100 plant species (Babu et al. 2003) over the last 20 years (Gelvin 2003). Eggplant also has been frequently transformed with *Agrobacterium* since 1988 (Guri and Sink 1988) and several reporter genes including *gus* (β-glucouronidase) (Billings et al. 1997); (Chen et al. 1995), *luciferase* (Komari 1989), *gfp* (Horowitz et al. 2002) and *gfp:gus* fusion genes (Kumar and Rajam 2005) have been used in conjunction with *npt*II as the selectable marker gene to improve *Agrobacterium* mediated transformation efficiency. In the present research, *gfp* and *nptII* were used for the first time to test eggplant transformation efficiency.

In the present research, we attempted to transform 12 different eggplant cultivars using the transformation protocol developed by Rotino and Gleddie (1990) to select the cultivars with the highest transformation and regeneration efficiency, to be transformed at a later date with artificially synthesized *cry*2Aa gene.

Materials and methods

Plant Material

The seeds of 12 cultivars (Table 1) of eggplant from three different countries – USA, India and the Philippines – obtained from the Plant Genetic Resources Conservation Unit (Griffin, Georgia). Two cultivars named Turbo and Domino that are known to have natural resistance to the borer were also

obtained from Thailand (East-West Seed Co., Inc., Bulacan, Philippines). These cultivars varied in fruit size, shape and color. A contrast in shape, size and color of the fruits is shown in Figure 5.

Seed Sterilization

Two seed sterilization techniques were used in the present research:

1) Twenty seeds were surface sterilized by dipping in 20% commercial bleach (Clorox) + 2-3 drops of 'Tween 20' for 20 min followed by five rinses for 5 min each and an overnight soak in water. These seeds were then treated with 10% Clorox for 15 min followed by three rinses of 5 min each (Chen et al. 1995).
2) The seeds were dipped in PPM (plant preservative mixture) (Plant Cell Technology, Inc, Washington, DC) solution for 1 min.

Seed Germination

The surface sterilized seeds were grown in 10 ml Murashige and Skoog (MS) basal medium (Murashige and Skoog 1962) in flat bottom test tubes (2.5 X 10 cm) capped with Magenta 2-way caps and sealed with parafilm. These tubes were kept at 25°C and 16 h light and 8 h dark period. However, the seeds treated with PPM were kept in continuous dark at 25°C until germination.

Preparation of Agrobacterium

Agrobacterium culture was preserved in Protect Bacterial Preservation System (Technical Service Consultant, Lancashire, England). One colony of *Agrobacterium* grown for 48 h at 28°C was picked by a sterile loop and a thick

suspension was made in the tube containing 20-25 treated beads in a cryopreservation fluid. After mixing the culture with the beads, this fluid was withdrawn and the tubes were stored at -20°C to preserve the bacteria.

One bead of *Agrobacterium tumefaciens* strain LBA 4404 from these tubes was used for growing *Agrobacterium* by dipping it in 40 ml LB medium supplemented with 50 mg/l kanamycin and 30 mg/l streptomycin on an incubator shaker at 28°C for 48 h at 250 rpm, until OD_{600} = 0.5.

Transformation

Agrobacterium tumefaciens strain LBA 4404 containing plasmid pHB2892 (Figure 6) was obtained from Molinier et al. (2000) and was used for transformation of eggplant. The plasmid pHB2892 contains *S-GFP* as the reporter gene controlled by a double 35S cauliflower mosaic virus promoter (CaMV 35S) and *npt* II (confers kanamycin resistance) as a selectable marker gene. The *S-GFP* gene has been optimized for human codon usage by Chalfie et al. (1994). Eggplant transformation was carried out using two types of explants: the cotyledonary leaves and the stumps. The following two procedures were followed for transformation:

1) *Leaf Disc Method*: Cotyledons collected after excision from seedlings (10 days old) were cut into 1 x 1 cm leaf discs with a sterilized scalpel and were cocultivated with *Agrobacterium tumefaciens* culture for 20 min. The discs were blotted dry on Whatman filter paper and allowed to grow on regeneration medium in petri plates (100 × 15 mm) for 2 days with the abaxial side in contact with the regeneration medium (Figure 7). The regeneration

medium (20 ml per petri plate) contained MS basal salts supplemented with sucrose (20 g/l), myoinositol (100 mg/l), thiamine (10 mg/l), gelrite (2 g/l) and zeatin (1 mg/l) which was added after autoclaving the medium. The pH of the medium was adjusted to 5.8 prior to autoclaving at 121°C for 45 min. After 48 h, leaf discs were transferred to selection medium to suppress bacterial growth. Selection medium is regeneration medium with the addition of kanamycin (100 mg/l) and cefotaxime (500 mg/l). Leaf discs were placed on fresh selection medium every 2 wks. The developing callus was observed for GFP expression under a fluorescence microscope after 4 wks of regeneration (Figure 8). When the callus was approximately 5 mm in diameter (6 wks), it was excised from the leaves and subcultured in petri plates (60 x 15 mm) containing the selection medium (100 mg/l kanamycin) and a lower concentration of cefotaxime (200 mg/l) (Rotino and Gleddie 1990). Developing callus was subcultured 2-3 times to fresh medium every 2 wks for shoot development.

The shoots developing from callus were observed for GFP expression under a fluorescent microscope (Figure 9). The transformed shoots were rooted on 10 ml MS medium containing IBA (500 μg/l) in flat bottom test tubes (2.5 X 10 cm) capped with Magenta 2-way caps and were sealed with parafilm. Rooting begin in 1 wk of transfer and the rooted shoots were transferred to the greenhouse for acclimatization (process of an organism adjusting to changes in its environment) (Chen et al. 1995). A selection of S-GFP expressing plants of Black Beauty are shown in Figure 10. These plants

were observed under an Olympus fluorescence microscope SZX-RFL3 (Olympus America, Melville, NY, USA) with a 100 W mercury burner (Olympus BH2-RFL-T3) and filter sets of GFP longpass (filter set number: 4018, Chroma Technology Corp, Rockingham, VT, USA). This lens gave a clear distinction between the GFP expressing and non-expressing parts of the plant by causing GFP expressing parts of the plant to appear green while the non-GFP chlorophyll expressing parts glowed red (Figure 10). The flowers of the primary tranformants (T_0) thus obtained were selfed to produce the T_1 progeny which were subsequently tested for GFP expression.

2) *Transformation of the stumps*: - The freshly cut stems (10 days old) were treated with a thick culture of *Agrobacterium tumefaciens* prepared after spinning 20 ml overnight grown culture in a microcentrifuge for 1 min. The stumps were allowed to regenerate for 2 days after which they were again treated with selection medium to restrict the growth of *Agrobacterium tumefaciens* (Figure 11) (Chen et al. 1995).

GFP testing of primary transformants

The growing callus was observed under a fluorescent microscope every 2 wks to cull the escape shoots and select GFP expressing ones. Selected shoots subsequently grew into plants which were transferred to the greenhouse. GFP expression in greenhouse plants was quantified with a GFP meter that is commercially available (Opti-sciences, Tyngsboro, MA).The GFP meter uses a fiber optic cable that detects fluorescence by emitting and detecting specific

wavelengths of light, i.e., 395 nm/495 nm (emit) and 530 nm (detect). It records GFP expression in relative units called tics.

Younger fully expanded leaves were selected for GFP analysis. Readings were taken from the leaf lamina, the midrib, the opposite side of the lamina, opposite side of midrib, flower buds (if present), open flowers and fruits. Control Black Beauty plants were used to record autofluorescence. Three readings were taken per tissue per plant and comparisons among plants were analyzed by SAS ANOVA.

PCR analysis

DNA samples from GFP expressing eggplant plants were extracted from young leaf tissue by the CTAB method (Doyle and Doyle 1987). PCR analysis was performed using the following *npt* II primers: forward primer, 5' GTCTGTGGCGGGAACTCC; reverse primer 5'GCGATGATCTCGTCGTGAC. These primers amplify a 600 bp fragment of the T-DNA sequence of pHB2892 plasmid.

The PCR reaction was performed in a 50 µl reaction containing the template genomic DNA (100 ng), forward and reverse primers (20 pmol each), dNTPs (200 µM), PCR buffer 1× (50 mM KCl, 1.1 mM $MgCl_2$, 10mM Tris-HCl pH 8.3 and 0.01% gelatin) and 1U REDTaq DNA polymerase (Sigma Aldrich, St. Louis, MO). Amplifications were carried out in a Robocycler Gradient 40 thermocycler (Stratagene, La Jolla, CA) using the following parameters:

1 cycle = 5 min 95°C; 35 cycles = 45 sec 95°C, 30 sec 53°C, 1 min 30 sec 72°C; 1cycle = 10 min 72°C. A10 µl aliquot from each reaction was used for electrophoresis on a 1% (w/v) agarose gel containing 5 µl ethidium bromide.

Results

Transformation results

The transformation efficiency of 12 different cultivars was evaluated using two different transformation methods (Table 2). Transformation efficiency was calculated as the ratio of number of transformants (GFP expressing explants) to the number of explants (leaf discs + cotyledons or stumps) cultivated.

Leaf explants started showing callus growth within 2 wks after cocultivation and friable callus could be seen growing profusely along the edges of the cotyledons and first primary leaves without any distinguishing shoots after 3-4 wks. Shoots appeared after 2-3 subcultures (10-12 wks post-inoculation) and rooted plants at 14-16 wks post-inoculation were transferred to the greenhouse. Explants from the negative control (explants not treated with *Agrobacterium* culture) did not regenerate, turned brown, and died within 1 wk after transfer to selective medium.

The transformed callus of Pusa Purple Long died due to contamination. However, transformed callus of 014 and Black Beauty developed shoots. These shoots were excised and transferred to root induction medium. The shoots from a single independent callus were considered as a single transformation event.

Five plants of Black Beauty were transferred to the greenhouse, of which three developed from different leaf discs (BB 1-T, BB 2-T and BB 3-T) and thus

were independent transformation events. BB 1-T callus developed two shoots (BB 1-1T and BB1-2T), BB 2-T also produced two shoots (BB 2-1T and BB 2-2T) and BB 3-T developed one shoot, i.e., BB 3-1T.

GFP expression results

Callus was observed under a fluorescent microscope 4 wks post-infection and continued every 2 wks. GFP expressing callus could be easily distinguished. The non-GFP parts did not glow and could not be seen. Of 12 cultivars used for transformation, four (014, Black Beauty, Pusa Purple Long and New Hampshire) developed GFP expressing callus (Figure 8). Only Black Beauty and 014 developed shoots (Figure 9). The GFP expression in greenhouse plants was estimated using a GFP meter. ANOVA revealed that genotype was a significant source of variation for S-GFP expression in all four tissues of T_0 plants (Table 4). The GFP expression in two Black Beauty plants, i.e., BB2-1T and BB3-1T was significantly greater than other three Black Beauty plants (Figure 12).

Integration of T-DNA in T_0 plants

Five T_0 Black Beauty plants were screened by PCR analyses for the integration of pHB2892 T-DNA in the plant genome using *npt* II primers. BB2-1T and BB3-1T, which were significantly greater in GFP expression, produced a 600 bp fragment corresponding to the *npt* II gene (Figure 13). This indicated the integration of *npt*II in plant genome. The other three putative transformants did not yield a PCR product. Further, southern blot analysis is required to confirm the genomic integration of pHB2892 T-DNA in the plant genome.

GFP expression in T_1 progeny

T_1 progeny was obtained by selfing BB 2-1T and BB 3-1T. Seeds from both transformed plants were collected and expression was measured in dry seeds, germinating seeds, seed coat, radicle and plumule. Strong GFP expression was detected in the radicle and plumule of the germinating seed while the seed coat did not show any expression (Figure 14).

Chi square analysis of 14 T_1 seedlings of BB2-1T and 13 T_1 seedlings of BB3-1T showed no significant difference between observed ratios (11:3 and 8:5 GFP^+/GFP^-, respectively) and the expected 3:1 Mendelian ratio for a single insert hemizygous transgene (Table 3).

Of the seedlings transferred to the greenhouse, nine progeny plants from each of the parents (BB2-1 T1-9 and BB3-1 T1-9) grew to maturity and were used to determine differences in GFP expression among the plants. Data were obtained from newly opened leaves and mature leaves at apical, middle and proximal ends of the leaves. A general increase in GFP expression was observed from apical towards the proximal end of the leaf. Also, newly developed leaves exhibited higher GFP expression than mature leaves (Figures 15 and 16).

ANOVA on nine BB2-1T1-9 plants (Table 5) showed that genotype was a significant source of variation for GFP expression of both new and mature leaves. Mean expression of 400 tics appeared to be the threshold value between non-GFP expressing and GFP expressing plants in new leaves while this value decreased to 200 in mature leaves. However, the plants having higher mean

expression in new leaves were consistently higher expressing in mature leaves as well.

Similar ANOVA analysis on nine BB3-1T1-9 plants (Table 6) showed a mean GFP expression threshold level of 225.3. Only three out of six plants showed consistency in expression between the new and mature leaves.

Discussion

GFP has been reported to reduce transformation efficiency in barley (Murray et al. 2004) and proper optimization is necessary for its expression in plants. Rouwendal et al. (1997) optimized GFP for expression in tobacco by increasing GC content on the third position in codons from 32 to 60% which increased the expression of high expressors in the transgene expressing plants without observing any increase in final GFP concentration. This was attributed to interference of high GFP expression with regeneration ability and thus low or minimum observed regeneration of transformed cells which have high GFP concentration. Similarly, Haseloff et al. (1997) reported GFP to interfere with regeneration in *Arabidopsis*. This interference of GFP is more likely due to transgene silencing and post-translational factors such as protein insability, degradation during modification and protein trafficking effects (Dai et al. 2005). In contrast, Molinier et al. (2000) used a human codon optimized *S-GFP* for expression in tobacco and analyzed its expression up to T_2 generation without observing any negative effects. Also, Joshi et al. (2005) has used enhanced green fluorescent protein (EGFP) and enhanced yellow fluorescent protein (EYFP) for expression in peanut. Similar results were obtained in the present

work of eggplant transformation. No reduction in transformation efficiency was observed. In fact, a significantly higher efficiency (11.4%; Black Beauty and 8.6%; Pusa Purple Long) was reported with GFP using the same protocol where *npt* II was used to report the integration of transgene in the eggplant genome (7.6%; Picentia) (Rotino and Gleddie 1990).

In the present study, expression of GFP was found to decrease with age of the plant. Strong GFP expression could be seen in germinating seedlings. However, this signal decreased as seedlings developed true leaves and even more as whole plants in the greenhouse. Similar results have been reported by Zhou et al. (2005) in rice, *Arabidopsis* and *Medicago*. Thus, it can be concluded that GFP expression is strongly age and genotype dependent and chlorophyll pigment interference should be taken in consideration while observing in vivo expression of the reporter gene.

Transformation and regeneration efficiency has been found to be genotype dependent in crops including cereals, i.e., rice (Hoque et al. 2005; Kumar et al. 2005), wheat (Huber et al. 2002) and barley (Murray et al. 2004) as well as solanaceous crops, i.e., tomato (Ellul et al. 2003) and potato (Heeres et al. 2002). There are no reports of previous work on comparison of different eggplant cultivars for their transformation and regeneration efficiency. In the present research, we tested 12 different cultivars from three different countries for the genotypic and transformation procedure dependence of transformation efficiency. Black Beauty and Pusa Purple Long were found exhibit the best efficiencies of 11.4% and 8.6%, respectively. No success was found with other

cultivars including Arka Sheel, Long Green and Cluster Purple. Thus, transformation efficiency was found to be genotype dependent in eggplant. Beauty has been previously used by Guri and Sink (1988) while Pusa Purple Long was used by Kumar et al. (1998) for production of *L. orbonalis* resistant eggplant. Earlier transformation work on eggplant by Rotino and Gleddie (1990) reported a 7.6% transformation efficiency following same protocol but cultivar Picentia. An increase in transformation efficiency in the present work can be attributed to the cultivar used for transformation. Thus, we propose that Black Beauty and Pusa Purple Long can be efficiently used for further transformation work and molecular analysis of eggplant.

Of the five Black Beauty T_0 plants, two were found to be positive for the *npt*II gene suggesting at least partial integration of pHB2892 T-DNA in the plant genome. This integration of T-DNA in the plant genome of BB2-1T and BB3-1T explains a significantly higher GFP expression during GFP quantification in the greenhouse. The other three plants, i.e., BB1-1T, BB1-2T and BB2-2T are suggested to be escapes (plants with no T-DNA insertion but they survive the antibiotic treatment during the tissue culture practices).

If we were to attempt to classify the T_1 progenies into genetic categories (non-transgenic:hemizygous:homozygous) from the data in Figures 15 and 16, the ratios on new leaves would be 3:2:4 and 3:3:3 for BB2-1T and BB3-1T, respectively, based on statistically different gene expression measured in new leaves using the GFP meter. If intermediate GFP expression is taken as expression of hemizygosity, there would appear to be a deficiency of

hemizygotes in both families. However, family size is small and GFP expression in the mature leaves is not in complete agreement, although generally it is similar. Chi-square analysis did not reveal a statistically significant difference between observed and expected (1:2:1) frequencies in either family ($\chi^2 = 3.0$, $p < 0.23$ and $\chi^2 = 1$, $p < 0.61$, respectively, for BB2-1T and BB3-1T). Further analysis of T_2 generation is required for segregation ratio analysis to estimate the gene inheritance.

References

Adang MJ, Brody MS, Cardineau G, Eagan N, Roush RT, Shewmaker CK, Jones A, Oakes JV, McBride KE (1993) The reconstruction and expression of a *Bacillus thuringiensis cry*3A gene in protoplasts and potato plants. Plant Mol Biol 21: 1131-1145.

Babu RM, Sajeena A, Seetharaman K, Reddy MS (2003) Advances in genetically engineered (transgenic) plants in pest management - an over view. Crop Prot 22: 1071-1086.

Barton KA, Whiteley HR, Yang NS (1987) *Bacillus thuringiensis* delta endotoxin expressed in transgenic *Nicotiana tabacum* provides resistance to lepidopteran insects. Plant Physiol 85: 1103-1109.

Billings S, Jelenkovic G, Chin CK, Eberhardt J (1997) The effect of growth regulators and antibiotics on eggplant transformation. J Am Soc Hort Sci 122: 158-162.

Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC (1994) Green fluorescent protein as a marker for gene expression. Science 263: 802-805.

Chen Q, Jelenkovic G, Chin CK, Billings S, Eberhardt J, Goffreda JC, Day P (1995) Transfer and transcriptional expression of coleopteran *cry*3B endotoxin gene of *Bacillus thuringiensis* in eggplant. J Am Soc Hort Sci 120: 921-927.

