Characterization of novel cassava genes involved in cyanide detoxification, β-cyanoalanine synthase and cysteine synthase, using *Arabidopsis thaliana* mutants

by

Kevin Babilonia Figueroa

A thesis submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

In

Biology

UNIVERSITY OF PUERTO RICO
MAYAGÜEZ CAMPUS
2014

Approved by:

Timothy Porch, PhD Member, Graduate Committee	Date
Rafael Montalvo, PhD Member, Graduate Committee	Date
Dimuth Siritunga, PhD President, Graduate Committee	Date
Aixa Rivera, PhD Representative of Graduate Studies	Date
Nanette Diffoot Carlo, PhD Chairman, Biology Department	Date

ABSTRACT

Cassava (*Manihot esculenta* crantz) is a perennial shrub of the family Euphorbiaceae and is one of the most important foods in the world. As a subsistence crop, it is highly competitive due to its advantages in production such as high yields per hectare. Cassava can be easily cultivated and is tolerant to extreme stress conditions such as drought. It is propagated very easily as cuttings with up to 20 cuttings being obtained from one plant. Its tuberous roots are rich in starch, being an excellent source of calories for humans.

Besides the advantages of cassava, it contains high amounts of cyanogenic glycosides, which are compounds that yield free cyanide. The cyanogenic glycoside of major abundance in cassava is linamarin, which is synthesized from valine in reactions that occur in the vacuole of the plant cell. Linamarin can be hydrolyzed by the cell wall associated linamarase, especially when there is tissue rupture, producing acetone cyanohydrin. Therefore, poorly processed cassava roots will accumulates toxic levels of acetone cyanohydrin, which can cause cyanide associated health problems such as Konzo, Tropical Ataxic Neuropathy and Goitre. Despite the fact that cassava has a cyanide detoxification pathway, it is not enough for the removal of all the cyanide in some varieties.

Recently, two genes from the cassava detoxification pathway, *MANes;BsasA* and *MANes;BsasB*, where isolated and based on enzyme assays *MANes;BsasB* may be involved directly in the cyanide detoxification pathway and *MANes;BsasA* in cysteine biosynthesis, which is also important in the cyanide detoxification pathway. Prior to utilizing these novel genes for genetic transformation of cassava they have been studied using mutants of the model plant *Arabidopsis thaliana*. Mutants have target genes in the cyanide detoxification pathway silenced thus enabling the functional complementation with the novel cassava genes.

Two different expression vectors (pKYLX and pB2GW7) were used in this study for the transformation of the plants with *Agrobacterium tumefaciens*. The floral dipping method was performed using many modifications since false positive transgenics were obtained in multiple experiments. However, the selection protocol employed was unable to confirm transgenic lines with the T-DNA integrated into its genome. It is believed that the concentration of the surfactant Silwet L-77 that was used for the transformation protocol could have interfered with the transformation mechanism. The plant stage can also be a factor for the lack of transgenic plants due to the importance of the right flowering stage in the floral dipping method. It is also possible that a mutation in any of the virulence genes from Agrobacterium resulted in no T-DNA transfer due to the importance of the virulence genes in this process. Furthermore, mutations or gene rearrangements in areas like the selectable marker or its promoter, or the T-DNA borders, could also affect the integration of the T-DNA into the plant genome. Detailed sequencing of the T-DNA region could be employed to verify that there are no mutations.

RESUMEN

La yuca es un arbusto perenne de la familia Euphorbiaceae y es uno de los cultivos más importantes del mundo. Este cultivo de subsistencia es altamente competitivo debido a sus ventajas en producción, como por ejemplo su alta producción por hectárea y su flexibilidad en cuanto a su sembradía y cosecha. La yuca es fácilmente cultivada y es tolerante a condiciones extremas como la sequía. La misma puede ser propagada fácilmente con estacas, obteniendo unas 20 estacas por planta. Sus raíces son ricas en almidón, lo que la hace una excelente fuente de calorías para los humanos.

A pesar de todas las ventajas que posee la yuca esta contiene cantidades altas de glucósidos cianogénicos, los cuales son compuestos que producen cianuro. El glucósido cianogénico de mayor abundancia en la yuca es linamarina, el cual es sintetizado en reacciones que ocurren en la vacuola de la célula. Linamarina puede ser hidrolizada por la enzima linamarasa, la cual se encuentra en la pared celular y entra en contacto con la linamarina, especialmente cuanto hay rompimiento del tejido, produciendo así cianuro. Por lo tanto, si la yuca no es procesada correctamente puede acumular niveles tóxicos de cianuro que pueden ser responsables de condiciones de la salud como “Konzo”, “Tropical Ataxic Neurophaty (TAN)” y “Goitre”. A pesar de que la yuca está equipada con un mecanismo para eliminar el cianuro, el mismo no es suficiente para remover los niveles tóxicos de cianuro en algunas variedades.

Recientemente se aislaron dos genes de yuca que están envueltos en el proceso de detoxificación de cianuro, *MANes;BsasA* y *MANes;BsasB*. Según los ensayos enzimáticos realizados se cree que *MANes;BsasB* funciona directamente en la detoxificación de cianuro, mientras que *MANes;BsasA* funciona principalmente en la formación de cisteína. Antes de utilizar los genes para la transformación genética de yuca, los mismos han sido estudiados utilizando mutantes de la planta modelo *Arabidopsis thaliana*. Los mutantes tienen genes del proceso de detoxificación de cianuro silenciados, permitiendo la caracterización de los genes de yuca.

En este estudio se utilizaron dos vectores de expresión (pKYLX y pB2GW7) para transformar Arabidopsis con *Agrobacterium tumefaciens*. Se utilizó el método de transformación conocido como “floral dipping”, pero no fue posible confirmar plantas transgénicas mediante nuestro protocolo de selección. Es posible que la concentración del detergente Silwet L-77, el cual es importante en la transformación debido a que remueve la tensión superficial y permite que la bacteria pueda penetrar fácilmente el tejido para la infección, haya interferido con el proceso. Otras posibles explicaciones son la edad de las plantas al momento de transformarse o que la solución de Agrobacterium con Silwet L-77 no se haya escurrido lo suficiente, provocando daños al tejido. También es posible que una mutación en alguno de los genes de virulencia de Agrobacterium resultara en un mecanismo de inserción del ADN defectuoso. Por otro lado, mutaciones o desplazamientos de los genes en otras partes de los vectores pudieron haber afectado la transferencia de nuestros genes de interés al genoma de la planta. En una próxima ocasión sería necesario secuenciar todos esos componentes para verificar que no hay mutaciones.

ACKNOWLEDGEMENTS

First of all I want to thank God because without him I would not be where I am today. Secondly, I want to thank my family for being there in the good and bad moments of my life. To my parents, Noemi Figueroa and William Babilonia, who have supported me both economically and emotionally, and to my brother Brian Babilonia, for being such a special person in my life. I had excellent moments with you and I can say that the three of you are a part of me.

I also want to thank my graduate committee. Thanks you to Dr. Dimuth Siritunga, who has given me the opportunity and the resources to work in his lab and be part of this and other research projects. Although I had many difficulties he always supported me and helped me until the end. Also, Dr. Rafael Montalvo and Dr. Timothy Porch for being available for me whenever I had questions or needed support.

Along with that are my lab partners, especially Lorraine Rodriguez, who has been the best lab partner in the world. Thank you for your help, but especially for your friendship, jokes and for bringing food to the lab to feed my hungry stomach. Thank you also to the ones who were my students and helped me a lot with my experiments: Clara Cruet, Gabriela del Mar and Shaimar Gonzalez. Thank you to the other lab members who have also helped me and became my friends: Cristina, Nate, Rocio S., Rocio M., Lorraine K., Derickson and Dianiris. I also want to thank Andrea McCue who helped me a lot with the constructs and with the protocols and ideas.

Thank you to the personal in the biology building which have contributed equipment or reagents. In conclusion, to any other person that I may have forgotten to mention here but that have helped me in one way or another, thank you.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

Introduction and Literature Review

1 Introduction

1.1 Cassava (*Manihot esculenta* Crantz)

1.1.1 The biology of cassava and its importance as a crop

Cassava (*Manihot esculenta* crantz) is a perennial shrub of the family Euphorbiaceae (Alves, 2002) and is one of the most important crops in the world (Figure 1). It is largely cultivated for its nutritional and, more recently, industrial importance. Cassava is used as a source of calories, feeding 500 million people worldwide. Cassava plays a particularly important role in farming in the developing countries because it grows well in poor soils and dry grounds. As a subsistence crop, it is highly competitive because it has advantages in production, such as high yields per hectare and the flexibility in planting and harvesting. The plant grows to a height of one to three meters (Tonukari, 2004) and can produce 8-15 roots of high starch content. It is widely cultivated in countries in America, Asia and Africa, but Nigeria (West Africa) is the world largest producer of the crop. In Nigeria cassava flour and a processed cassava product called gari are consumed daily. Depending on the country cassava is named differently, for example mandioca, maniot, manioc, pai, cui, yucca, cassava, etc. Cassava can be prepared for consumption as flour, bread (Casabe), pudding, torta, fried, beverages (yaraque, cachiri) and snacks.

1.2.2 Advantages of the crop

Cassava is a plant that can be easily cultivated and is tolerant to extreme stress conditions such as drought, thus fitting well with traditional farming systems. Some of

the responses that allow the plant to tolerate water stress are control of stomatal closure and osmotic adjustment (Alves, 2002). Furthermore, cassava yields well even in poor soils without fertilizer and the root can be left in the ground for up to three years as a reserve source of food (Bradbury and Denton, 2010).

Cassava is propagated very easily as cuttings with up to 20 cuttings being obtained from one single plant. This type of propagation minimizes the unintentional spread of the crop by humans (Chavarriaga and Halsey, 2005). Cuttings are taken from the mature part of the stems and should have 4 to 5 nodes with viable buds (at least six inches long) that serve as the origins of shoots and roots. Cuttings produce roots within a few days and typically, harvesting can begin from eight to ten months after planting.

Figure 1. Cassava plant and roots. (Source: http://www.sacredearth.com)

Another advantage of the crop is the high content of starch in tuberous roots, which comprises about 90% of the dry weight. Among the principal starchy crops in the world, its carbohydrate production is about 40% higher than rice and 25% more than maize, meaning that cassava is the cheapest source of calories for humans (Tonukari, 2004). Starch is one of the most important plant products for humans because it plays an important role in human diet, providing a large quantity of daily intake of calories (Tonukari, 2004). Because of the high starch content, cassava also has other

applications such as in the production of adhesives, dyes, food thickeners, animal foods, ethanol, among others.

Similar to other plants, cassava has developed an elaborate chemical defense to protect itself against herbivores. This is accomplished by the production of secondary metabolites known as cyanogenic glycosides, which are molecules that contain hydrogen cyanide bound to a sugar molecule (See section 1.2 for further details). When an organism attacks the plant, those compounds are activated and free cyanide is liberated (together with a sugar and a ketone compound), poisoning the attacking organism by interfering with cellular respiration (Tattersall *et al*., 2001). Although cyanogenic glycosides are beneficial for plant defense against herbivores, they can also be dangerous to humans if cassava is not fully processed before being consumed.

1.1.3 Cassava economy

Cassava is mostly consumed in developing countries and since 1984 its use as a food has increased by millions of tons per year. The growth is reflecting a dynamic expansion of cassava production, especially in Africa, where the contribution of cassava to human calorie intake averaged approximately ten percent (FAO, 2000). For example, cassava production in Nigeria is the highest worldwide, with a half more production than in Brazil. However, cassava production in other African countries like the Congo, Madagascar, Mozambique and Uganda, is far lower as compared to Nigeria (Figure 2 and 3; FAO, 2004) .