Dai ZY, Hooker BS, Quesenberry RD, Thomas SR (2005) Optimization of *Acidothermus cellulolyticus* endoglucanase (E1) production in transgenic tobacco plants by transcriptional, post-transcription and post-translational modification. Transgenic Res 14: 627-643.

Doyle JJ, Doyle JV (1987) A rapid DNA isolation procedure for small amounts of leaf tissue. Phytochem Bull 19: 810-815.

Ellul P, Garcia-Sogo B, Pineda B, Rios G, Roig LA, Moreno V (2003) The ploidy level of transgenic plants in *Agrobacterium* mediated transformation of tomato cotyledons (*Lycopersicon esculentum* L.Mill.) is genotype and procedure dependent. Theor Appl Genet 106: 231-238.

Fischhoff DA, Bowdish KS, Perlak FJ, Marrone PG, Mccormick SM, Niedermeyer JG, Dean DA, Kusanokretzmer K, Mayer EJ, Rochester DE, Rogers SG, Fraley RT (1987) Insect tolerant transgenic tomato plants. Bio/Technology 5: 807-813.

Franklin G, Sita GL (2003) *Agrobacterium tumefaciens* mediated transformation of eggplant (*Solanum melongena* L.) using root explants. Plant Cell Rep 21: 549-554.

Gelvin SB (2003) *Agrobacterium*-mediated plant transformation: the biology behind the "gene-jockeying" tool. Microbiol Mol Biol Rev 67: 16-37.

Guri A, Sink KC (1988) *Agrobacterium* transformation of eggplant. J Plant Physiol 133: 52-55.

Haseloff J, Siemering KR, Prasher DC, Hodge S (1997) Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic *Arabidopsis* plants brightly. Proc Nat Acad Sci USA 94: 2122-2127.

Heeres P, Schippers-Rozenboom M, Jacobsen E, Visser RGF (2002) Transformation of a large number of potato varieties: genotype-dependent variation in efficiency and somaclonal variability. Euphytica 124: 13-22.

Hoque ME, Mansfield JW, Bennett MH (2005) *Agrobacterium* mediated transformation of indica rice genotypes: an assessment of factors affecting the transformation efficiency. Plant Cell Tiss Org Cult 82: 45-55.

Horowitz S, Freeman S, Sharon A (2002) Use of green fluorescent protein-transgenic strains to study pathogenic and nonpathogenic lifestyles in *Colletotrichum acutatum*. Phytopathology 92: 743-749.

Huber M, Hahn R, Hess D (2002) High transformation frequencies obtained from a commercial wheat (*Triticum aestivum* L. cv. 'Combi') by microbombardment of immature embryos followed by GFP screening combined with PPT selection. Mol Breed 10: 19-30.

Joshi M, Niu C, Fleming G, Hazra S, Chu Y, Nairn CJ, Yang HY, Ozias-Akins P (2005) Use of green fluorescent protein as a non-destructive marker for peanut genetic transformation. In Vitro Cell Dev Biol Plant 41: 437-445.

Kashyap V, Kumar SV, Collonnier C, Fusari F, Haicour R, Rotino GL, Sihachakr D, Rajam M (2003) Biotechnology of eggplant. Sci Hort 97: 1-25.

Komari T (1989) Transformation of callus cultures of nine plant species mediated by *Agrobacterium*. Plant Sci 60: 223-229.

Koziel MG, Beland GL, Bowman C, Carozzi NB, Crenshaw R, Crossland L, Dawson J, Desai N, Hill M, Kadwell S, Launis K, Lewis K, Maddox D, McPherson K, Meghji MR, Merlin E, Rhodes R, Warren GW, Wright M, Evola SV (1993) Field performance of elite transgenic maize plants expressing an insecticidal protein derived from *Bacillus thuringiensis*. Bio/Technology 11: 194-200.

Kumar KK, Maruthasalam S, Loganathan M, Sudhakar D, Balasubramanian P (2005) An improved *Agrobacterium* mediated transformation protocol for recalcitrant elite indica rice cultivars. Plant Mol Biol Rep 23: 67-73.

Kumar PA, Mandaokar A, Sreenivasu K, Chakrabarti SK, Bisaria S, Sharma SR, Kaur S, Sharma RP (1998) Insect-resistant transgenic brinjal plants. Mol Breed 4: 33-37.

Kumar SV, Rajam MV (2005) Enhanced induction of *vir* genes results in the improvement of *Agrobacterium* mediated transformation of eggplant. J Plant Biochem Biotechnol 14: 89-94.

Macrae TC, Baur ME, Boethel DJ, Fitzpatrick BJ, Gao AG, Gamundi JC, Harrison LA, Kabuye VT, McPherson RM, Miklos JA, Paradise MS, Toedebusch AS, Viegas A (2005) Laboratory and field evaluations of transgenic soybean exhibiting high-dose expression of a synthetic *Bacillus thuringiensis cry*1A gene for control of Lepidoptera. J Econ Entomol 98: 577-587.

Molinier J, Himber C, Hahne G (2000) Use of green fluorescent protein for detection of transformed shoots and homozygous offspring. Plant Cell Rep 19: 219-223.

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol Plant 15: 473-497.

Murray F, Brettell R, Matthews P, Bishop D, Jacobsen J (2004) Comparison of *Agrobacterium*-mediated transformation of four barley cultivars using the *GFP* and *GUS* reporter genes. Plant Cell Rep 22: 397-402.

Perlak FJ, Deaton RW, Armstrong TA, Fuchs RL, Sims SR, Greenplate JT, Fischhoff DA (1990) Insect resistant cotton plants. Biotechnology (NY) 8: 939-943.

Rotino GL, Gleddie S (1990) Transformation of eggplant (*Solanum melongena* L) using a binary *Agrobacterium tumefaciens* vector. Plant Cell Rep 9: 26-29.

Rouwendal GJA, Mendes O, Wolbert EJH, deBoer AD (1997) Enhanced expression in tobacco of the gene encoding green fluorescent protein by modification of its codon usage. Plant Mol Biol 33: 989-999.

Zhou X, Carranco R, Vitha S, Hall TC (2005) The dark side of green fluorescent protein. New Phytol 168: 313-321.

Table 1: Eggplant cultivars used for transformation and their countries of origin

Cultivar	Country
New Hampshire	United States
Florida Special	United States
Black Beauty	India
Long Green	India
New York Improved	India
Cluster Purple	India
Pusa Purple Long	India
Arka Sheel	India
Tarung	Philippines
CA-64-017	Philippines
014	Philippines
Dingras3	Philippines

Table 2: Transformation efficiency results of 12 different cultivars collected from the USA, India and the Philippines using two different transformation protocols, leaf disc inoculation and stump inoculation, with *Agrobacterium tumefaciens* strain LBA 4404 and plasmid vector pHB2892

Cultivar	Seeds planted	Seeds germinated	Leaf discs + cotyledon pieces		Stumps		Transformation efficiency (%)	
			Cultured	Transformed	Cultured	Transformed	Leaf discs+ Cotyledons	Stumps
Pusa Purple Long	10	10	70	6	10	0	8.6	0.0
Florida Special	10	1	0	0	0	0	0.0	0.0
Dingras 3	10	10	65	0	10	0	0.0	0.0
014	20	19	120	6	19	0	5.0	0.0
New Hampshire	20	20	220	15	20	0	6.8	0.0
Black Beauty	20	19	175	20	19	0	11.4	0.0
Tarung	20	16	140	0	16	0	0.0	0.0
NewYork Improved	20	15	125	0	15	0	0.0	0.0
CA-64-017	20	18	150	0	18	0	0.0	0.0
Cluster Purple	20	20	95	8	20	0	8.4	0.0
Long Green	20	17	135	0	17	0	0.0	0.0
Arka Sheel	20	12	65	0	12	0	0.0	0.0

Table 3: Chi square values of segregation ratios in T_1 progeny seedlings of two independent regenerates of Black Beauty. Phenotypic characterization of T_1 seedlings was based on GFP expression observed under fluorescent microscope

Genotype	df	Expected ratio (GFP expressing:non-expressing)	Observed ratio (GFP expressing:non-expressing)	Chi square
BB2-1T	1	3:1	11:3	0.095ns
BB3-1T	1	3:1	8:5	1.256ns

Table 4: ANOVA of S-GFP expression based on tic values measured with a GFP meter in adaxial and abaxial sides of leaves, midribs and flower buds of T_0 plants. Four plants in the greenhouse were used for analysis and data were taken from three leaves per plant. A non-transformed Black Beauty plant was used as the negative control. Asterisks; *, ** denote significant source of variation at $p < 0.05$ and $p < 0.01$, respectively.

Source		S-GFP expression in adaxial side of leaves	S-GFP expression in abaxial side of leaves	S-GFP expression in midribs	S-GFP expression in flower buds
	df	MS	MS	MS	MS
Genotype	4	169950**	576771**	255612*	1718168**
Error	10	12165	52139	58697	87889

Table 5: ANOVA of S-GFP expression based on tic values measured with a GFP meter in new and mature leaves of T_1 progeny plants of GFP expressing BB 2-1T. Nine plants in the greenhouse were used for analysis and data were taken from apical, middle and proximal ends of adaxial side of three leaves per plant. A non-transformed Black Beauty plant was used as the negative control. Asterisks (**) following a value denotes a significance at $p < 0.0001$.

Source		**S-GFP expression in new leaves**	**S-GFP expression in mature leaves**
	df	**MS**	**MS**
Genotype	**9**	**2497706****	**657471****
Error	**20**	**63651**	**7256**

Table 6: ANOVA of S-GFP expression based on tic values measured with a GFP meter in new and mature leaves of T_1 progeny plants of GFP expressing BB 3-1T. Nine plants in the greenhouse were used for analysis and data were taken from apical, middle and proximal ends of adaxial side of three leaves per plant. A non-transformed Black Beauty plant was used as the negative control. Asterisks (**) following a value denotes a significance at $p < 0.0001$.

Source		**S-GFP expression in new leaves**	**S-GFP expression in mature leaves**
	df	**MS**	**MS**
Genotype	**9**	**450944****	**333430****
Error	**20**	**8634**	**157554**

CA-64-017 (The Philippines)

New Hampshire (USA)

Long Green (India)

Cluster purple (India) **Black Beauty (India)**

Figure 5. Fruit of five eggplant cultivars used in transformation studies showing variation in their size, shape and color. Fruit shown are fully mature at seed harvest stage rather than at their more typical culinary use stage.

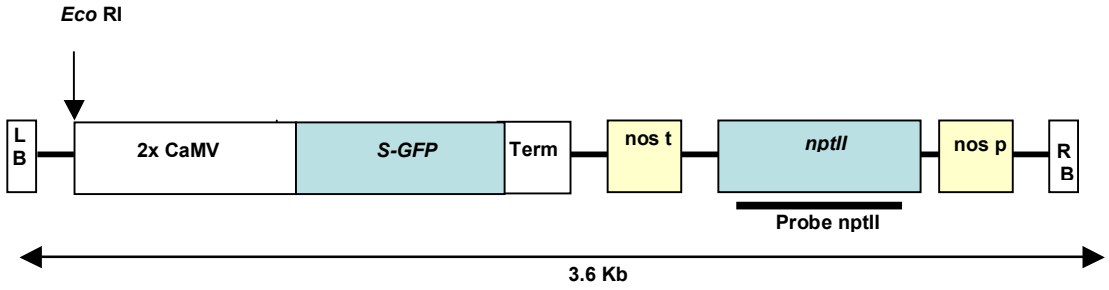

Figure 6. Structure of introduced T-DNA from plasmid pHB2892 (Molinier et al. 2000). Expression of S-GFP was controlled by 2x CaMV promoter which gave strong expression in transformed plants. A 0.6 kb fragment of the *nptII* gene was used as probe for southerns.

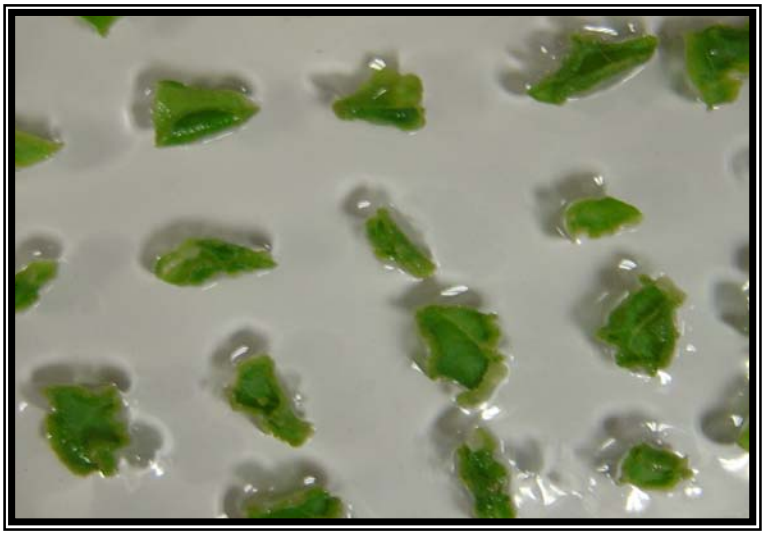

Figure 7. Leaf disc regeneration method. Callus growth of *Solanum melongena* cv. Pusa Purple Long, 1 wk after inoculation is shown. Cotyledons from 10-15 day old seedlings were cut into 1 × 1 cm segments and treated with *Agrobacterium* strain LBA 4404 containing pHB2892 vector.

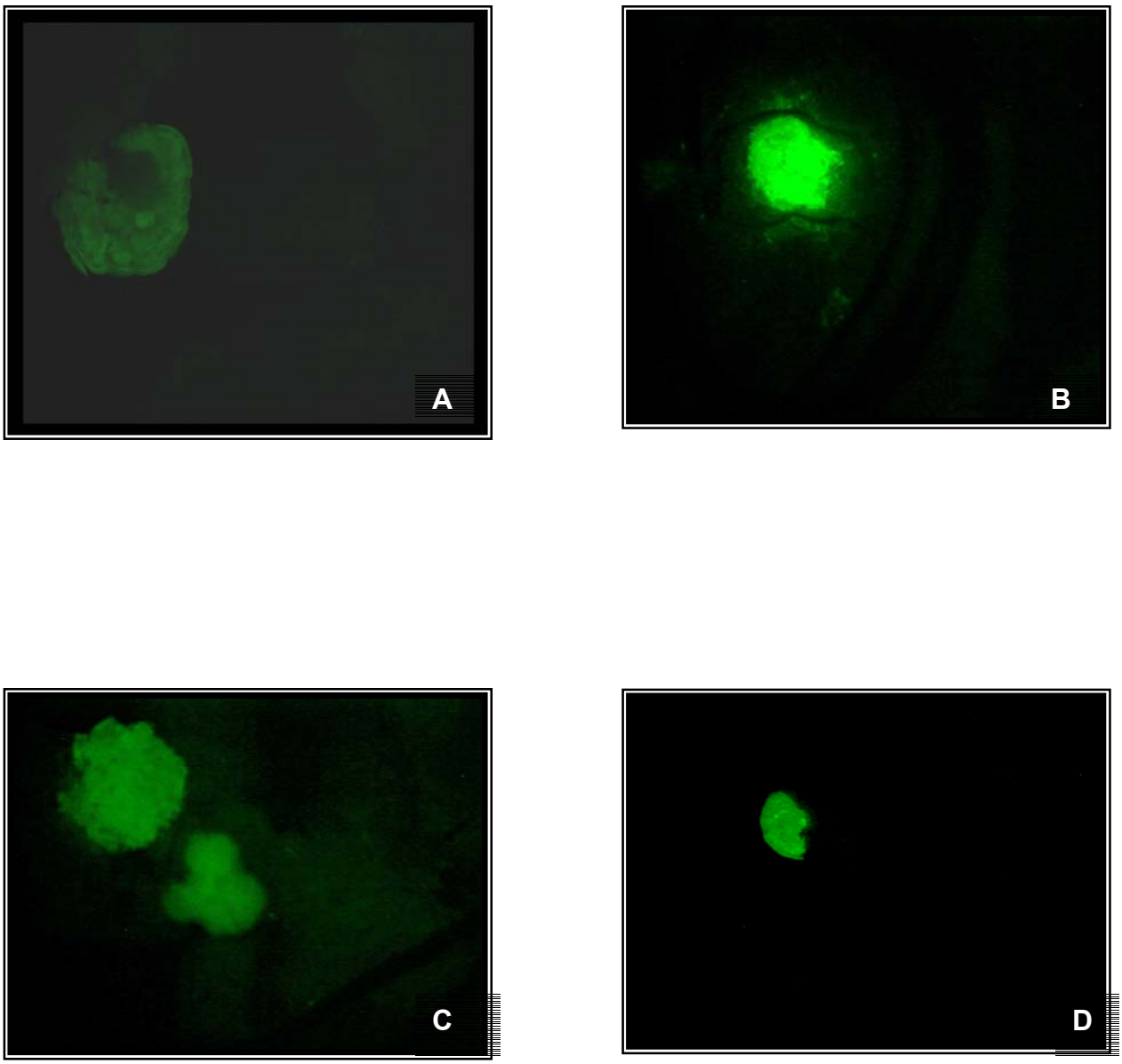

Figure 8. S-GFP expression in developing callus of: A) Pusa Purple Long; B) New Hampshire; C) Black Beauty; D) 014 observed under an Olympus fluorescence microscope SZX-RFL3 (Olympus America, Melville, NY, USA) with a 100 W mercury burner (Olympus BH2-RFL-T3) and filter sets of GFP longpass (filter set number: 4018, Chroma Technology Corp, Rockingham, VT, USA).