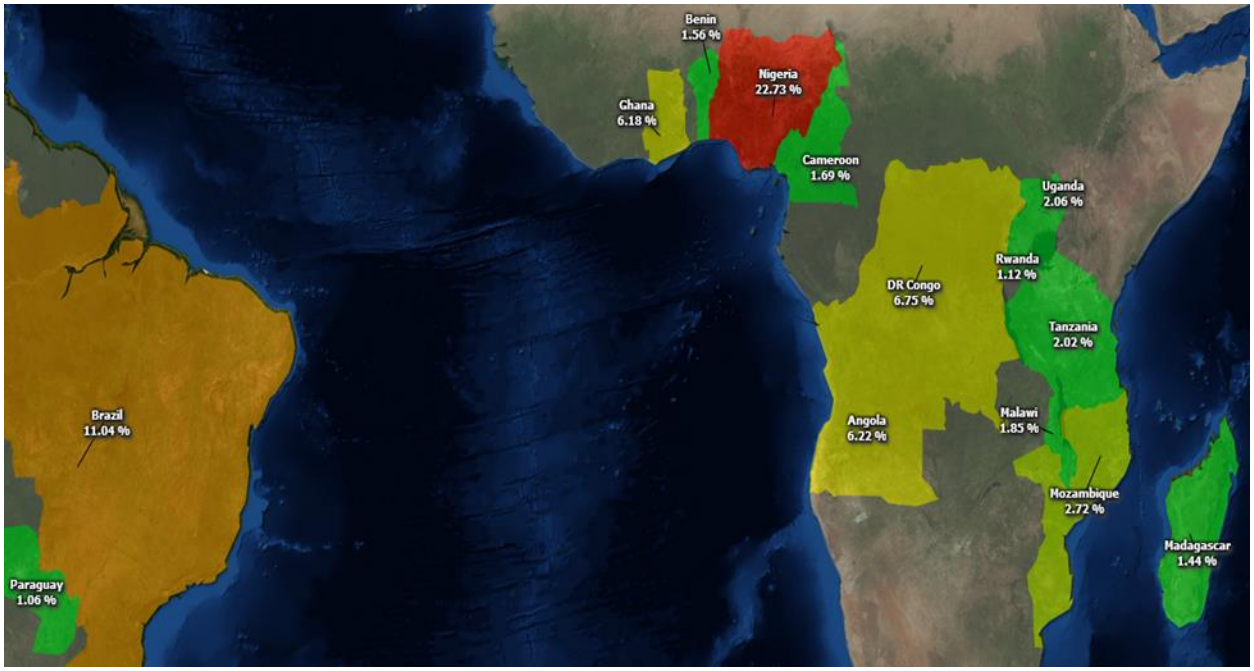

Figure 2. Cassava world production in 2011. (Source: http://www.targetmap.com/viewer.asp)

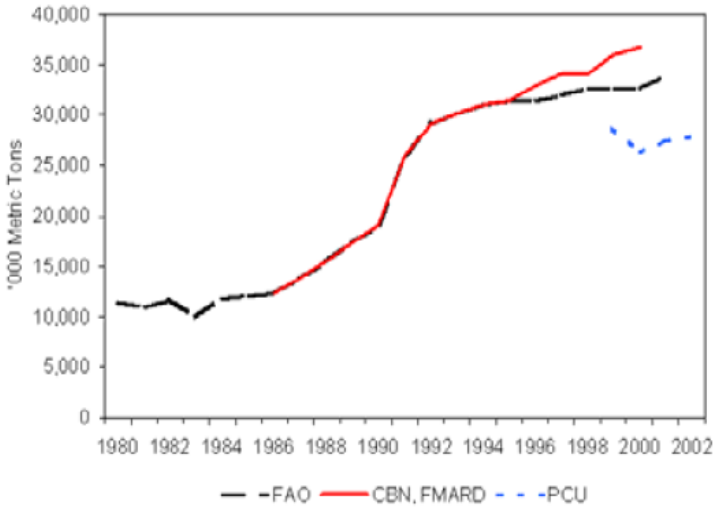

Figure 3. Three estimates for cassava production in Nigeria from 1996 to 2002. (Source: *Food and Agriculture Organization* (FAO) of the United Nations)

1.2 Cyanide in cassava

1.2.1 Cyanogenic glycosides

Besides the advantages of cassava, it contains high amounts of cyanogenic glycosides, which are compounds that yield free cyanide. The cyanogenic glycoside of major abundance in cassava is Linamarin, and is accumulated in all of its tissues (Nambisan, 2011; Marrero *et al*., 2010). It constitutes more than 95% of total cyanogens in plants. The other 5% is constituted by the cyanogenic glycoside Lotaustralin. These molecules serve as defense mechanisms against herbivores and pathogens, as discussed earlier, and also as storage compounds of reduced nitrogen which can be incorporated in other metabolic pathways (Siritunga and Sayre, 2004; Sayre et al., 2011). The total concentration of cyanogenic glycoside will be dependent on the cultivar, the environment and plant age (Alves, 2002), but on average cassava leaves and roots contain 5 g/kg and 100 – 500 mg/kg, respectively (White *et al*., 1998). Typically within the root, cortex can contain more than 400 mg/kg of HCN and the root parenchyma less than 100 mg/kg. For this reason, poorly processed cassava is very dangerous for humans. For example, in Mozambique the mean total cyanide in cassava flour can be 45 mg/kg (Bradbury, 2009), which is above the FAO safe level of 10 mg/kg (Nambisan, 2011).

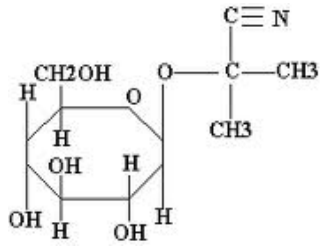

Figure 4. Linamarin molecule

1.2.2 Linamarin synthesis

Linamarin is synthesized in the leaves and transported to the roots. It is synthesized from valine in reactions that occur in the vacuole of the plant cell (Gallo and Sayre, 2009). The first step in linamarin synthesis is the hydroxylation of L-valine to N-hydroxyvaline, which is then is converted to 2-methylpropanol oxime. These two first steps are catalyzed by the two multifunctional cytochrome P450s, CYP79D1 and CYP79D2, respectively (Jorgensen *et al*., 2010; Gallo and Sayre, 2009; Siritunga and Sayre, 2007). Then a putative CYP71E catalyze the reaction of acetone cyanohydrin synthesis from 2-methylpropanol oxime. The next step is the glycosilation of acetone cyanohydrin by the enzyme indoxyl-uridine diphosphate glucose (UDPG)-glucosyltransferase, which finally produces linamarin (Siritunga and Sayre, 2007; Jorgensen *et al*., 2005). If the produced linamarin is not metabolized or transported then it is store in the vacuole of the plant cell (Siritunga and Sayre, 2004).

1.2.3 Cyanogenesis

Because of the different subcellular localizations of linamarin and the enzymes that convert it to cyanide, it has been concluded that cyanogenesis is initiated when there is tissue rupture (Gallo and Sayre, 2009). That way, the rupture of the cell vacuole liberates linamarin and it can then be hydrolyzed by the cell wall associated linamarase (White *et al*., 1994) producing acetone cyanohydrin. Linamarase is a very stable β-glucosidase enzyme and is also abundant in laticifers (O'Brien and Ip, 2008; Siritunga and Sayre, 2004). Acetone cyanohydrins can be converted to cyanide and acetone spontaneously, at pH >5.0 and at temperatures greater than 35°C (White *et al*., 1994), or they can be cleaved by the enzyme hydroxynitrile lyase (HNL). In the leaves, HNL

and linamarase have similar catalytic efficiencies, but in the roots there is much lower linamarase activity and very little or no HNL activity, meaning that poorly processed cassava roots will accumulates toxic levels of acetone cyanohydrin (Siritunga and Sayre, 2007).

1.3 Cyanide associated health problems

The cyanogen levels of cassava roots (>400mg/kg) and leaves are higher than the maximum levels recommended for foods, thus cassava has to be well processed before being consumed (Nambisan, 2011). In Africa, several cyanide health disorders are associated with eating poorly processed cassava and the severity of the disorders depend on the amount or frequency of cyanide exposure. The most common symptoms of cyanide intoxication are headache, nausea, vomiting, dizziness, stomach pain, diarrhea, exacerbated goiter and cretinism, even death (Bradbury *et al.*, 2011).

A common cyanide health disorder in Africa is named “Konzo”, characterized by an irreversible, permanent spastic paralysis, especially in the legs (Bradbury and Denton, 2011). The disease is known to predominantly affect children and young women. Cliff *et al*. (2011) found that since 1981, poor rural communities in some areas of Mozambique and other countries in Africa have suffered from Konzo, both in epidemic and persistent form. They found that Konzo has been reported from more districts in the decade of 2000-2009 than previous decades, and that epidemics occur mostly during agricultural crises, like drought, when the population is more dependent on a diet involving insufficiently processed cassava. This can happen because people take short cuts in processing cassava because waiting several days for consumption after proper processing is difficult due to hunger. Compounding this effect is the failure

of other crops during drought, leaving cassava the predominantly surviving crop during drought. They concluded that konzo is spreading to new areas in Africa as cassava cultivation increases and is occurring outside major agricultural crises with persistent cases and epidemics.

Another cyanide health disorder is Tropical Ataxic Neuropathy (TAN), characterized by unsteady walking, loss of sensation in hands, deafness, loss of vision and weakness (Bradbury, 2009). The disease is common among older people in Nigeria who have consumed ‘gari' and other cassava products over many years. Many of the cases do not respond to improvements in diet or vitamin supplementation. Oluwole *et al.* (2000) determined the prevalence of TAN in Ososa, Nigeria, and found that 323 of 3428 subjects aged 10 years and above were positive for the disease (Table 1). They concluded that TAN incidence is still a problem since 30 years ago in Nigeria and deserves both public health and scientific attention.

Table 1. Age and sex distribution of subjects screened for TAN in Nigeria (Oluwole *et al.,* 2000)

Age group (y)	*Screened*		*TAN patients*		*Prevalence (%)*	
	Male	*Female*	*Male*	*Female*	*Male*	*Female*
10–19	566	508	3	3	0.5	0.6
20–29	222	356	1	4	0.5	1.1
30–39	186	272	1	10	0.5	3.7
40–49	161	199	5	19	3.1	10.0
50–59	134	193	5	36	3.7	18.7
60–69	153	168	24	41	15.7	24.4
70–79	81	108	14	24	17.3	22.2
>80	44	77	8	8	18.0	10.4
Total	1547	1881	61	145	3.9	7.7

Goitre is another cyanide-related health disorder common in Nigeria (Figure 5). Goitre is characterized by an enlarged thyroid gland, which can result in an abnormality

of thyroid function. The ingested cyanide competes with iodine in the thyroid gland *(El Ghawabi et al.*, 1975), leading to iodine deficiency which is the most common cause for goiter. Cyanide inhibits the uptake of iodine by the thyroid gland, and when iodine supply is below the recommended daily intake, goiter occurs.

Figure 5. Person with Goitre. (Source: http://www.asnom.org)

1.4 Cyanide detoxification

Cassava also is equipped with a pathway to metabolize the cyanide produced to aspartate and ammonium. The first step is catalyzed by the enzyme β-cyanoalanine synthase (β-CAS), which converts cyanide to β-cyanoalanine with the addition of a cysteine (Figure 6). Then, β-cyanoalanine is converted to asparagine by the β-cyanoalanine hydratase, while the deamination of asparagine by asparaginase releases aspartate and ammonia (Marrero *et al*., 2010).