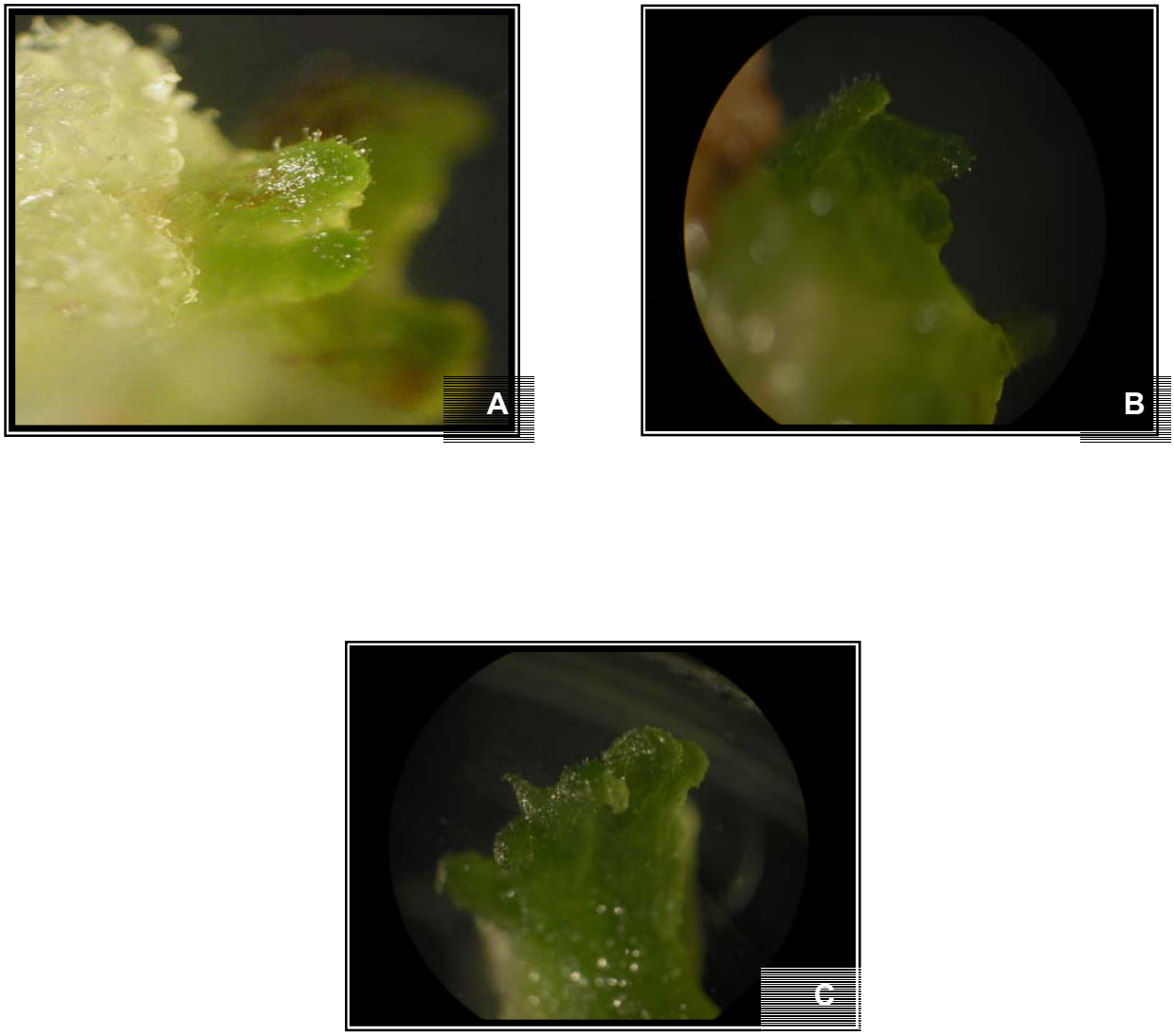

Figure 9. S-GFP expression observed in developing shoots on leaf explants of: A) and B) Black Beauty; C) 014; 10 wks after inoculation with *Agrobacterium* strain LBA 4404 under a fluorescence microscope.

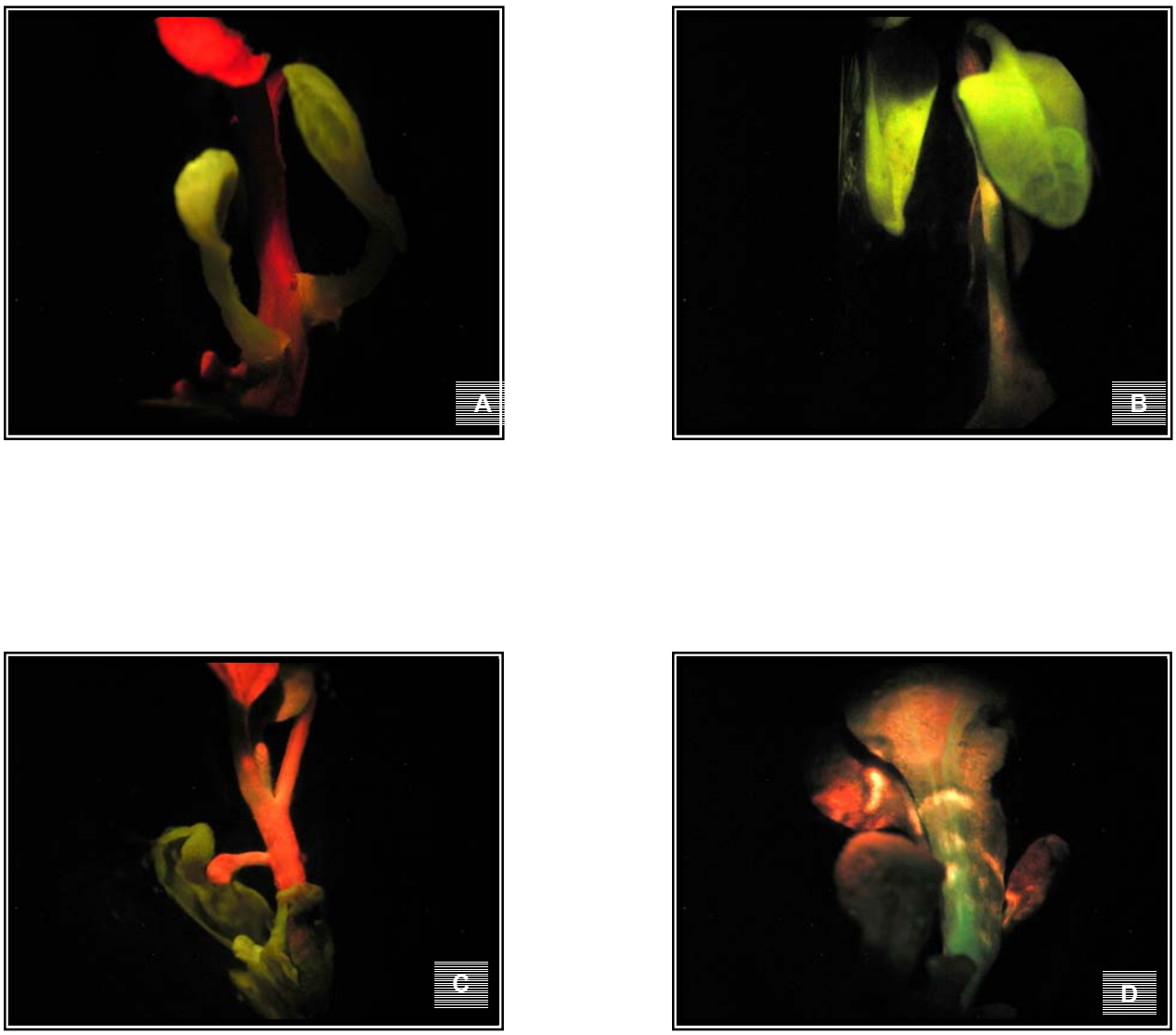

Figure 10. S-GFP expressing T_0 plants of A) and B) BB2-1T; C) and D) BB3-1T. GFP expressing tissues appear green under long pass lens while chlorophyll (non-GFP) expressing tissues appear red.

Figure 11. Stump regeneration method. Epicotyls of 10-15 day old seedlings were excised and the hypocotyls were treated with *Agrobacterium* paste. Regenerants from the top of hypocotyls were expected to express GFP if transformed with plasmid vector pHB2892.

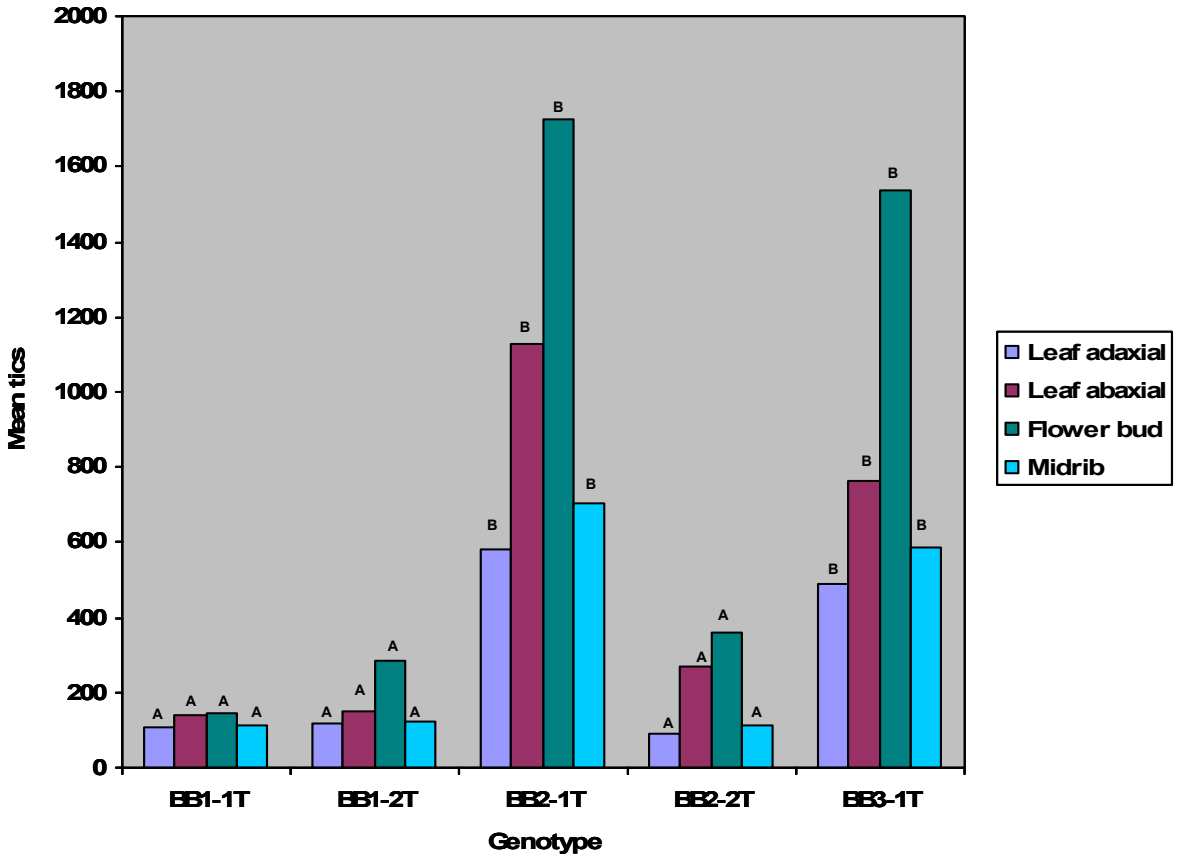

Figure 12. Mean GFP expression (n=3) measured in tics using a GFP meter for five T_0 plants. BB2-1T and BB3-1T showed a significantly greater GFP expression compared with three other plants.

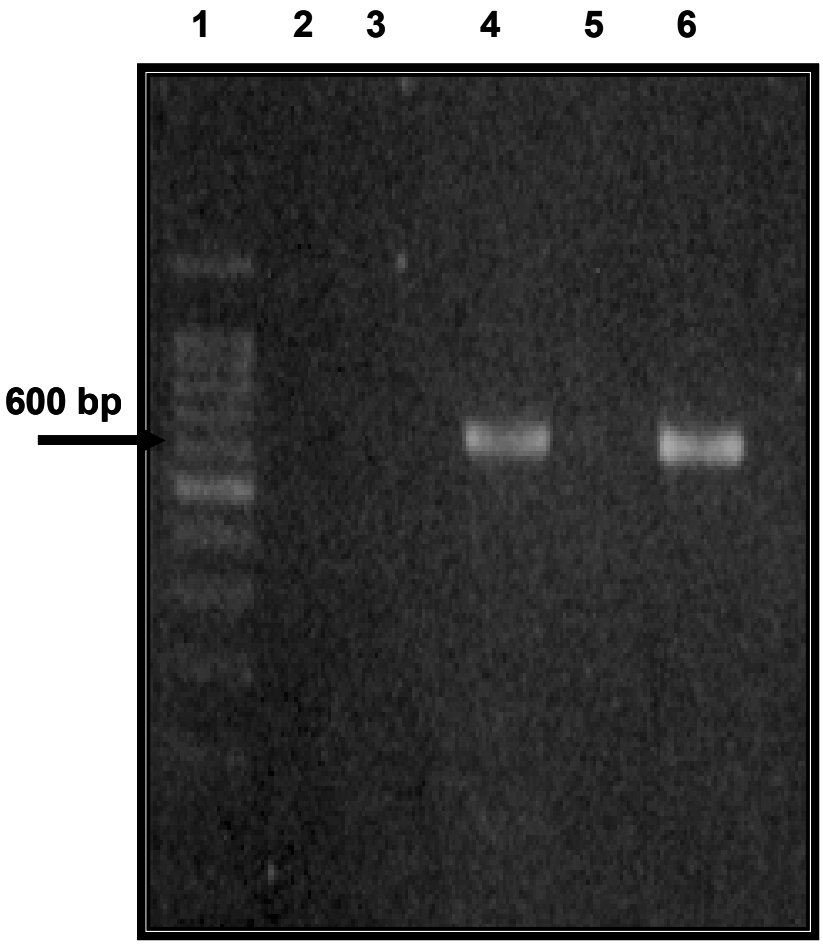

Figure 13. PCR detection of integration of T-DNA segment of pHB2892 in T_0 Black Beauty plants. The amplification products were analysed on 2% agarose gel. Lane 1: 100bp DNA ladder; 2: BB1-1T; 3: BB1-2T; 4: BB2-1T; 5: BB2-2T; 6: BB3-1T. A 600 bp PCR product for npt II primers was generated only for BB2-1T and BB3-1T plants which were significantly higher in GFP expression than other plants.

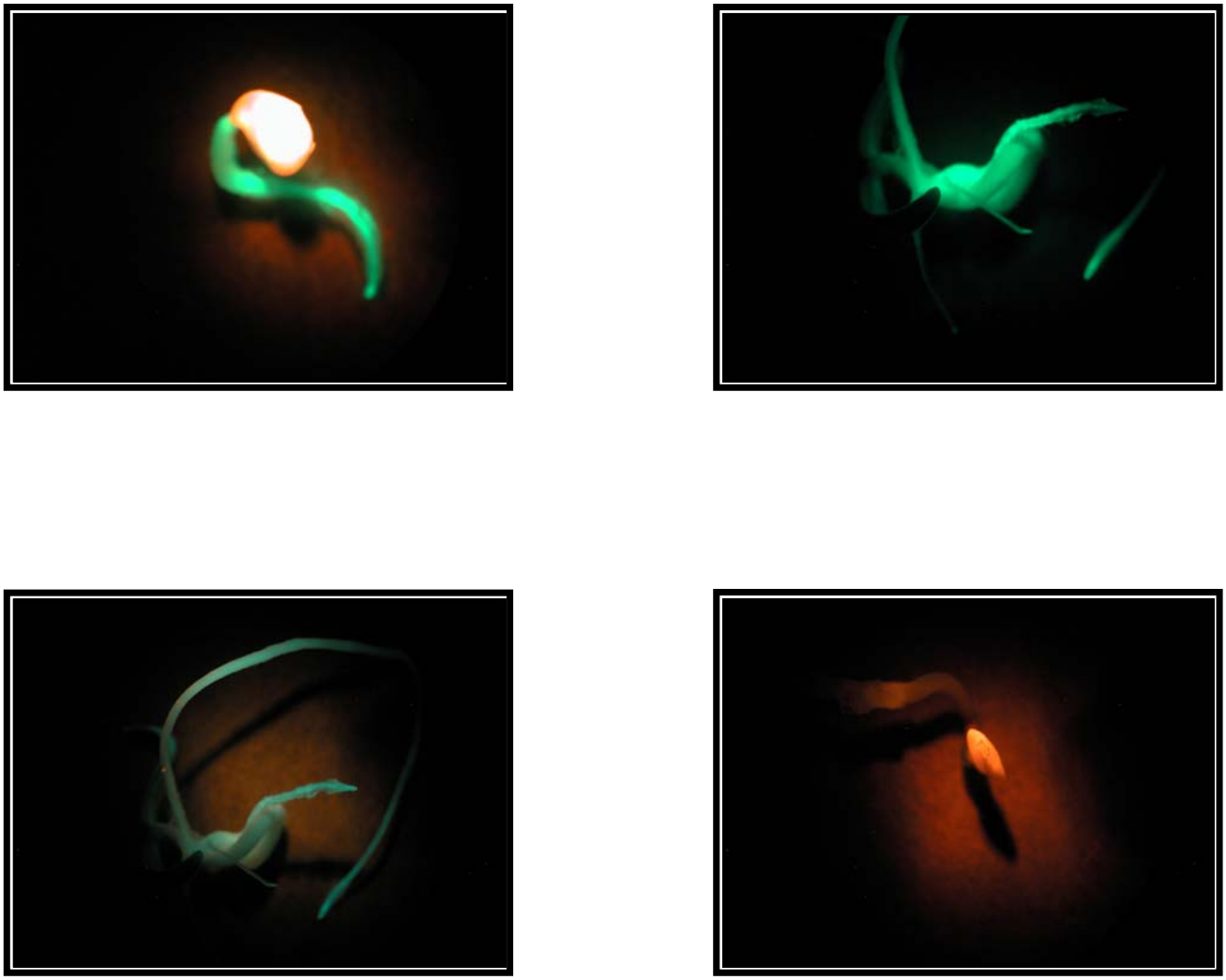

Figure 14. S-GFP expression in the seedcoat, plumule and radicle of T_1 generation seedlings by fluorescence microscopy. Non-GFP seedlings appear red.