β-CAS, the most common enzyme involved in the cyanide detoxification pathway, release sulfide in the reaction, which is recycled back by cysteine synthase (CS) (Lai *et al*., 2009; Marrero-Degro *et al*., 2010), the enzyme that catalyze the last step in cysteine biosynthesis. The activity of β-CAS has been detected in bacteria, insects, and plants and sub cellular fractionation studies indicate that it is located in the mitochondria (Warrilow and Hawkesford, 1998; Hatzfeld *et al*., 2000a). For example,

Hatzfeld *et al*. (2000a) found that β-CAS activity in *Arabidopsis thaliana* and *Spinacia oleracea* is predominantly found in the mitochondria. Also, β-CAS activity has been demonstrated in mitochondrial fractionation studies in barley leaves (Wurtele *et al*., 1985) and blue lupin seedlings (Akopyan *et al*., 1975). Having mitochondrial β-CAS facilitates the effective removal of cyanide to protect the oxidative phosphorylation process in the mitochondria. Furthermore, the pH in the matrix of mitochondria is 8.5, the optimal for β-CAS activity (Lai *et al*., 2009). In plants the main cyanide detoxification process is by β-CAS enzymatic activity. The role of β-CAS in cyanide detoxification has been demonstrated by feeding plants with labeled cyanide and observing the incorporation of the radioactive carbon initially to β-cyanoalanine and then into asparagine (Lai *et al*., 2009).

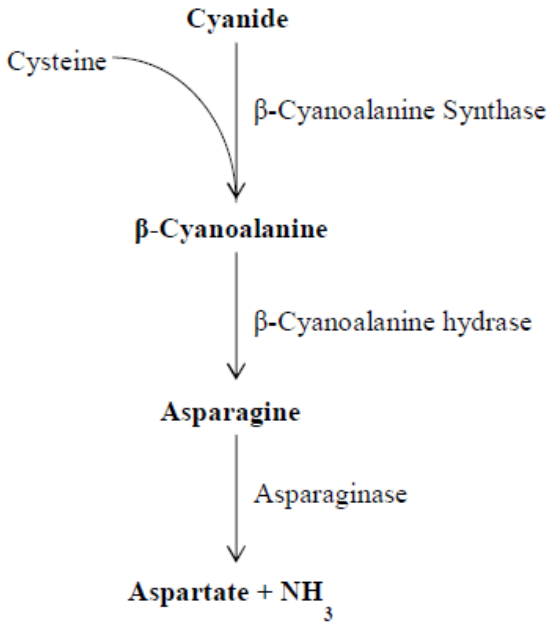

Figure 6. Cyanide detoxification in higher plants. The first step of cyanide detoxification is catalyzed by β-CAS, which converts cyanide to β-cyanoalanine with the addition of a cysteine. β-cyanoalanine is then converted to asparagine by the β-cyanoalanine hydratase, followed by the deamination of asparagine by asparaginase releasing aspartate and ammonia. β-CAS, β-cyanoalanine synthase; NH_3, ammonia.

The requirement of cysteine in the first step of cyanide detoxification makes cysteine a limiting factor because β-CAS is not able to metabolize cyanide under cysteine deficiency. In humans, cysteine is also important in cyanide detoxification in the gut when foods with high cyanide levels are consumed. If cyanide is consumed under a low protein diet, the probability of getting a cyanide-related disorder increases. Interestingly, the composition of the β-CAS enzyme is very similar to that of CS enzyme, which catalyzes the formation of cysteine from O-acetylserine and bisulfide. Both enzymes belong to the β-substituted alanine synthases (*Bsas*) family of enzymes, which catalyze reactions of β-replacement producing alanine derived amino acids (Marrero-Degro *et al*., 2010).

According to kinetic analysis, the proteins encoded by the *Bsas* gene family could be divided into two groups: the true CS protein, which exhibit bi-functional CS and β-CAS activities, and the CS-like proteins, which preferentially catalyze the formation of β-Cyanoalanine (Jost *et al*., 2000; Warrilow and Hawkesford, 2000; Burandt *et al*., 2002). Within the *Bsas* family it is impossible to predict function from amino acid sequence only because of the high homology in sequence between organisms (Jost *et al*., 2000). The true CS proteins probably do not function in cyanide detoxification or secondary metabolism, having high K_m values for KCN, while the K_m values for CS-like proteins are low (Jost *et al*., 2000). An authentic CS can be defined by its ability to interact with serine acetyltransferase (SAT), which has been demonstrated using different approaches (Droux *et al*., 1998; Bonner *et al*., 2005; Heeg *et al*., 2008). Purified β-CAS exhibit detectable CS activity, while CS displays detectable CAS activity, but each enzyme is capable of catalyzing its corresponding activity at a higher extent

(Maruyama *et al*., 1998; Hatzfeld *et al*., 2000a; Marrero-Degro *et al*., 2010). For example, in some plants β-CAS activity is present in the cytosol and chloroplast fractions, but it is attributed to CS. Because enzymes catalyze a balance between forward and backward reactions or between substrate and product formation, the enzymes of the *Bsas* family can use and release H_2S and OAS in a back reaction (Burandt *et al*., 2002). *Arabidopsis* genome codes for nine *Bsas* genes, and only one is the true β-CAS localized in the mitochondria (Table 1) (Watanabe *et al*., 2008). Until now, *Arabidopsis* is the only plant identified with a mitochondrial CS in addition to β-CAS, probably due to a duplication event (Hesse *et al*., 1999; Jost *et al*., 2000).

1.5 Cysteine in plants

1.5.1 Cysteine biosynthesis

In plants, cysteine biosynthesis is a fundamental process that provides the entry of sulfur into cellular metabolism for protein biosynthesis and anabolic pathways (Jost *et al*., 2000). Cysteine biosynthesis can be divided in three major steps: (1) assimilation of sulfide by sulfate activation and reduction; (2) synthesis of the cysteine precursor O-acetylserine (OAS); and (3) combination of OAS and sulfide to yield cysteine (Figure 7a). The sulfate reduction and assimilation pathways are fairly clear but the different cellular compartments of the enzymes involved and the fact that the enzymes belong to multigenic families makes it complex. It is known that chloroplasts contain the entire pathway for cysteine biosynthesis from inorganic sulfide, thus it is the primary site for cysteine biosynthesis (Figure 7b) (Shibagaki *et al*., 2002). Some of the sulfur assimilation enzymes exist in other cellular compartments outside of plastids, but the function of those enzymes is not well understood. Sulfide is incorporated into the β-

position of the amino acid skeleton of OAS by the enzyme CS to yield cysteine and acetate in a β-replacement reaction (Figure 7a) (Tai and Cook, 2000; reviewed in Droux, 2003). CS contains a pyridoxal phosphate as a prosthetic group. The carbon and nitrogen-containing precursor of cysteine, OAS, is formed from serine and acetyl-CoA by serine acetyltransferase (SAT) in a reaction specific to cysteine biosynthesis (Mendoza-Cózatl *et al*., 2005; Heeg *et al*., 2008). In the reaction, SAT transfers an acetyl moiety from acetyl-CoA to serine, and that is subsequently replaced with sulfide by the CS to release cysteine. There are several possible sources of acetyl-CoA in plants. Serine is synthesized from 3-phosphoglycerate and glycine, formed by the Calvin cycle in chloroplasts and photorespiration in the mitochondria, respectively (Linka and Weber, 2005).

1.5.2 Importance of cysteine

Cysteine is an indispensable proteinogenic amino acid (Heeg *et al*., 2008) in secondary and tertiary protein structures due to the covalent disulfide bond formed between thiol groups of two cysteine residues. Cysteine also plays an important role in defense against abiotic and biotic stresses (Leustek *et al*., 2000; Noctor *et al*., 2002). Glutathione (GSH), which is synthesized from cysteine, is involved in many aspects of growth and development, including redox control, storage and transport of reduced sulfur, metabolism of herbicides, and response to environmental stresses. GSH can also be used as a storage form of reduced sulfur in the vacuole (Mendoza-Cózatl*et al*., 2005). Phytochelatins are low-molecular-mass thiols derived from GSH that play a role in protection against toxic heavy metals due to the binding of the phytochelatins to metal ions using the thiol group as the ligand (reviewed in Cobbett, 2000). Other sulfur-

containing compounds formed by plants are thioethers, sulfoxides, methylsulfonium compounds, sulfate esters, sulfamates, and sulfonic acids. Main examples are vitamins and coenzymes such as coenzyme A, S-adenosyl-L-methionine, thiamine, biotin and S-methylmethionine (vitamin U). In addition, some signaling molecules also contain sulfur.

1.5.3 Compartmentalization

Cysteine synthase is localized in three compartments: cytosol, mitochondria and chloroplasts, and the ratio among these isoforms vary between compartment (Figure 7b) (Wirtz *et al*., 2001; Jost *et al*., 2000). The genome of *Arabidopsis thaliana*, the best investigated plant in respect to sulfur metabolism, codes for eight different CS (Wirtz *et al*., 2004) genes, similar to other species (Wirtz and Hell, 2006) (Table 2). The other CS in the genome of Arabidopsis is classified as β-CAS, which is in the mitochondria. Of these at least three cDNA for each enzyme have been isolated (Noji *et al*., 1998; Jost *et al*., 2000). It is accepted that cysteine is exchangeable between compartments (Zhang *et al*., 2008), but plastids, mitochondria and cytosol contains independent cysteine biosynthesis proteins, thus cysteine could also be synthesized in each of these compartments (Figure 7b) (Heeg *et al*., 2008; Haas *et al*., 2008). However, the fact that sulfate reduction is in the chloroplast implies that sulfide must be transported from chloroplast to other cell compartments (Figure 7b) (Mendoza-Cózatl *et al*., 2005). Because different CS isoforms exist in the three cell compartments, different subcellular cysteine amounts are produced (Alvarez *et al*., 2009). However, the interaction between the subcellular locations and the contribution to the overall cysteine synthesis are still unknown (Heeg *et al*., 2008).