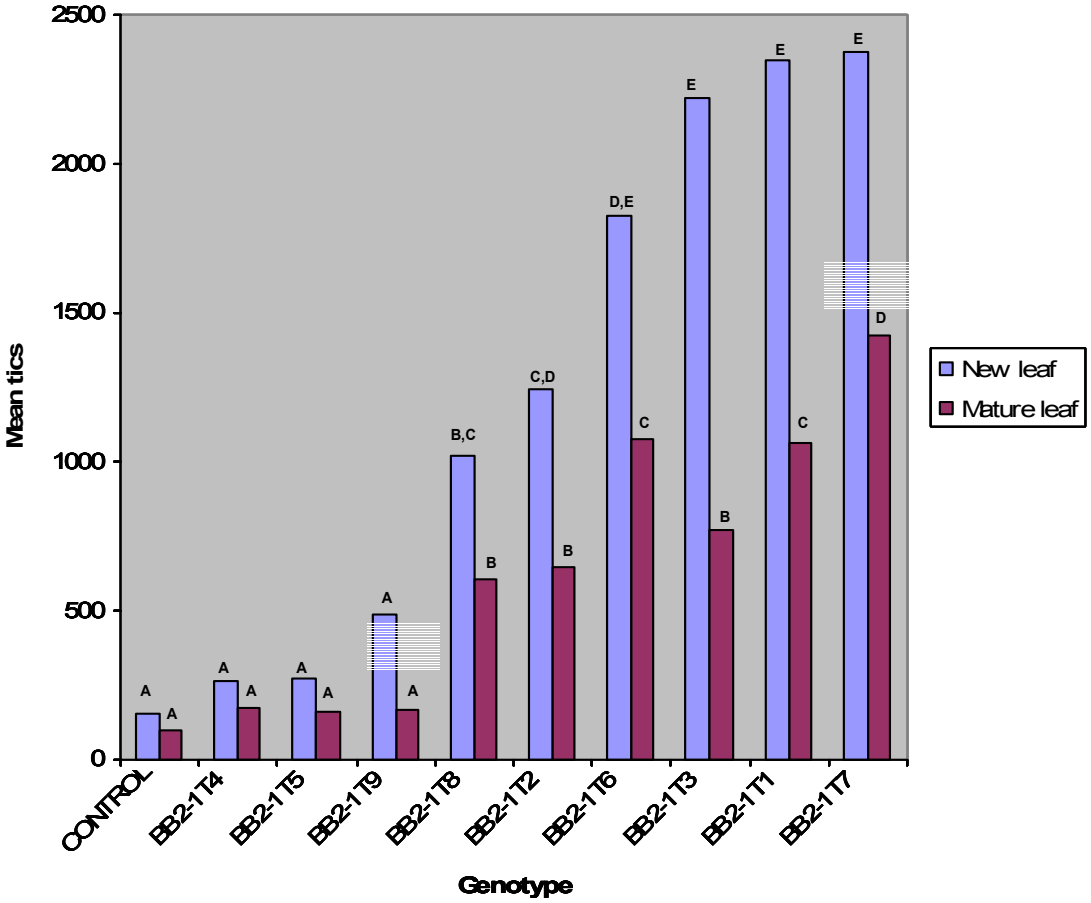

Figure 15. Mean GFP expression (n=3) measured in tics using a GFP meter for nine BB2-1T T_1 progeny lines derived from BB2-T. The plants were significantly different regarding expression in new and mature leaves.

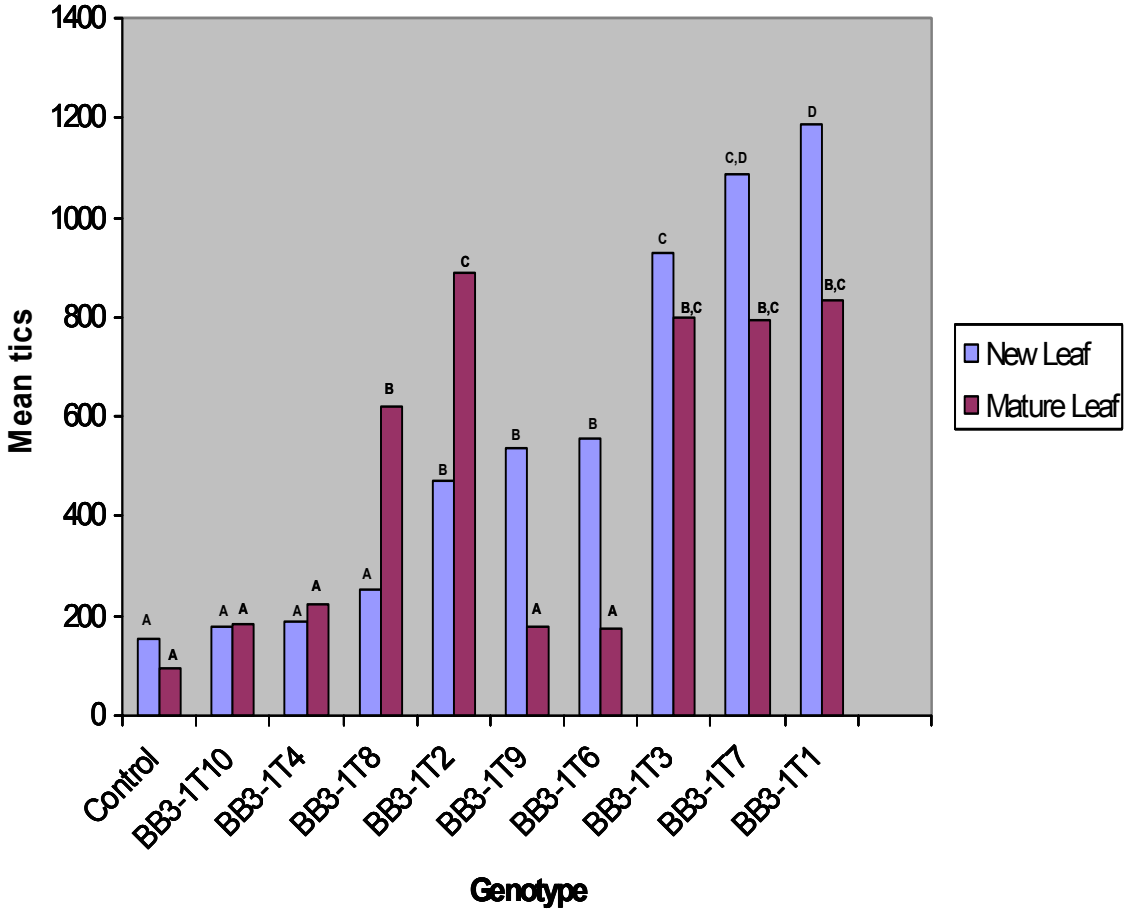

Figure 16. Mean GFP expression (n=3) measured in tics using a GFP meter for nine BB3-1T T_1 progeny lines derived from BB2-T. The plants were significantly different regarding expression in new and mature leaves.

Chapter 3: PCR - mediated optimization of cry2Aa gene for expression in eggplant

Introduction

Proper codon optimization is essential for improving expression of foreign bacterial or viral genes in plants as well as animals. It has been widely used over the last two decades to increase heterologous protein expression in host systems including both plants and animals (Gustafsson et al. 2004). Various crops including rice (Cheng et al. 1998; Fujimoto et al. 1993), barley (Horowitz et al. 2002; Jensen et al. 1996), soybean (Macrae et al. 2005), cotton (Perlak et al. 1990; Wilson et al. 1992), maize (Koziel et al. 1993) and tobacco (Perlak et al. 1991) have been transformed with codon optimized genes to achieve desired levels of gene expression. Codon optimization involves the removal of structures like ATTTA sequences that are recognized as mRNA destabilizing in plants, removal of polyadenylation sequences or cryptic splicing sites from the gene sequence, and increasing the overall GC content. These steps remove the *in vivo* blockade of the translational process(Iannacone et al. 1997) and codons in the optimized gene correspond to the most abundant isoaccepting tRNAs (Fennoy and Baileyserres 1993). The use of optimal codons increases the heterologous gene expression in the plant system. A 22-fold increase in GFP expression was reported following optimization for expression in human beings (Zolotukhin et al. 1996). Similarly, expression of HIV Gag protein, during development of a successful vaccine against HIV, was increased by more than 322-fold following optimization for expression in humans (zur Megede et al. 2000).

Cry genes including *cry*1Ab and *cry*1Ac have been optimized for a 100-fold expression increase in tomato (Perlak et al. 1991). A 15-fold increase in expression of an edible vaccine, i.e., cholera toxin B subunit, was reported after optimization for expression in tobacco plants (Kang et al. 2004). A 1.8 kb *cry*3A gene was optimized and assembled by Adang et al. (1993) for expression in potato.

Eggplant has been previously transformed using a wild type *cry*3B gene (Chen et al. 1995) (for resistance to Colorado potato beetle; CPB) and a rice optimized *cry*1Ab gene (Kumar et al. 1998) (for resistance to BSFB). Transformation of eggplant with an unoptimized (wild type) *cry*3B gene was unsuccessful in developing resistance against CPB and use of a rice optimized *cry*1Ab gene could not confer complete protection to the crop against BSFB. However, transformation with a codon optimized version of *Bt*43 (belonging to *cry*3B class) (Iannacone et al. 1997) and synthetic *cry*3A (Jelenkovic et al. 1998) resulted in successful production of fully CPB resistant eggplant cultivars. Complete resistance to BSFB still has not been achieved. In the present research, we are attempting to confer complete resistance to the borer in eggplant. For this, the gene sequence of a wild type *cry*2Aa was analyzed for codon-usage pattern similarities with eggplant and two related solanaceous crops, i.e., tomato and pepper, to replace the bacterial biased codon sequences with plant biased sequences. For effective incorporation of these changes as well as the removal of destabilizing structures and cryptic introns, the changes were spread throughout the sequence which required *de novo* gene synthesis.

In the present work, a PCR based approach used by Young and Dong (2004) that combined the DA and OE PCR methods for an effective two step gene synthesis of up to 1.2 kb size was used to synthesize the codon optimized version of the wild type *cry*2Aa with the goal of production of fully insect resistant transgenic eggplant.

Materials and methods

The codon usage tables for eggplant, tomato and pepper were obtained from www.kazusa.or.jp/codon/ (Nakamura et al. 2000). For optimization, the following points were considered:

1) The DNA sequence was modified for preferred codon-bias without changing any amino acid sequence and
2) Sequences that generated A+T or G+C strings, i.e., the sequences with A+T or G+C > 6 were removed.
3) The sequence was scanned for putative polyadenylation signals, cryptic splicing sites, self-dimerizing structures, secondary structures and mRNA destabilizing structures such as AATAA, ATTA, AATAAT, AACCAA, ATATAA, AATCAA, ATACTA, ATAAA, ATGAAA, AAGCAT, ATTAAT, ATACAT, AAAATA, ATTAAA, AATTAA, AATACA, GGCGCC, CATAAA, WGGTAA (W=A/T), TRYAG (R = A/G, Y = C/T) (Joshi 1987).
4) Nucleotide changes were made at specific locations in the wild type gene to incorporate unique restriction sites at desirable locations in the optimized version to enable synthetic gene synthesis and subsequent cloning into

appropriate vectors and expression cassettes. The restriction enzymes used and their unique restriction sites are shown in Figure 17.

Synthetic gene design

During codon optimization, the gene sequence was designed to have unique *Xho* I, *Mfe* I, *Bgl* II, *Bsu36* I and *Xma* I sites for ligation of independently synthesized segments (called synthons). Also, these restriction enzymes were chosen such that they were unique in the cloning vector as well. This facilitated proper ligation to the vector and subcloning into the expression cassette. An *Xho* I site was added at the 5' end of the gene before the ATG start codon. Similarly, an *Xma* I was added at the 3' end after the terminating TAG codon. The *Xho* I and *Xma* I sites flanked the *cry* gene and were used to subclone the 1.9 kb *cry*2Aa gene into the pFGC5941 expression cassette. Each of the synthons – A,B,C and D was synthesized individually, cloned into pGEM T- Easy vector, sequenced to identify any mismatches, insertions or deletions and to check the orientation. The clones with no errors were ligated to their flanking synthons by complementary restriction sites.

Design of a synthon

The synthon design and synthesis were based on the protocol used by Young and Dong (2004) in which they combined dual asymmetric (DA) and overlap extension (OE) PCR methods for the synthesis of genes as large as 1.2 kb. In this approach, each synthon was further dissected into 25 bp oligo-nucleotides with a 10 nt overlap at each 5' end and a 15 nt overlap at 3' end

without leaving any nucleotide gaps (Figure 18). Also, synthons were designed such that each DA PCR product had a 35 bp overlap with the adjacent product (Figure 19). The number of the oligonucleotides varied with length of each synthon. A total of 152 oligos was used for synthesis of the gene.

Synthon synthesis

Synthon B synthesis: A total of 49 oligos (25 bp, 10 nM salt-free oligos; Operon biotechnologies) was used for synthesis and 24 DA PCR reactions were carried out to synthesize 60 bp intermediate DA PCR products. The PCR reactions were carried out in 50 µl consisting of 1× ThermoPol reaction buffer (20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1% Triton X-100, pH 8.8 @ 25°C), 200 µM dNTPs (Promega, Madison, WI), and 5 U native Vent DNA polymerase (New England Biolabs, Ipswich, MA). The inner oligos were added for a final concentration of 40 nM flanked by outer oligos at five times that concentration, i.e., 200 nM. The PCR program consisted of following cycles: 1 cycle = 2 min 94°C; 30 cycles = 20 sec 94°C, 30 sec 44°C, 1 min 72°C; 1 cycle = 5 min 72°C. Equal volumes (5 µl) of all 24 DA PCR (B1 – B24) products were pooled for phenol chloroform-isomyl alcohol (25:24:1) extraction and precipitated in three volumes of ethanol. The DNA pellet from the extraction was resuspended in an equal volume of water as the starting volume (120 µl) (Young and Dong 2004).

In the second step, this mixture was subjected to OE PCR: 84 of 120 µl of dissolved DA PCR product was mixed with 1× native *pfu* plus buffer, 200 µM

dNTP mix and 5 U native *pfu* DNA polymerase to a final volume of 100 µl of PCR reaction. The PCR program consisted of the following cycle:

1 cycle = 2 min 94°C; 30 cycles = 30 sec 94°C, 30 sec 55°C, 90 sec 72°C; 1 cycle = 5 min 72°C. The OE PCR product (1 µl) was used for further amplification with two end primers (20 pmol each), 1× native *pfu* plus buffer, 200 µM dNTP mix and 5 U native *pfu* DNA polymerase in a final volume of 25 µl. The following PCR conditions were used:

1 cycle = 2 min 94°C; 30 cycles = 30 sec 94°C, 30 sec 55°C, 90 sec 72°C; 1 cycle = 5 min 72°C. The amplification product was excised from a 1% agarose gel and purified by freeze 'N squeeze spin column (Bio-Rad Laboratories, Hercules, CA). The purified fragment was cloned into pGEM-T Easy vector system (Promega) following dA tailing procedure as recommended in the protocol. Ligation reactions were transformed in high efficiency competent DH 5α cells; plated on LB plates supplemented with ampicillin (100 µg/ml), 100 µl IPTG (100 mM) and 20 µl X-Gal (50 mg/ml); the white colonies were incubated overnight in an incubator shaker at 37ºC and 250 rpm in LB medium supplemented with 100 µg/ml ampicillin. The cultures were screened by PCR and the plasmid was isolated from the positive cultures by Wizard® *Plus* Minipreps DNA Purification System (Promega). The PCR reaction consisted of a total volume of 25 µl with two end primers (20 pmol each) and 100 ng DNA template (overnight culture) in puRe Taq Ready-To-Go PCR beads (Amersham Biosciences, Piscataway, NJ). The following PCR conditions were used:

1 cycle = 3 min 94°C; 35 cycles = 45 sec 94°C, 30 sec 55°C, 90 sec 72°C; 1 cycle = 10 min 72°C. Five plasmid DNA templates (100 ng/µl) were submitted to Virginia Bioinformatics Institute (VBI, Blacksburg, Virginia) for sequencing. T7 and SP6 were used as the forward and reverse primers, respectively.

Synthons A, C and D

Synthon C was constructed from 27 oligos (25 bp) and a total of 13 (C1-C13) DA PCR products was pooled for OE PCR. To construct synthon D, 35 oligos (25 bp) were used to construct 17 DA PCR products (D1 – D17) and these were pooled in an OE PCR reaction. Similar PCR conditions, DNA purification, cloning, ligation and sequencing steps were performed for both C and D.

Synthesis of the 525 bp synthon A was based on 41 oligos (25 bp). Twenty DA PCR reactions were carried out to obtain the first stage products (A1 – A20) and the products were pooled in an attempt to obtain OE PCR product at annealing temperatures between 48 - 55°C.

T7 endonuclease I treatment

The protocol was followed from Young and Dong (2004). The synthon D OE PCR product was subjected to phenol-chloroform isomyl alcohol extraction and the pellet was dissolved in 50 µl T7 1× NEB buffer 2 (10 mM Tris-HCl, 50 mM NaCl, 10 mM $MgCl_2$, 1 mM dithiothreitol, pH 7.9 @ 25°C). This mixture was denatured at 94°C for 3 min and annealed at 75°C for 5 min. After this, 30 U T7 endonuclease I was added and the reaction was incubated at 37°C followed by 55°C for 1 h each.

Sequence analysis

The chromatogram of synthons B, C and D was obtained by Chromas Lite (Technelysium Pty Ltd, Tewantin Qld, Australia) and the sequence was analyzed for any mismatches, orientation, insertions or deletions by CLUSTAL W tool in Biology Workbench (Thompson et al. 1994).

Restriction enzyme digestion and ligations

Ligation of synthons was first done with synthons C and D; followed by ligation of B to the C+D segment and finally ligating synthon A to B+C+D segment.

Restriction enzyme digestion of synthons C and D

Plasmid DNA of synthon C (500 ng) and synthon D (500 ng) were restriction enzyme digested by treatment with 1.5 µl 1× NEB buffer 3 (New England Biolabs) and 0.5 µl (5 U) of *Nco* I (New England Biolabs) in separate reactions in a total volume of 15 µl at 37°C water bath for 4 h. This reaction was stopped by heating at 65°C for 20 min before cooling the reactions to room temperature. This reaction was further subjected to a second digestion by treating with 3 µl 1× NEB buffer 3, 0.3 µl 1× bovine serum albumin (BSA) and 0.5 µl (5 U) *Bsu36* I (New England Biolabs) in a total volume of 30 µl reaction at 37°C water bath for 4 h that was stopped by heating the reaction to 80°C for 20 min.

Ligation of synthons C and D

The reaction products (10 µl) were analyzed on a 1% agarose gel to verify digestion. The plasmid with synthon C was the donor plasmid (insert) and

plasmid D was acceptor plasmid (vector). DNA bands stained with ethidium bromide and visualized under UV light were excised from the gel and purified by Freeze 'N Squeeze spin column (Bio-Rad Laboratories). The purified DNA mixture thus obtained was used for ligation reactions. The ligation mixture consisted of 14 µl mixture of insert + vector (3:1 molar ratio), 15 µl 1× T4 DNA ligase reaction buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM ATP,
10 mM dithiothreitol, 25 µg/ml BSA, pH 7.5 @ 25°C) and 1 µl T4 DNA ligase (New England Biolabs). This reaction was incubated overnight at 4°C to achieve maximum ligation efficiency. The ligation reactions were transformed into highly competent DH 5α cells by heat shock treatment at 42°C. Colonies were grown overnight at 37°C in LB medium supplemented with 100 µg/ml ampicillin and positive colonies were selected by PCR screening. The PCR reaction consisted of a total volume of 25 µl with two end primers (20 pmol each) and 100 ng DNA template (overnight culture) in puRe Taq Ready-To-Go PCR beads (Amersham Biosciences). The following PCR conditions were used:
1 cycle = 3 min 94°C; 35 cycles = 45 sec 94°C, 30 sec 55°C, 90 sec 72°C; 1 cycle = 10 min 72°C.

Plasmids isolated from positive colonies were purified by Wizard® *Plus* Miniprep DNA Purification System (Promega) and sequenced to verify fidelity of ligations. Reactions with successful ligation and without any mutations were further ligated with synthon B where synthon B was used as an insert.

Restriction enzyme digestion of synthon B and plasmid C+D

Five µl DNA (500 ng) of synthon B and 5 µl DNA (500 ng) of C+D were treated with 1.5 µl 1× NEB buffer 3 (New England Biolabs) and 0.5 µl (5 U) *Nco* I (New England Biolabs) in separate reactions as described above. These reactions were further subjected to digestion by treating with 3 µl 1× NEB buffer 3 and 0.5 µl (5 U) *Bgl* II (New England Biolabs) each in a total volume of 30 µl reaction at 37°C water bath for 4 h. Since, *Bgl* II cannot be heat inactivated, the reaction was subjected to gel electrophoresis at the end.

Ligation of synthon B and plasmid C+D

Synthon B was used as the insert and plasmid C+D was used as the acceptor vector for ligations. The same protocol was followed for ligation and selection of positive colonies as described for ligation of synthons C and D. Plasmids isolated from the positive colonies were sent for sequencing and eventually used for ligating synthon A to plasmid B+C+D.

Restriction enzyme digestion of synthon A and plasmid B+C+D

Synthon A in plasmid pUC57 and plasmid B+C+D was digested with enzymes *Aat* II and *Mfe* I using NEB buffer 4 as described above. The reaction was stopped by heating at 65°C for 20 min and subsequent cooling to room temperature.

Ligation of synthon A and plasmid B+C+D

Ten µl of the products from ligation reactions were analyzed and purified from a 1% agarose gel and subjected to a ligation reaction consisting of 14 µl

purified DNA mixture (containing A as insert and B+C+D as acceptor plasmid), 15 µl 1× T4 DNA ligase reaction buffer and 1 µl T4 DNA ligase. The reaction was incubated overnight at 4°C; positive colonies were selected, screened, purified and sequenced following the same methods as described previously.

Gene cloning

Cloning in pFGC5941

The artificially synthesized *cry*2Aa gene was cloned in the expression cassette pFGC5941 (Figure 20) after verification of sequence. The gene was cloned between *Xho* I and *Xma* I sites in pFGC5941. This was accomplished following steps:

The gene in pGEM-T Easy vector was restriction enzyme digested using *Xho* I and *Xma* I thus producing compatible 5' and 3' ends. 500 ng of the gene was treated with 1.5 µl NEB buffer 4 and 0.5 µl (5 U) of *Xho* I in the first step for 3-4 h at 37°C in water bath. The reaction was stopped by increasing the temperature to 65°C for 20 min. The reaction temperature was cooled to room temperature. Then, reaction mixture was treated with 3 µl NEB buffer 4 and 0.5 µl (5 U) of *Xma* I for another 3-4 h at 37°C in a water bath followed by termination of reaction at 65°C for 20 min; pFGC5941 was also digested simultaneously following the same procedure.

Ten µl of each of the reactions were analyzed on a 1% agarose gel, the bands were excised, purified as described above and purified DNA was used for ligation. The ligation mixture consisted of 14 µl purified DNA, 15 µl 1× T4 DNA

ligase reaction buffer and 1 µl T4 DNA ligase. The reaction was incubated overnight at 4°C; the ligation reactions were transformed into highly competent DH 5α cells, colonies were grown on 30 ml LB plates supplemented with 50 mg/l kanamycin, positive colonies were selected, grown overnight in an incubator shaker at 37°C and 250 rpm. The cultures were screened using PCR for the presence of an insert in the vector. Plasmid DNA was isolated from positive cultures following the protocol mentioned above and used for the subcloning of *cry*2Aa under the CaMV 35S promoter by transcriptional fusion in expression vector pCAMBIA 1302 (CAMBIA, Black Mountain, Australia).

Results

Codon optimization

Proper codon optimization was one of the major goals of the present research. The native *Bacillus thuringiensis cry*2Aa sequence was modified from AT rich to a more GC rich plant optimized *cry*2Aa gene to attempt to produce fully BSFB resistant eggplant cultivars. A total of 472 nucleotides was replaced in the native 1.9 kb *cry*2Aa gene. The GC content was increased from 34.3% in the native gene to 41.3% in the optimized gene. This GC content falls in between the AT-biased *Bt* genome and GC-biased monocot genome (Adang et al. 1993).

Also, the use of rare codons like CG and TA dinucleotides was altered with frequently used codons at codon positions 2 and 3 (Murray et al. 1989). It has been well proven that the greater the number of rarely used codons in expression host system, especially near to the transcription start site, the less likely is the satisfactory expression of the heterologous protein (Kane 1995).

Gene synthesis

Synthetic genes are becoming a major tool in biotechnology because they allow the scientist to manipulate codon usage pattern for a particular host, remove repetitive sequences and place restriction sites as desired for subsequent manipulations. The low cost and rapid turn around of short oligo primers make the DA PCR/OE PCR gene synthesis method attractive. The number of 25 bp overlapping primers needed to construct over 1.9 kb codon-modified *cry*2Aa gene was 41, 49, 27 and 35 for synthons A, B, C and D, respectively. The T_m of the primers ranged from 68 to 79°C and the GC content from 30 to 48%. T_m and GC content of primers for the four synthons differ significantly. It is obvious from Figures 23 and 24 that some DA PCR products resulted in stronger bands than others.

DA PCR products

The DA PCR products (60 bp) for synthons – A (Figure 21), B (Figure 22), C (Figure 23), and D (Figure 24) were analyzed on a 4% agarose gel and visualized under UV light after staining with ethidium bromide. Most of the DA PCR products were obtained in a single reaction. However, some DA PCR products were only obtained after repeated PCR attempts (3 for synthons B and C; 1 for synthon D) without changing any PCR conditions.

OE PCR products

Four OE PCR products were analyzed on a 1% agarose gel and visualized by ethidium bromide staining to verify the product size of each synthon (Figure 25). The products were of correct size, i.e., synthon A – 525 bp (Figure 25; Lane 2), synthon B – 615 bp (Figure 25; Lane 3), synthon C – 348 bp (Figure

25; Lane 5), and synthon D – 432 bp (Figure 25; Lane 6). The initial OE PCR product for synthon A was lost. However, on repeating the reaction, no OE PCR product could be obtained after pooling DA PCR products of synthon A. A range of annealing temperatures between 48 - 55°C was tried but without any success. After eight subsequent attempts, this part of the gene was purchased from Celtek genes (Celtek Bioscience, LLC, Nashville, TN) in pUC57 vector cloned between *Xho* I and *Hind* III sites. We also attempted to break synthon A into 200 bp and 325 bp pieces and use OE PCR on the first eight and last 12 subsets of the DA PCR products but we were unsuccessful in synthesizing either of the smaller segments. The company synthesized synthon A was analyzed on a 1% agarose gel and found to be the correct size, i.e., 525 bp (Figure 25; Lane 2).

A smear was seen after OE PCR of the pooled DA PCR products in synthon D. This smear indicated the presence of mismatches, incomplete products and dimers in the product (Young and Dong 2004). These mismatches were removed after T7 endonuclease I treatment of the OE PCR product (Figure 25; Lane 8).

Sequencing results

The synthons were sequenced after each OE PCR to verify the fidelity and orientation of cloning. The sequences without any mismatches, deletions and insertions were selected for further restriction enzyme digestion and ligation reactions. No gaps were left between the oligos and high fidelity proofreading *pfu* polymerase was used for PCR to ensure the highest possible efficiency of the PCR products. An overlap of 35 bp with the adjacent DA PCR products ensured

no non-specific annealing even though all the products were pooled in a single PCR reaction. Use of *pfu* polymerase compromised our goal of low cost gene synthesis, thus another high fidelity Vent DNA polymerase was used. Synthon B was synthesized using Vent DNA polymerase and synthons C and D were synthesized by *pfu* polymerase. The efficiency of perfect sequence ranged from 20% (synthon B) and 100% (synthon D) (Table 7). Vent DNA polymerase is known to have a higher error rate (2.8×10^{-6}) than *pfu* polymerase (1.3×10^{-6}) and this was clear from the higher error frequency in synthon B. Also, the length of synthon B may be another factor causing the higher error frequency. The most common errors were deletions (eight of 16 sequence variants), with insertions, duplications and mismatches occurring less frequently (Table 7). Some sequences had multiple errors.

Ligation reaction results

The four synthons with perfect sequences were restriction enzyme digested from their cloning vectors and ligated using the complementary restriction sites to yield 1920 bp *cry* gene product (Figure 26).

After synthesis of the codon optimized *cry* gene, we ligated it into expression vector pFGC5941 next to the CaMV35S promoter with bar resistance gene. This vector was used to attach a constitutive promoter in front of the gene before subcloning it in the pCAMBIA 1302 expression cassette.

Binary vector pCAMBIA1302 has been widely used in plant transformation experiments because of its high throughput nature, small size, ability to accept

large gene inserts and the GFP reporter gene. In the present research, the synthetic gene with CaMV35S promoter was cloned into multiple cloning site in front of GFP.

Discussion

Our optimized version of a native *cry*2Aa gene that was isolated from an Indian isolate of *Bt* is clearly different from other *cry*2Aa sequences as is clear from the phylogenetic tree analysis comparing the native *cry*2Aa gene with plant codon-optimized *cry*2Aa and nine other related *cry*2Aa genes (Figure 27). This optimization was achieved without any changes in the amino acid sequence coded by the gene.

Exchange of plasmid between *Bt* strains leads to genetic diversity which can further increase toxic potential of the strain. Thus, identification of indigenous *Bt* strains can lead to isolation of novel *cry* genes toxic to specific insect groups. Such diversity in *Bt* has already been found in Turkey (Apaydin et al. 2005), the Philippines (Theunis et al. 1998), Japan (Lee et al. 2002) and Spain (Iriarte et al. 2000). The Indian subcontinent is rich in biodiversity and genetic resources. The identification of indigenous isolates of *Bt* and isolation of *cry* genes from them is a valuable tool for identifying genes toxic to the pests prevalent in that area (Kumar and Udayasuriyan 2004). The *cry*2Aa gene used in this research is known to be toxic against Lepidopteran insect larvae (Lenin et al. 2001). Eggplant has been previously transformed with *cry*1Ab; although protein expression was observed to be high, fully resistant cultivars could not be developed. Protein expression has been used as an indicator of relative efficacy

of a gene but cannot predict the absolute expression (Adang et al. 1993). Also, the Cry1 class of proteins is significantly larger (approx. 137 KDa) than Cry2Aa protoxins (72 KDa). Gene size is known to affect the expression pattern; the larger the gene, the more limited is its expression thus reducing potential lethal affect (Morse et al. 2001). Also, *cry* genes express as protoxins (Kota et al. 1999) instead of toxins which are known to be more genetically stable but need highly specific conditions for activation. This makes them specific to a target insect with minimum damage to nontarget species.

The goal of redesigning a native version of the *cry* gene was to make this gene specifically toxic to a major pest belonging to the lepidopteran group, BSFB which causes extensive losses to eggplant production every year. Codon optimization was achieved by making specific sequence modifications to resemble the codon bias of eggplant, tomato and pepper. Since these modifications were spread throughout the gene sequence, *de novo* gene synthesis was considered as a better approach than site directed mutagenesis where codon bias is compromised. Several codon optimizing tools are available and various companies have developed different proprietary softwares to adjust the codon bias in accordance with a host. Some of these tools and companies are listed below:

a) Upgene: A web based codon optimization algorithm (Gao et al. 2004).

b) Codon optimizer (Fuglsang 2003)

c) Software Leto 1.0 (Entelechon GmbH, Regensburg, Germany)

d) Software Genemaker (Blue Heron Bio, BOTHELL, WA, USA)

e) GCUA: General Codon Usage Analysis (McInerney 1998)

f) DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis (Hoover and Lubkowski 2002)

g) Synthetic Gene Designer (Wu et al. 2005)

However, none of the above software could be used for optimization of a gene specifically for eggplant. This is because the codon usage table of eggplant is based on only 14 coding sequences (CDS's) and use of any software on such limited data is not expected to produce reliable results. Thus, we used, in addition to eggplant, related solanaceous crops, i.e., tomato with 1249 CDS's (543566 codons) and pepper with 225 CDS's (69676 codons) in their database for optimal optimization (Nakamura et al. 2000).

Artificial gene synthesis of individual segments (synthons) followed by ligation of synthons to produce the 1.9 kb *cry*2Aa gene was achieved for three of four synthons. In synthon A, the exception, no PCR product was obtained after pooling of products from DA PCR. Some of the possible reasons for this follow:

1) Yield of some of the DA PCR products was not high enough compared to other products. Insufficiency of some of these products can be explained as a cause of failure to yield an OE PCR product.
2) The T_m of successive DA PCR products should be similar for successful OE PCR. Wide variation between T_m of some of the products can be another cause of inability of products to amplify.
3) Some of the oligos had high GC content (For example: TATGTGGCCC CAGTGGTTGGTACTG) and AT stretches (A8 - CCTTTTGAAGAAGGT).

The use of 25 bp primers with no gap between oligos promised decreased errors but it could not totally eliminate them. Certain mutations like deletions, insertions, mismatches and duplications were also observed in the sequencing results. We adopted the same protocol used by Young and Dong (2004) but more mutations were observed in our gene synthesis. This can be attributed to the following:

1) Only synthon D was subjected to T7 endonuclease treatment which eliminated all the mutated and heteroduplex products and thus no error was observed in sequencing results (100% efficiency). Synthon D was selected for T7 endonuclease treatment because the OE PCR product occured as a smear on the gel which indicated the presence of heteroduplexes and dimers in the product. Treatment with T7 endonuclease introduced cuts at sites of mismatches and thus heteroduplexes were cleaved into small pieces with only homoduplexes left in the correct size range. Other synthons, i.e., B and C showed a clear single band and sequencing results showed mutations in the product with efficiency in the range of 20% - 33%.
2) Our goal was to achieve gene synthesis at minimal cost. Due to the high cost of *pfu* DNA polymerase, we used Vent DNA polymerase for the synthesis of synthon B which is significantly cheaper and is a high fidelity proofreading enzyme but with an error rate higher than *pfu* polymerase. Also, synthon B was the longest synthon. A combination of these factors led to lowest efficiency of gene sequence fidelity for synthon B.

3) More PCR cycles, i.e., 30 cycles were used for DA PCR compared to 15-20 cycles in the Young and Dong (2004) protocol for DA PCR and 30 cycles for amplification of synthons by OE PCR. More PCR cycles can lead to increased error rates during polymerase activity.

This optimized version of gene was 20% different in its nucleic acid sequences and encoded 633 amino acids which were identical to the amino acid sequence encoded by native *cry* gene.

Future prospects

The eggplant-specific codon modified *cry*2Aa gene has been cloned under the transcriptional control of the 35S promoter into the pCAMBIA 1302 binary expression vector. The clones will be transformed into *Agrobacterium* and the presence of recombinant plasmid in *Agrobacterium* will be checked by PCR analysis. The positive clones will be used for eggplant transformation. The putative transgenic plants will be tested by PCR analysis and southern blot analysis. Expression of the *cry*2Aa gene will be confirmed by western blot analysis. The transgenic plants will be used for insect feeding bioassays using the wt and codon optimized *cry*2Aa gene to determine the difference in entomocidal activity of the Cry2Aa protein in transgenic plants using two different versions of the *cry* gene. The pesticidal activity of the protein will be tested in further generations. The transgenic eggplant lines that confer complete protection against BSFB can be released as a commercial cultivar in southeast Asia.

References

Adang MJ, Brody MS, Cardineau G, Eagan N, Roush RT, Shewmaker CK, Jones A, Oakes JV, McBride KE (1993) The reconstruction and expression of a *Bacillus thuringiensis cry*3A gene in protoplasts and potato plants. Plant Mol Biol 21: 1131-1145.

Apaydin O, Yenidunya AF, Harsa S, Gunes H (2005) Isolation and characterization of *Bacillus thuringiensis* strains from different grain habitats in Turkey. World J Microbiol Biotechnol 21: 285-292.

Beitz E (2000) TeXshade:shading and labeling of multiple sequence alignments using LaTeX2e. . Bioinformatics 16.

Chen Q, Jelenkovic G, Chin CK, Billings S, Eberhardt J, Goffreda JC, Day P (1995) Transfer and transcriptional expression of coleopteran *cry*3B endotoxin gene of *Bacillus thuringiensis* in eggplant. J Am Soc Hort Sci 120: 921-927.

Cheng XY, Sardana R, Kaplan H, Altosaar I (1998) *Agrobacterium*-transformed rice plants expressing synthetic *cry*IA(b) and *cry*IA(c) genes are highly toxic to striped stem borer and yellow stem borer. Proc Nat Acad Sci USA 95: 2767-2772.

Felsenstein J (1989) PHYLIP -- Phylogeny Inference Package (Version 3.2). Cladistics 5: 164-166.

Fennoy SL, Baileyserres J (1993) Synonymous Codon Usage in *Zea mays* L nuclear genes is varied by levels of C-ending and G-ending codons. Nucleic Acids Res 21: 5294-5300.

Fuglsang A (2003) Codon optimizer: a freeware tool for codon optimization. Protein Expr Purif 31: 247-249.

Fujimoto H, Itoh K, Yamamoto M, Kyozuka J, Shimamoto K (1993) Insect resistant rice generated by introduction of a modified delta-endotoxin gene of *Bacillus thuringiensis*. Biotechnology (NY) 11: 1151-5.

Gao W, Rzewski A, Sun H, Robbins PD, Gambotto A (2004) UpGene: Application of a web-based DNA codon optimization algorithm. Biotechnol Prog 20: 443-8.

Gustafsson C, Govindarajan S, Minshull J (2004) Codon bias and heterologous protein expression. Trends Biotechnol 22: 346-353.

Hoover DM, Lubkowski J (2002) DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. Nucleic Acids Res 30: e43.