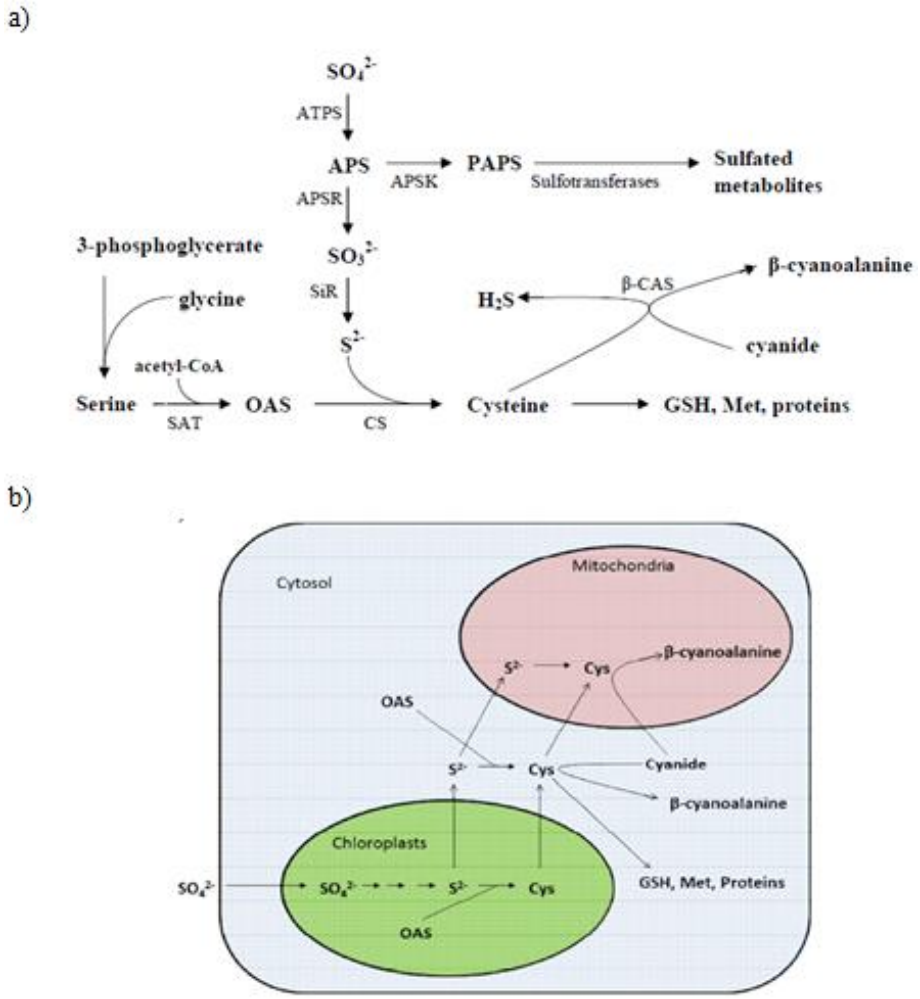

Figure 7. Cysteine Biosynthesis pathway and its compartmentalization. (a) After assimilation through the roots into the plant cell, sulfate is converted to APS in a reaction catalized by ATPS. APS can be converted to PAPS by APSK for the production of sulfated metabolites, but it mostly feed the sulfate reduction pathway, which consists of its reduction to free sulfite by the enzyme APSR. Sulfite is reduced to sulfide by SiR to then be incorporated into OAS by CS to yield cysteine. OAS is produced from serine and acetyl-CoA by the enzyme SAT. Serine is synthesized from 3-phosphoglycerate formed in the Calvin cycle and glycine formed in photorespiration. Cysteine can be used to produce GSH, Met and proteins, or it can be used to detoxify cyanide generated in the ethylene cycle or by metabolization of cyanogenic glucosides in higher plants. Cyanide is detoxified to β-cyanoalanine by the enzyme β-CAS, releasing hydrogen sulfide. (b) Sulfate is transported from the outside to the inside of the cell, specifically into chloroplasts, being the only organelle equipped with the complete pathway for sulfate reduction. Sulfide formed in chloroplast can be transported to cytosol and mitochondria, where cysteine biosynthesis also occurs, or it can be incorporated into OAS to produce cysteine in chloroplast. Cysteine can also be transported between cell compartments. In the cytosol, cysteine can be used for the synthesis of GSH, Met and proteins or for cyanide detoxificacion, while in mitochondria it can also detoxify cyanide to protect the oxidative phosphorilation process. SO_4^{2-}, sulfate; ATPS, ATP sulfurylase; APS, adenosine 5'-phosphosulfate; APSK, APS kinase; PAPS, 3'-phosphoadenosine 5'-phosphosulfate; APSR, APS reductase; SO_3^{2-}, sulfite; SiR, sulfite reductase; S^{2-}, sulfide; CS, cysteine synthase; SAT, serine acetyltransferase; β-CAS, β-cyanoalanine synthase; H_2S, hydrogen sulfide; GSH, glutathione; Met, methionine. (Source: Babilonia and Siritunga, 2012)

The reason why cysteine is biosynthesized in different cell compartments is still unknown and different suggestions exist about cysteine compartmentalization. Haas *et al*. (2008) confirmed through Arabidopsis knockout mutants that cysteine must be exchangeable between cytosol and organelles. Minimally, cysteine should be synthesized in mitochondria to protect the respiratory chain due to the need for a steady supply of reduced sulfur from cysteine for numerous pathways in mitochondria (reviewed in Droux, 2004), including cyanide detoxification. Some of the enzymes involved in the assimilatory sulfate pathway are encoded by small nuclear gene families and the proteins are transported to different cell compartments (Leustek *et al*., 2000). However, sulfite reductase protein, which is encoded by a single copy gene, is localized only in chloroplast in *A. thaliana* (Bork *et al*., 1998). Therefore, the complete cysteine biosynthesis pathway is located only in the chloroplast in higher plants (Figure 7b). Heeg *et al*. (2008) investigated the impact of CS in different cell compartments by using *A. thaliana* mutants and concluded that sulfide, in addition of cysteine, must be also exchanged between organelles and cytosol. Furthermore, the organelles can in fact compensate for cysteine deficiency.

Table 2. Proteins involved in cysteine biosynthesis in plants

Protein function	Protein location	Protein Name	Gene location	References
Sulfate Transporter	Chloroplast	SULTR1;1	At4g08620	Takahashi et al., 2000
	Plasma membrane	SULTR1;2	At1g78000	Shibagaki et al., 2002
	-	SULTR1;3	At1g22150	Leustek, 2002
	Plasma membrane	SULTR2;1	At5g10180	Takahashi et al., 1997
	Plasma membrane	SULTR2;2	At1g77990	Takahashi et al., 1996
	-	SULTR3;1	At3g51895	Takahashi et al., 2000
	Plasma membrane	SULTR3;2	At4g02700	Takahashi et al., 2000
	Plasma membrane	SULTR3;3	At1g23090	Takahashi et al., 2000
	Plasmodesma	SULTR3;4	At3g15990	Yamaguchi et al., 1997
	Membrane	SULTR3;5	At5g19600	Leustek, 2002
	Membrane	SULTR4;1	At5g13550	Takahashi et al., 2000
	Membrane	SULTR4;2	At3g12520	Salanoubat et al., 2000
	Vacuole	SULTR5;1	At1g80310	Gong et al., 2001
	Vacuole	SULTR5;2	At2g25680	Carter et al., 2004