Horowitz S, Freeman S, Sharon A (2002) Use of green fluorescent protein-transgenic strains to study pathogenic and nonpathogenic lifestyles in *Colletotrichum acutatum*. Phytopathology 92: 743-749.

Iannacone R, Grieco PD, Cellini F (1997) Specific sequence modifications of a *cry*3B endotoxin gene result in high levels of expression and insect resistance. Plant Mol Biol 34: 485-496.

Iriarte J, Porcar M, Lecadet MM, Caballero P (2000) Isolation and characterization of *Bacillus thuringiensis* strains from aquatic environments in Spain. Curr Microbiol 40: 402-408.

Jelenkovic G, Billings S, Chen Q, Lashomb J, Hamilton G, Ghidiu G (1998) Transformation of eggplant with synthetic *cry*3A gene produces a high level of resistance to the Colorado potato beetle. J Am Soc Hort Sci 123: 19-25.

Jensen LG, Olsen O, Kops O, Wolf N, Thomsen KK, von Wettstein D (1996) Transgenic barley expressing a protein-engineered, thermostable (1,3-1,4)-beta-glucanase during germination. Proc Natl Acad Sci USA 93: 3487-91.

Joshi CP (1987) Putative polyadenylation signals in nuclear genes of higher plants - a compilation and analysis. Nucleic Acids Res 15: 9627-9640.

Kane JF (1995) Effects of rare codon clusters on high level expression of heterologous proteins in *Escherichia coli*. Curr Opin Biotechnol 6: 494-500.

Kang TJ, Loc NH, Jang MO, Yang MS (2004) Modification of the cholera toxin B subunit coding sequence to enhance expression in plants. Mol Breed 13: 143-153.

Kota M, Daniell H, Varma S, Garczynski SF, Gould F, Moar WJ (1999) Overexpression of the *Bacillus thuringiensis* (Bt) Cry2Aa2 protein in chloroplasts confers resistance to plants against susceptible and Bt-resistant insects. Proc Natl Acad Sci USA 96: 1840-5.

Koziel MG, Beland GL, Bowman C, Carozzi NB, Crenshaw R, Crossland L, Dawson J, Desai N, Hill M, Kadwell S, Launis K, Lewis K, Maddox D, McPherson K, Meghji MR, Merlin E, Rhodes R, Warren GW, Wright M, Evola SV (1993) Field performance of elite transgenic maize plants expressing an insecticidal protein derived from *Bacillus thuringiensis*. Bio/Technology 11: 194-200.

Kumar PA, Mandaokar A, Sreenivasu K, Chakrabarti SK, Bisaria S, Sharma SR, Kaur S, Sharma RP (1998) Insect-resistant transgenic brinjal plants. Mol Breed 4: 33-37.

Kumar S, Udayasuriyan V (2004) Cloning of *cry*2Aa and *cry*2Ab genes from new isolates of *Bacillus thuringiensis* and their expression in recombinant *Bacillus thuringiensis* and *Escherichia coli* strains. World J Microbiol Biotechnol 20: 11-17.

Lee DH, Machii J, Ohba M (2002) High frequency of *Bacillus thuringiensis* in feces of herbivorous animals maintained in a zoological garden in Japan. Jpn J Appl Entomol Zool 37: 509-516.

Lenin K, Mariam MA, Udayasuriyan V (2001) Expression of a *cry*2Aa gene in an acrystalliferous *Bacillus thuringiensis* strain and toxicity of Cry2Aa against *Helicoverpa armigera*. World J Microbiol Biotechnol 17: 273-278.

Macrae TC, Baur ME, Boethel DJ, Fitzpatrick BJ, Gao AG, Gamundi JC, Harrison LA, Kabuye VT, McPherson RM, Miklos JA, Paradise MS, Toedebusch AS, Viegas A (2005) Laboratory and field evaluations of transgenic soybean exhibiting high-dose expression of a synthetic *Bacillus thuringiensis cry*1A gene for control of Lepidoptera. J Econ Entomol 98: 577-587.

McInerney JO (1998) GCUA: general codon usage analysis. Bioinformatics 14: 372-3.

Morse RJ, Yamamoto T, Stroud RM (2001) Structure of Cry2Aa suggests an unexpected receptor binding epitope. Structure 9: 409-417.

Murray EE, Lotzer J, Eberle M (1989) Codon usage in plant genes. Nucleic Acids Res 17: 477-98.

Nakamura Y, Gojobori T, Ikemura T (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucleic Acids Res 28: 292.

Perlak FJ, Deaton RW, Armstrong TA, Fuchs RL, Sims SR, Greenplate JT, Fischhoff DA (1990) Insect resistant cotton plants. Biotechnology (NY) 8: 939-943.

Perlak FJ, Fuchs RL, Dean DA, McPherson SL, Fischhoff DA (1991) Modification of the coding sequence enhances plant expression of insect control protein genes. Proc Natl Acad Sci USA 88: 3324-3328.

Theunis W, Aguda RM, Cruz WT, Decock C, Peferoen M, Lambert B, Bottrell DG, Gould FL, Litsinger JA, Cohen MB (1998) *Bacillus thuringiensis* isolates from the Philippines: habitat distribution, delta-endotoxin diversity, and toxicity to rice stem borers (Lepidoptera : Pyralidae). Bull Entomol Res 88: 335-342.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-80.

Wilson FD, Flint HM, Deaton WR, Fischhoff DA, Perlak FJ, Armstrong TA, Fuchs RL, Berberich SA, Parks NJ, Stapp BR (1992) Resistance of cotton lines containing a *Bacillus thuringiensis* toxin to pink bollworm (Lepidoptera, Gelechiidae) and other insects. J Econ Entomol 85: 1516-1521.

Wu G, Nabila B-B, Stephen F (2005) The synthetic gene designer: a flexible web platform to explore sequence space of synthetic genes for heterologous expression.

Young L, Dong QH (2004) Two-step total gene synthesis method. Nucleic Acids Res 32.

Zolotukhin S, Potter M, Hauswirth WW, Guy J, Muzyczka N (1996) A "humanized" green fluorescent protein cDNA adapted for high-level expression in mammalian cells. J Virol 70: 4646-54.

zur Megede J, Chen MC, Doe B, Schaefer M, Greer CE, Selby M, Otten GR, Barnett SW (2000) Increased expression and immunogenicity of sequence-modified human immunodeficiency virus type 1 gag gene. J Virol 74: 2628-35.

Table 7: Data showing the total number of clones of each synthon that were sequenced and error analysis including deletions, insertions, duplications and mismatches

Synthon	Total clones sequenced	Correct	Deletions	Insertions	Duplications	Mismatches
A	--	--	--	--	--	--
B	5	1	7	3	2	2
C	3	1	1	0	1	0
D	3	3	0	0	0	0

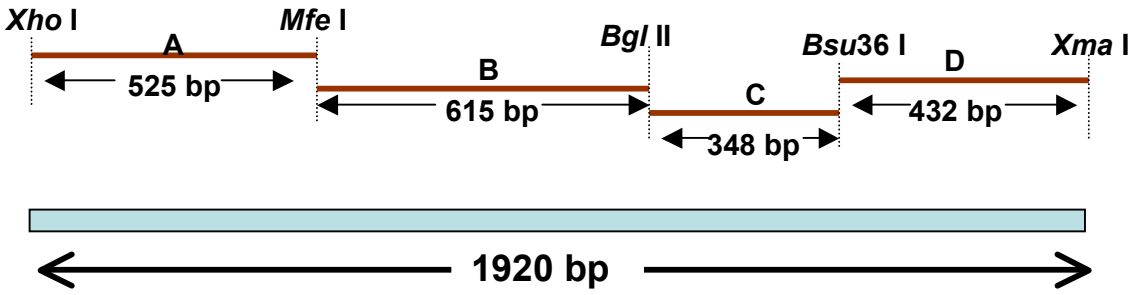

Figure 17. The 1920 bp *cry*2Aa gene partitioned into four synthons – A, B, C and D with the indicated restriction sites that would be used to anneal the synthons into a functional codon optimized gene

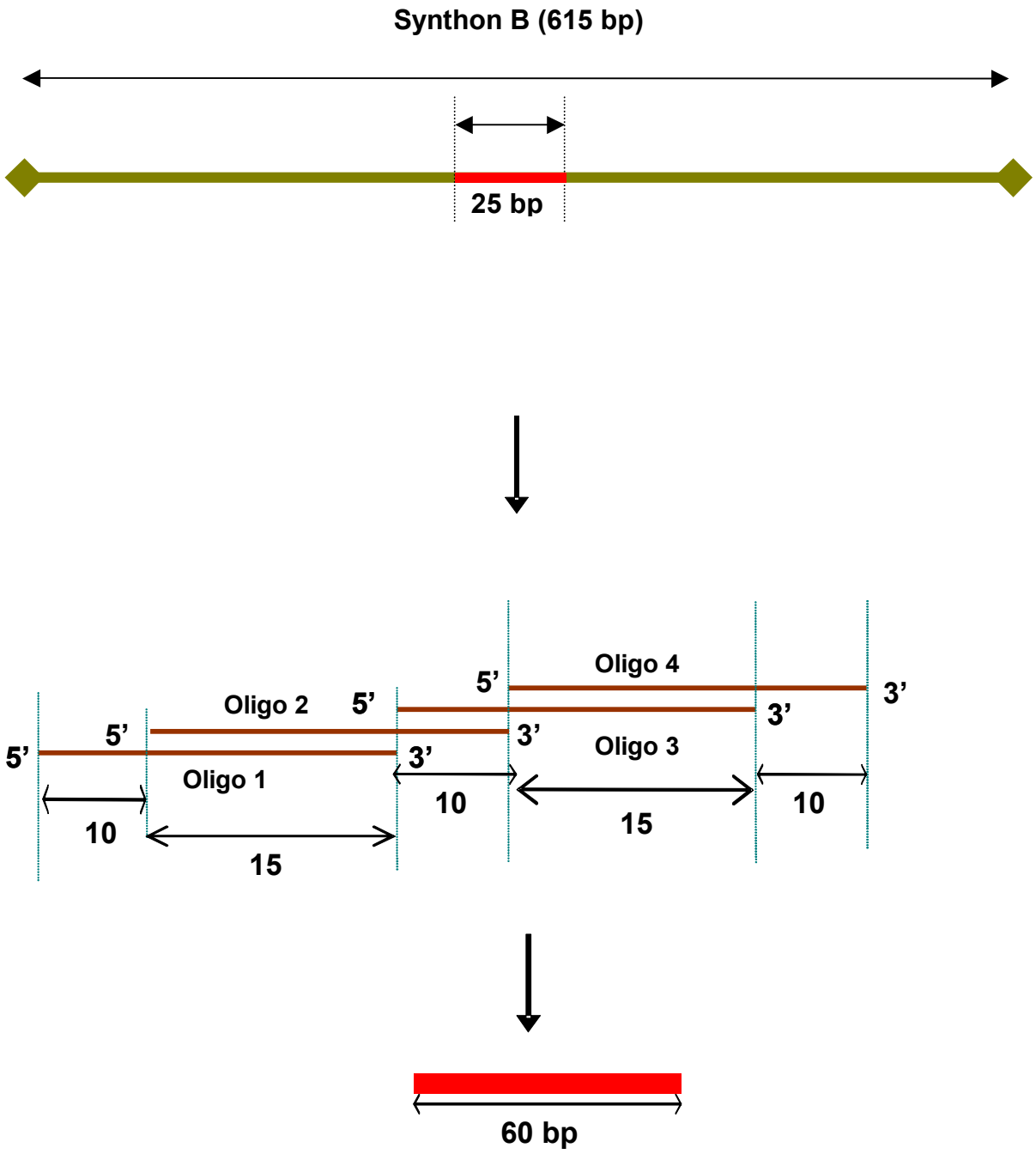

Figure 18. Synthon B was made from 49 25 bp oligos used in sets of four for DA PCR. The four oligos in one reaction had a 10 bp overlap at 5' end and 15 bp overlap at 3' end resulting in a 60 bp DA PCR product.

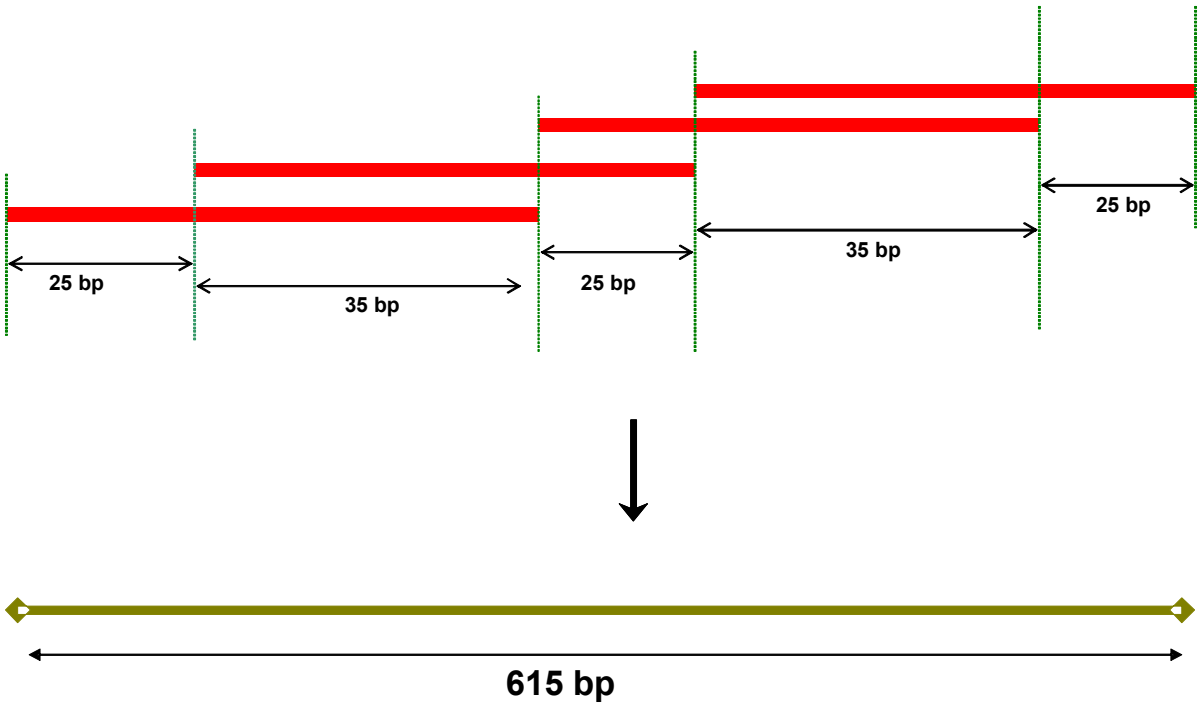

Figure 19. After DA PCR, synthon B could be constructed from 24 60 bp DA PCR products (4 of 24 shown in figure) with 35 bp overlaps used in OE PCR.

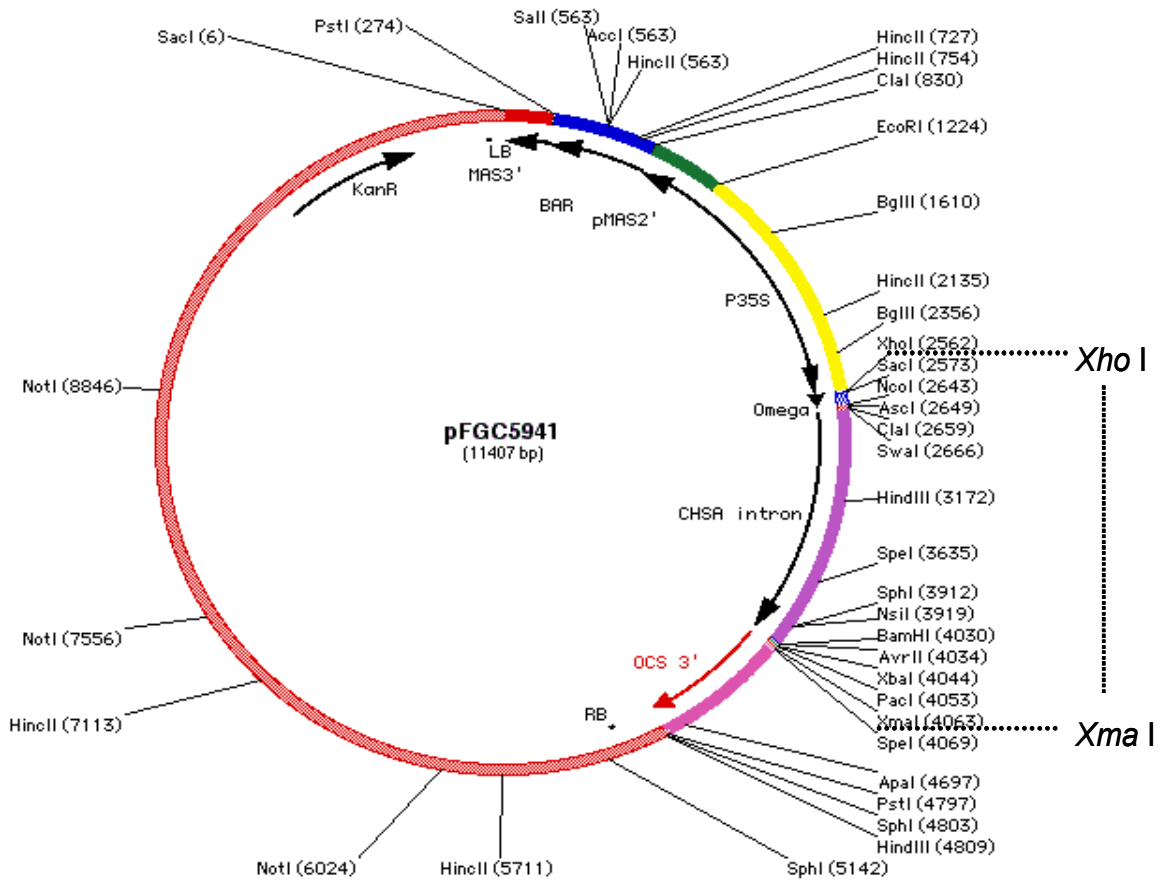

Figure 20. The codon optimized *cry*2Aa gene was cloned between *Xho* I and *Xma* I sites downstream of CaMV 35S promoter. This vector had kanamycin resistance for bacterial selection and the *bar* gene for plant selection with Basta resistance.

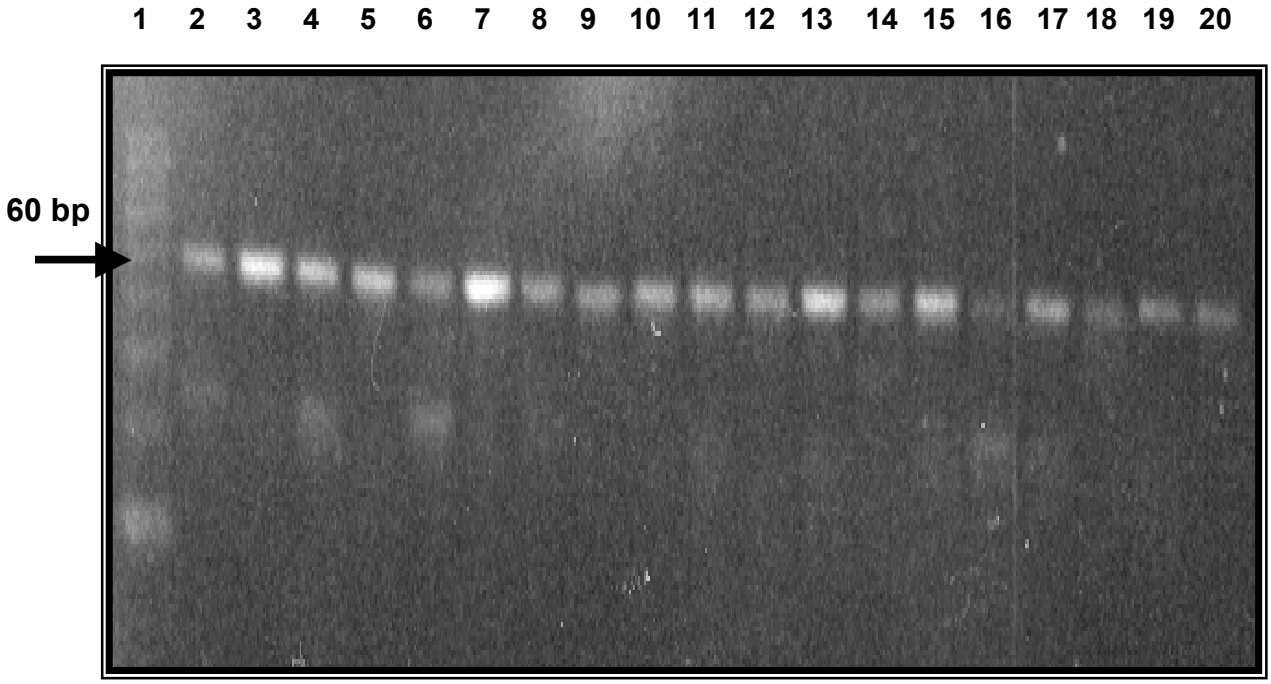

Figure 21. 60 bp DA PCR products of synthon A, resolved on a 4% agarose gel. Lane 1: 10 bp DNA step ladder; lanes 2-20; DA PCR products A1-A19. Synthon A20 was synthesized in a separate PCR reaction (not shown).

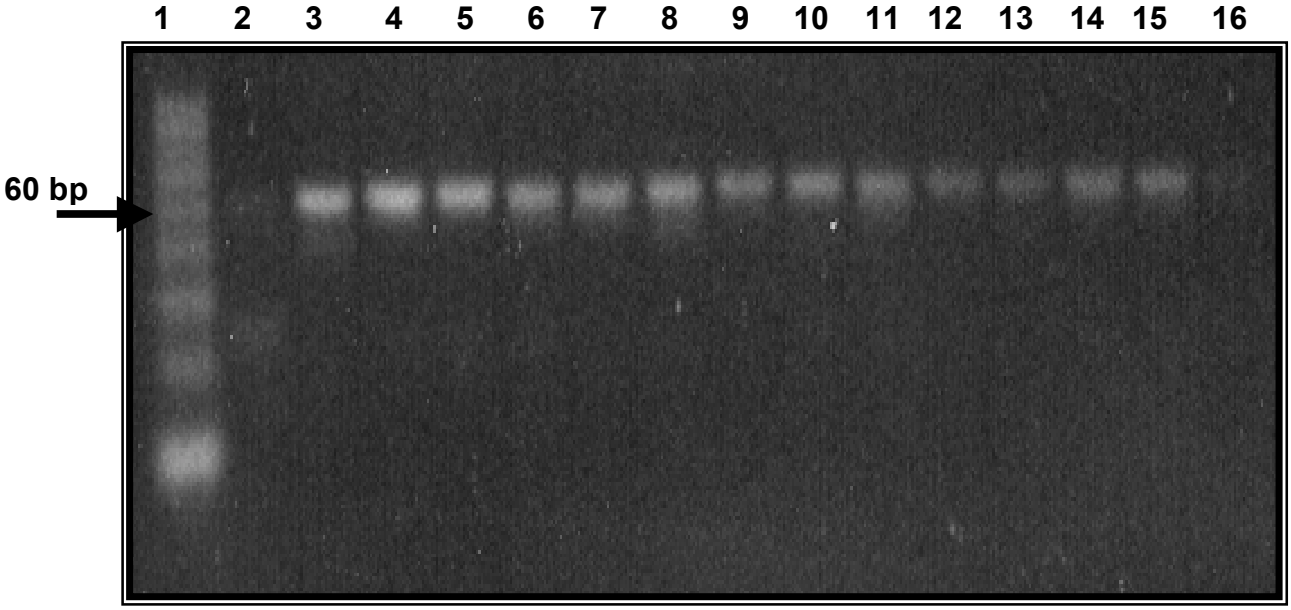

Figure 22. 4% agarose gel electrophoresis of DA PCR products of synthon B. Lane 1: 10 bp DNA step ladder; lanes 2-16: 60 bp DA PCR products B10-B24. Products B1-B9 could not be synthesized in a single PCR and thus were synthesized in independent PCRs (not shown). Similarly, the products in low amounts (B10 and B24) were also synthesized in separate reactions to ensure equal amount of products to be pooled in the OE PCR.

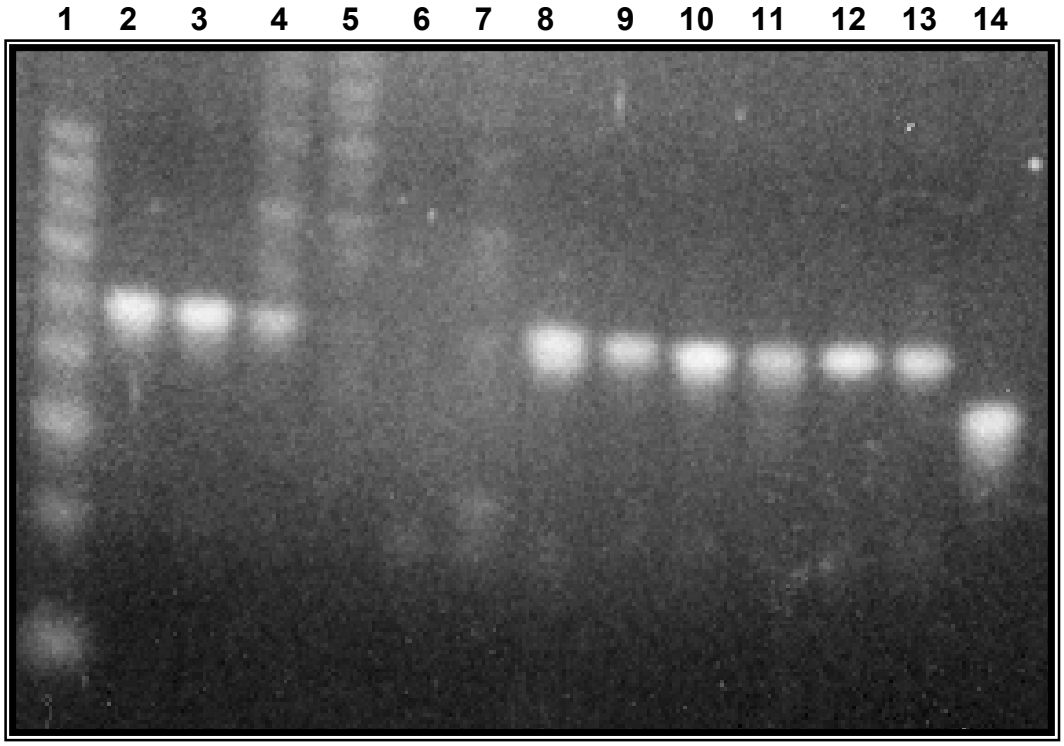

Figure 23. 4% agarose gel electrophoresis of DA PCR products of synthon C. Lane 1: 10 bp DNA step ladder; lanes 2-14: 60 bp DA PCR products C1-C13. C13 was a 50 bp product. Products C3-C6 could not be synthesized in single PCR and thus were synthesized in independent PCRs (not shown).

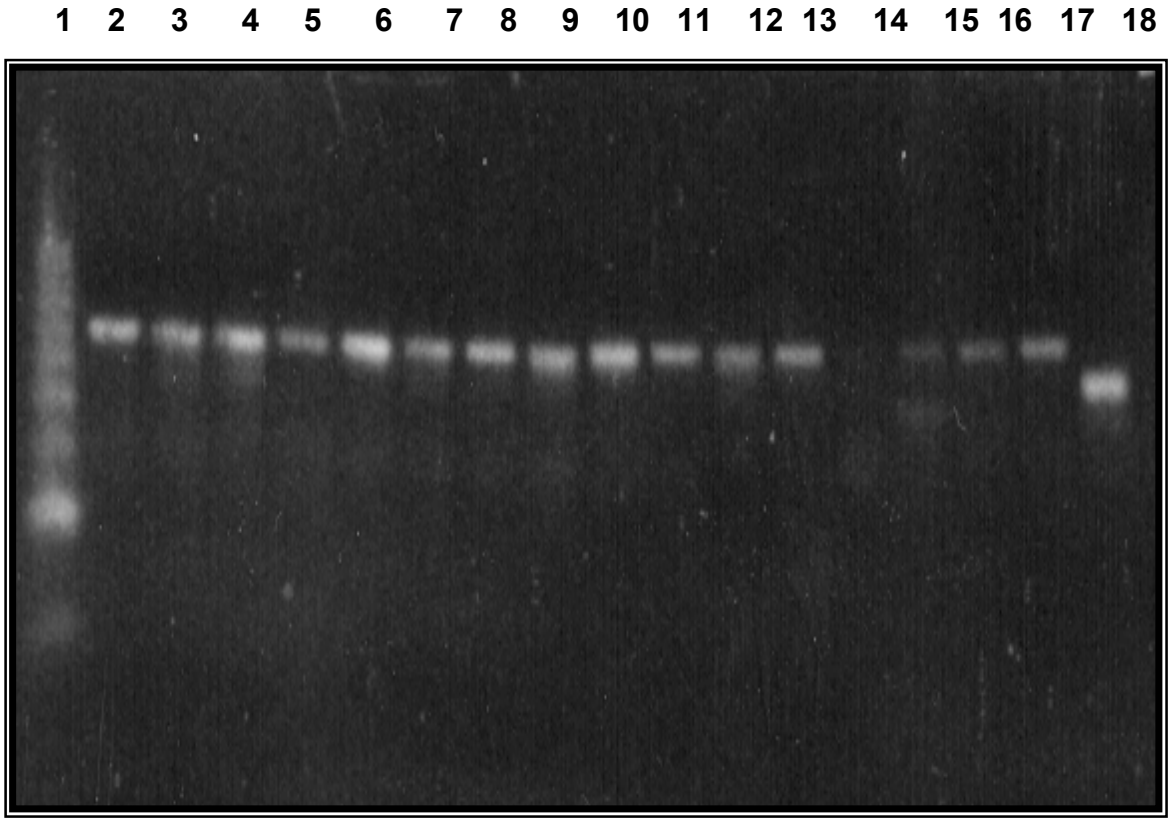

Figure 24. 4% agarose gel electrophoresis of DA PCR products of synthon D. Lane1: 10 bp DNA step ladder; lanes 2-18: 60 bp DA PCR products D1-D17. D14 was synthesized in an independent PCR reaction (not shown). 17 DA PCR products were pooled for OE PCR.

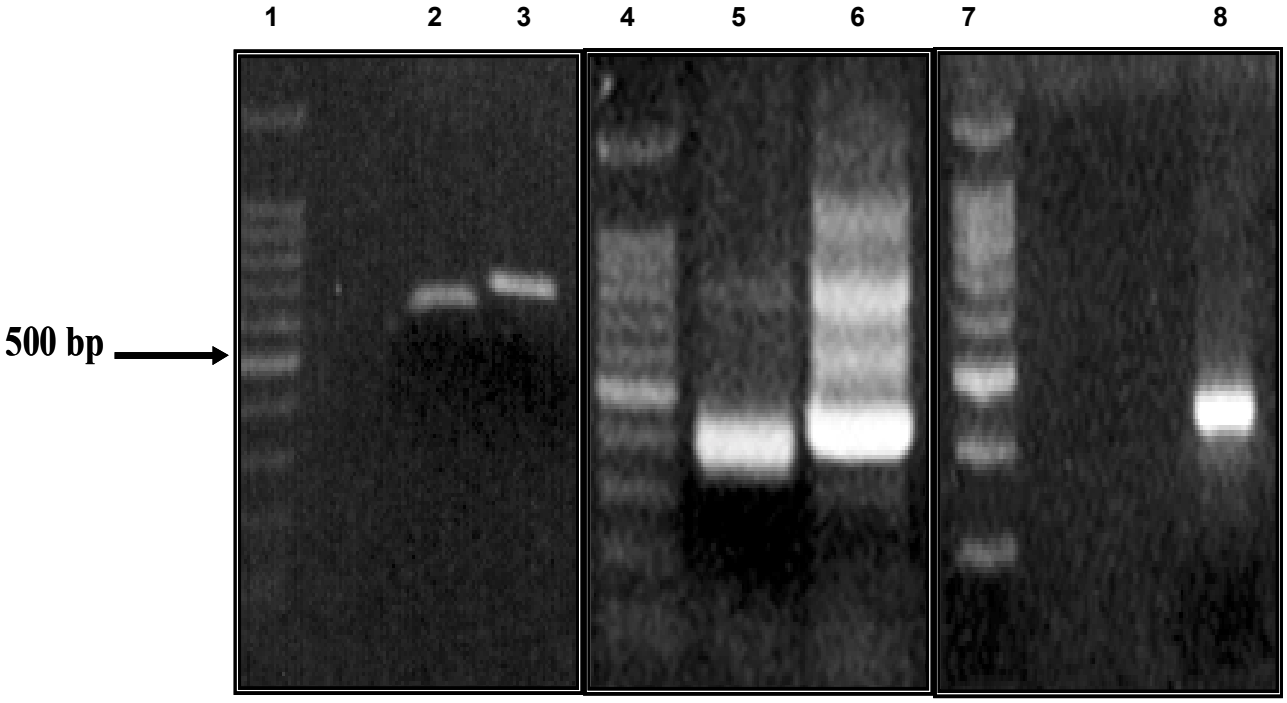

Figure 25. 1% agarose gel electrophoresis of segments of the codon-optimized *cry*2Aa gene after synthesis by OE-PCR of the 60 bp products of DA-PCR. Lane 1: 100 bp DNA ladder; 2: Synthon A – 525 bp; 3: Synthon B – 615 bp; 4: 100 bp DNA ladder; 5: Synthon C – 348 bp; 6: Synthon D – 432 bp; 7: 100 bp DNA ladder; 8: Synthon D after T7 endonuclease I treatment to remove the mismatch sites

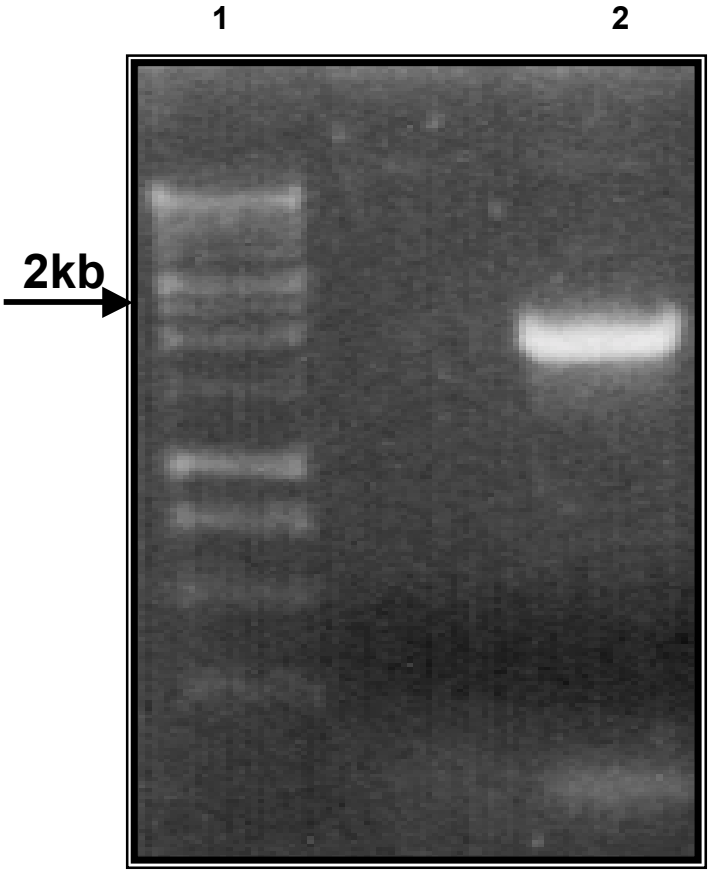

Figure 26. 1% agarose gel electrophoresis analysis of *cry*2Aa gene with 1 kb DNA ladder. The synthetic gene was at the correct size of 1.9 kb.

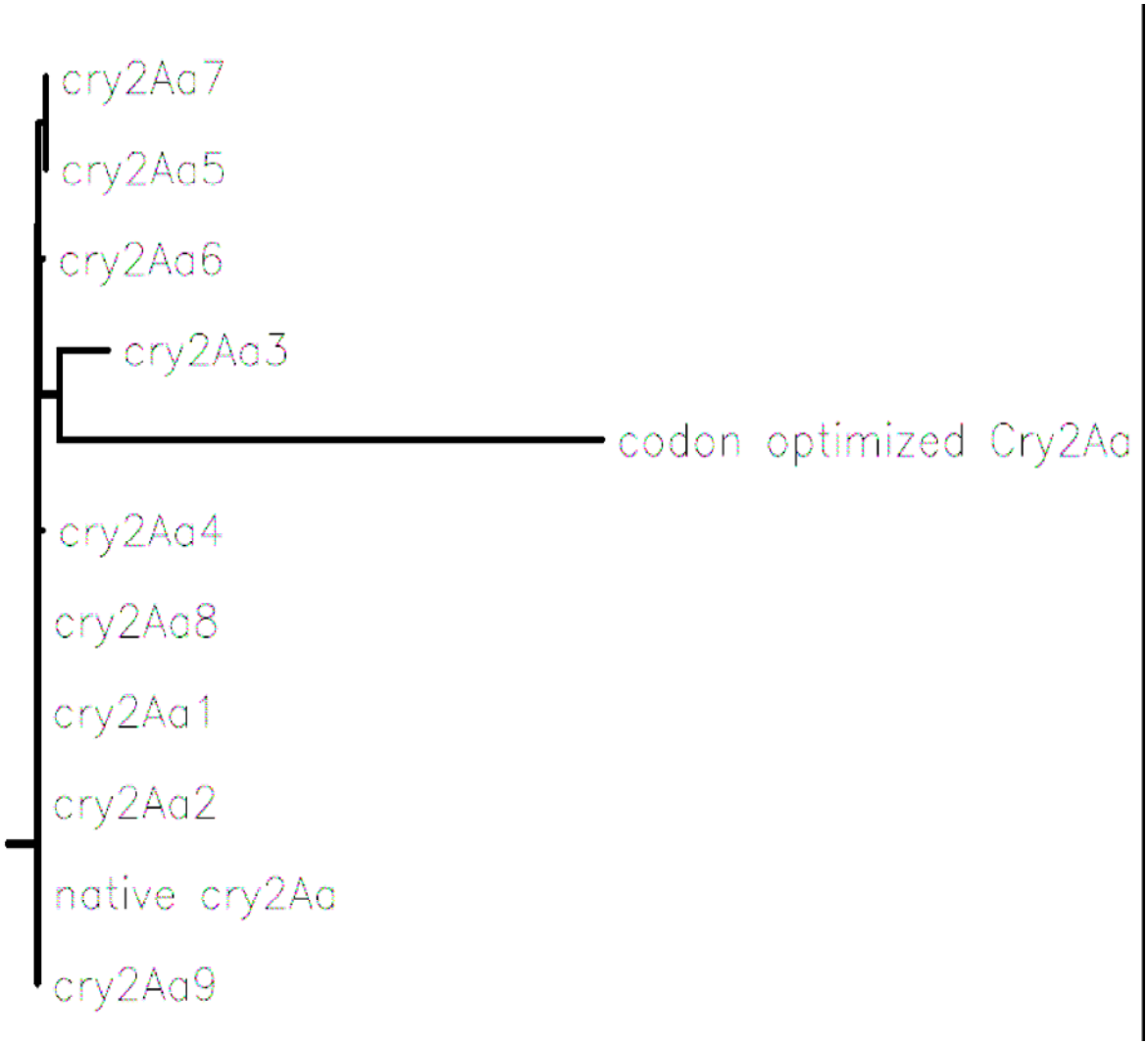

Figure 27. Phylogenetic tree analysis was performed on nucleotide sequence of *cry*2Aa genes using DRAWGRAM program in Biology workbench (Felsenstein 1989). The eggplant codon optimized *cry*2Aa gene was compared with native (wild type) *cry*2Aa and nine other *cry*2Aa genes.

```
                10        20        30        40        50        60
  1  CTCGAGATCAATAATTGTATTGAATAGTGGAAGAACAAGTATTTGTGATGGCTATAATGTA  wild_type_
  1  CTCGAGATCAATAACGTTCTCAATTCAGGACGAACTAGATTTTGTGATGGCTATAACGTC  Optimized_

                70        80        90       100       110       120
 61  GTAGGCGATGATCGATTTAGTTTTGAACATAAATGATTAGATAGCATCCAAAAGAATGG  wild_type_
 61  GTGGGACACGATCGTTTCTCTTTTGAACACAAATGCTTGGACACTATCCAGAAGGAATGG  Optimized_

               130       140       150       160       170       180
121  ATGGAGTGGAAAAGAAGAGATCATAGTTTATATGTAGGTCCTGTAGTCGGAACTGTGTCT  wild_type_
121  ATGGAGTGGAAGAGCAGCGATCATAGTCTTATGTGGCCGCAGTCGTTGGTACTGTTTCT  Optimized_

               190       200       210       220       230       240
181  AGTTTTTTGCTAAAGAAAGTGGGGAGTCTTATTGGAAAAAGCATATGCAGTGAATTATGG  wild_type_
181  TCTTTCCTTTGAAGAAGGTGGGCATCTCTTATCGGCAAAAGAATTCTTACTGAGTTGTGG  Optimized_

               250       260       270       280       290       300
241  GGCTTAATATTTTCCTAGTGGTAGTAGAATCTAATGCAAGATATTTTAAGGGAGACAGAA  wild_type_
241  GGTTTCATCTTTTCCTTCAGGTTCTAGTAACTTGATGCAAGTCATTTTGCGTGAACAGAG  Optimized_

               310       320       330       340       350       360
301  CAATTCGTAAATCAAGACTTAATAGAGATAGCCTTGGTCGTGTAAATGGGGAATTGATA  wild_type_
301  CAGTTTTCTCAATCAGAGATTGAATACTGAGCACATTGGGAAAGGTAAACGGTGAATTGATT  Optimized_

               370       380       390       400       410       420
361  GGGCTCCAAGGGAATTAAAGGGAGTTTAATCAAGAAGTAGATAATTTTTTAAACGCTAGT  wild_type_
361  GGTTTGCAGGCTAACATCAGGGAGTTTAACCAGCAGGTAGATAATTTCTTGAACCCAACT  Optimized_

               430       440       450       460       470       480
421  CAAAACCCTGTTCCTTTATCAATAACTTCTTCCGTTAATTACAATGCAGCAAATTATTCTA  wild_type_
421  CAAAACCCTGTTCCTTTGAGTATCACTTGGAGTGTAAACACAATGCAACAGTTGTTCCTG  Optimized_

               490       500       510       520       530       540
481  AATTAGATTAGGCGAGTTCGAGATAGAAGGATACGAGTTGTTATTATTTACGTTTATTTGCA  wild_type_
481  AATCGATTGCCACAATTCAAATTCAAGCTTATCAATTGTTGCTTTTCGATTGTTTGCT  Optimized_

               550       560       570       580       590       600
541  CAGGCAGGCAATATGCATCTTTCTTTATTAGAGATGTTATTGTTAATGCATGAATGAATGG  wild_type_
541  CAAGCAGCTAATATGCATCTTTCATTTATTAGAGATGTCATTTTGCAATGCTGACGAATGG  Optimized_
```

```
                 610       620       630       640       650       660
601  GGTATTTGAGGAGGAACATTACGTACGTATCGAGATTACCGCAGAAATTATRCAAGAGAT  wild_type_
601  GGTATCTGCGCGTGCCACTTTGCGTACATATAGAGACTATTTAAGCAACTACACACGAGAC  Optimized_

                 670       680       690       700       710       720
661  TATTCTAATTATTGTATAAATAGGTATCAAACTGCCTTTAGAGGCTTAAACACGCGTTTTA  wild_type_
661  TATTCTAACTACTGCATCAACACAATATCAAACTGCATTCAGCGGATTGAATACGAGGCTT  Optimized_

                 730       740       750       760       770       780
721  CACGATATGTTAGAATTTAGAACATATATGTTTTTAAATGTATTTGAATATGTATCCATT  wild_type_
721  CATGATATGCTTGAATTCAGAACTTATATGTTTTTGAATGTGTTTGAATATGTGTCCATA  Optimized_

                 790       800       810       820       830       840
781  TGGTCATTGTTTAAATATCAGAGTCTTATGGTATCTTCTGGCGCTAATTTATATGCTAGG  wild_type_
781  TGGAGTTTGTTCAAGTACCAGTCGTTGATGGTTTCATCTGGTGCTAATTTGTATGCTTCG  Optimized_

                 850       860       870       880       890       900
841  GGTAGTGGACCACAGCAGAGACATCATTTAGAGGACAAAACTGGGCATTTTTAATTCT  wild_type_
841  GGATCAGGACCACAGCAAACTCAGTCTTTCAGCGCTCAAAATTGGGCTTTCTTGTATTCT  Optimized_

                 910       920       930       940       950       960
901  CTTTTCGAAGTTAATTCGAATTATATATTATCTGGTATTAGTGGTAGTAGGCGTTCTATT  wild_type_
901  TTGTTCCAAGTCAATAGCAACTATATTCTATCCGGTATCAGCGGAACCAGATTGAGTATC  Optimized_

                 970       980       990       1000      1010      1020
961  AGCTTCCCTAATATTGGTGGTTTACGGGGTAGTACTAGAAGTCATTCATTGAATAGTGCC  wild_type_
961  ACTTTTGCTAACATTGGCGGACTTCCAGGAAGTACAAGTAGACATTCATTGAATTCTGCC  Optimized_

                 1030      1040      1050      1060      1070      1080
1021 AGGGTTAATTATAGCGGAGGAGTTTCATCTGGTCTCATAGGGGCGACTAATCGCAATCAC  wild_type_
1021 AGAGTCAACTATTCTGGAGGTGTTTCAACTGGATTGATTGGGGCAACAAATTTGAATCAT  Optimized_

                 1090      1100      1110      1120      1130      1140
1081 AACTTTAATTGCAGGACGTCCTCCTCTTTATCAAGAGCATTTGTTAGAAGTTGGCTG  wild_type_
1081 AACTTCAACTGTTCGACAGTTTTGCCTCCATTGTGCAGTGCTTTTGTGAGATCTTGGCTT  Optimized_

                 1150      1160      1170      1180      1190      1200
1141 GATTCAGGTACAGATCGAGAGGCGTTGCTAGCTCTAGCAATTGGCAGAGAATGCTTT  wild_type_
1141 GATTCAGGAACAGATCGTGAAGCAGTTGCTACATCTAGTAATTGGCAAACTGAATGATTT  Optimized_
```

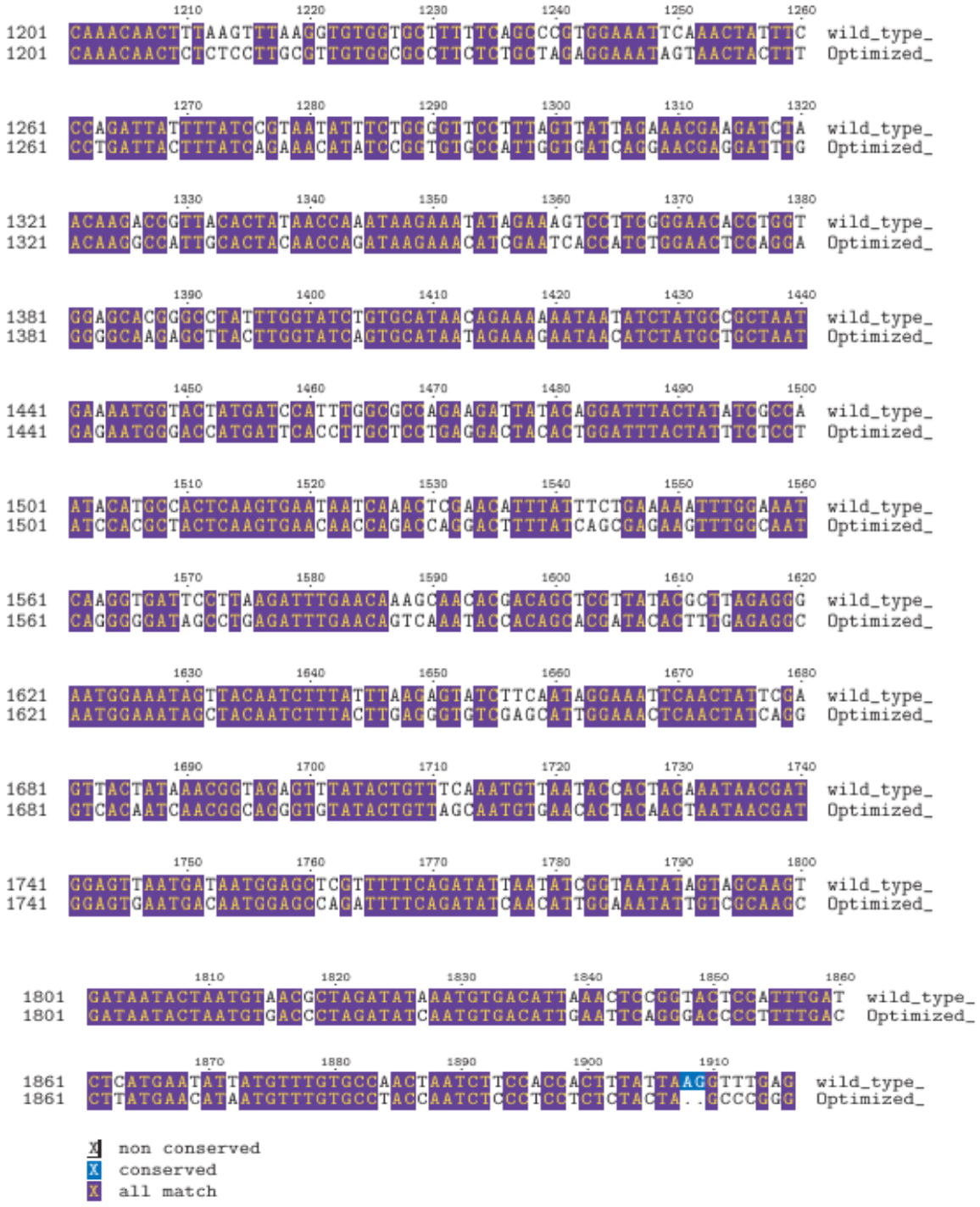

Figure 28. Gene sequence comparison of eggplant codon optimized *cry*2Aa and wild type (native) *cry*2Aa sequence. Texshade analysis was performed to compare the nucleotide sequence of native *cry*2Aa gene with the eggplant codon optimized *cry*2Aa gene. A total of 472 nucleotides were changed in the optimized version (Beitz 2000).

Eggplant optimized *cry*2Aa sequence

ATGAATAACGTTCTCAATTCAGGACGAACTACAATTTGTGATGCCTATAACGTCGTGGCACACGATCCTT
TCTCTTTTGAACACAAATCGTTGGACACTATCCAGAAGGAATGGATGGAGTGGAAGAGGACCGATCATAG
TCTTTATGTGGCCCCAGTGGTTGGTACTGTTTCTTCTTTCCTTTTGAAGAAGGTGGGATCTCTTATCGGG
AAAAGAATTCTTAGTGAGTTGTGGGGTATCATCTTTCCTTCAGGTTCTACTAACTTGATGCAAGACATTT
TGCGTGAAACAGAGCAGTTTCTCAATCAGAGATTGAATACTGACACATTGGCAAGGGTAAACGCTGAATT
GATTGGTTTGCAGGCTAACATCAGGGAGTTTAACCAGCAGGTAGATAATTTCTTGAACCCAACTCAAAAC
CCTGTTCCTTTGAGTATCACTTCGAGTGTAAACACAATGCAACAGTTGTTCCTGAATCGATTGCCACAAT
TTCAAATTCAAGGTTATCAATTGTTGCTTTTGCCATTGTTTGCTCAAGCAGCTAATATGCATCTTTCATT
TATTAGAGATGTGATTTTGAATGCTGACGAATGGGGTATCTCCGCTGCCACTTTGCGTACATATAGAGAC
TATTTAAGGAACTACACACGAGACTATTCTAACTACTGCATCAACACATATCAAACTGCATTCAGGGGAT
TGAATACCAGGCTTCATGATATGCTTGAATTCAGAACTTATATGTTTTTGAATGTGTTTGAATATGTGTC
CATATGGAGTTTGTTCAAGTACCAGTCGTTGATGGTTTCATCTGGTGCTAATTTGTATGCTTCCGGATCA
GGACCACAGCAAACTCAGTCTTTCACCGCTCAAAATTGGCCTTTCTTGTATTCTTTGTTCCAAGTCAATA
GCAACTATATTCTATCCGGTATCAGCGGAACCAGATTGAGTATCACTTTTCCTAACATTGGCGGACTTCC
AGGAAGTACAACTACACATTCATTGAATTCTGCCAGAGTCAACTATTCTGGAGGTGTTTCAAGTGGATTG
ATTGGGGCAACAAATTTGAATCATAACTTCAACTGTTCCACAGTTTTGCCTCCATTGTCCACTCCTTTTG
TGAGATCTTGGCTTGATTCAGGAACAGATCGTGAAGGAGTTGCTACATCTACTAATTGGCAAACTGAATC
ATTTCAAACAACTCTCTCCTTGCGTTGTGGCGCCTTCTCTGCTAGAGGAAATAGTAACTACTTTCCTGAT
TACTTTATCAGAAACATATCCGGTGTGCCATTGGTGATCAGGAACGAGGATTTGACAAGGCCATTGCACT
ACAACCAGATAAGAAACATCGAATCACCATCTGGAACTCCAGGAGGGGCAAGAGCTTACTTGGTATCAGT
GCATAATAGAAAGAATAACATCTATGCTGCTAATGAGAATGGGACCATGATTCACCTTGCTCCTGAGGAC
TACACTGGATTTACTATTTCTCCTATCCACGCTACTCAAGTGAACAACCAGACCAGGACTTTTATCAGCG
AGAAGTTTGGCAATCAGGGGGATAGCCTGAGATTTGAACAGTCAAATACCACAGCACGATACACTTTGAG
AGGCAATGGAAATAGCTACAATCTTTACTTGAGGGTGTCGAGCATTGGAAACTCAACTATCAGGGTCACA
ATCAACGGCAGGGTGTATACTGTTAGCAATGTGAACACTACAACTAATAACGATGGAGTGAATGACAATG
GAGCCAGATTTTCAGATATCAACATTGGAAATATTGTCGCAAGCGATAATACTAATGTGACCCTAGATAT
CAATGTGACATTGAATTCAGGGACCCCTTTTGACCTTATGAACATAATGTTTGTGCCTACCAATCTCCCT
CCTCTCTACTAG

Vita

Rahul Gupta
Department of Horticulture
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24060
Email: rahulg@vt.edu

Personal History
Born in Mukatsar, India on 25th January 1981

Education

- B.S., Agriculture, Punjab Agricultural University, Ludhiana, India, June 2003
- M.S., Horticuture, Virginia Polytechnic Institute and State University, Blacksburg VA, Dec. 2005

Professional Experience

Employer: Virginia Polytechnic Institute and State University
Type of Employment: Graduate Research Assistantship
Inclusive Dates: Aug. 2003-Dec. 2005