Protein function	Protein location	Protein Name	Gene location	References
ATP sulfurylase (ATPS)	Chloroplast	APS1	At3g22890	Hatzfeld et al., 2000b
	Chloroplast, Cytosol	APS2	At1g19920	Logan et al., 1996
	Chloroplast	APS3	At4g14680	Hatzfeld et al., 2000b
	Chloroplast	APS4	At5g43780	Heazlewood et al., 2004
APS kinase (APSK)	Chloroplast	APK1	At2g14750	Lee and Leustek, 1998
	-	APK2	At4g39940	-
	-	APK3	At3g03900	-
	-	APK4	At5g67520	-
APS reductase (APSR)	Chloroplast	APR1	At4g04610	Setya et al., 1996
	Chloroplast	APR2	At1g62180	Weber et al., 2000
	Chloroplast	APR3	At4g21990	Gutierrez-Marcos et al., 1996
Sulfite reductase (SiR)	Chloroplast	SIR	At5g04590	Yonekura-Sakakibara et al., 1998
Serine acetyl-transferase (SAT)	Chloro, Cytosol	SAT1	At1g55920	Dominguez-Solis et al., 2008
	Cytosol	SAT2	At2g17640	Wirtz and Droux, 2005
	Mitochondria	SAT3	At3g13110	Roberts and Wray, 1996
	Cytosol	SAT4	At4g35640	Wirtz and Droux, 2005
	Cytosol	SAT5	At5g56760	Howarth et al., 1997
Cysteine synthase (CS)	Cytosol	OASA1	At4g14880	Jost et al., 2000
	Plastids	OASB	At2g43750	Jost et al., 2000
	Mitochondria	OASC	At3g59760	Jost et al., 2000
	Mitochondria	ATCYSC1	At3g61440	Hatzfeld et al., 2000a
	Cytosol	ATCYSD1	At3g04940	Hatzfeld et al., 2000a
	Cytosol	ATCYSD2	At5g28020	Yamaguchi et al., 2000
	Chloroplast	CS26	At3g03630	Nikiforova et al., 2003
	Cytosol*	OAS-TL, putative	At3g22460	Jost et al., 2000
	Cytosol*	OAS-TL, putative	At5g28030	-

(Source: Babilonia and Siritunga, 2012)

The cysteine level in plant cells are not exactly known, but it has been shown to be below 10 µmol/L (Giovanelli *et al*., 1985). The reason for the low cysteine pool in cells may be due to their function as a substrate for several pathways and enzymes (Burandt *et al*., 2002). For example, cystathionine γ-synthase is the initial enzyme in methionine biosynthesis in higher plants and has a K_m for cysteine of 180 µmol/L in spinach and 460 µmol/L in *Arabidopsis* (Ravanel *et al*., 1995; 1998). γ-glutamylcysteine synthetase catalyze the first step in GSH biosynthesis, which have a K_m of 190µmol/L for cysteine (Hell and Bergmann, 1990).

1.6 Genetic improvement of Cassava

1.6.1 Overexpression of β-Cyanoalanine synthase

Despite the cyanide detoxification pathway that is present in cassava, it is not enough for effective cyanide removal in bitter varieties. Naturally, linamarin is converted to cyanide, but when tissue rupture occurs, much greater amounts of cyanide are formed because of the direct interaction of linamarin and the enzymes involved in cyanogenesis. Siritunga and Sayre (2004) explored two strategies to reduce cyanide in cassava: inhibition of the expression of cytochrome P450s (CYP79D1 and CYP79D2) and overexpression of HNL in roots. However, inhibition of CYP79D1/D2 resulted in plants unable to grow on soil without ammonium because the plants need linamarin for amino acid synthesis and incorporation in other metabolic pathways. Overexpression of HNL in roots accelerates cyanogenesis and cyanide volatilization. Other strategies should be explored to reduce cyanide in cassava roots, such as the overexpression of β-CAS, the key enzyme in the cyanide detoxification pathway. That way the cyanide produced can be immediately incorporated into amino acid synthesis.

1.6.2 Overexpression of cysteine synthase

Overexpression of CS can also reduce cyanide because of the similarities between β-CAS and CS, but at the same time higher cysteine levels can be achieved. On a high cysteine diet the probability of having cyanide related disorders is lower. Furthermore, plants with high cysteine content can be classified as high nutritional value plants. In humans, cysteine can be synthesized from the essential amino acid methionine by trans-sulfurylation in the cytosol (Haas *et al*., 2008). Methionine is first

converted to S-adenosyl methionine to produce homocysteine, which reacts with serine to yield cysteine. Contrary to humans, plants synthesize methionine from cysteine. According to the National Academy of Science, 25 mg of cysteine should be consumed for every gram, or minimally for every 400-600 mg, of protein daily. Cysteine and methionine are found in many high protein foods such as animal protein, fish, chicken, turkey, eggs, granola, garlic, onions, yogurt, dark leafy vegetables, corn, and oat flakes. For this reason, people with diets strongly dependent on cassava, such as in Africa, are more vulnerable to cysteine deficiencies due to the very low levels of protein in this cassava. The same deficiency may be seen in vegetarians as plants, overall, are not a high source of sulfur-containing amino acid. Parcell (2002) reported that further analysis of sulfur in the diets is necessary for vegan athletes, children and patients with HIV, due to the increased risk of sulfur-containing amino acid deficiencies in these groups of individuals. Cysteine deficiency symptoms include skin lesions, muscle loss, liver damage, loss of pigmentation in hair, slowed growth in children, decreased levels of proteins in serum, and impaired function of the immune system (Hunter and Grimble, 1997; Dröge, 1999). Cysteine deficiency can also result in reduced production of essential molecules such as GSH, coenzyme A, taurine and inorganic sulfur (Stipanuk *et al*., 2006).

In humans, GSH is synthesized from cysteine and is needed in many critical cell processes such as maintenance and regulation of the thiol-redox status of the cell (reviewed in Ballatori *et al*., 2009), detoxification of xenobiotics and carcinogens, biosynthesis of DNA, proteins and leukotrienes (Ristoff and Larson, 2007), and maintaining hepatic function by neutralization of free radicals. GSH also play an

important role in cell death, influencing caspase activity, transcription factor activation, ceramide production, among others (Ballatori *et al*., 2009). GSH deficiency can increase the susceptibility to oxidative stress, thus resulting in a number of human diseases which can be characterized by hemolytic anemia, neurological symptoms, spinocerebellar degeneration, myopathy and aminoaciduria (Ristoff and Larson, 2007; Ballatori *et al*., 2009). In addition, GSH is associated with age-related pathologies (Samiec *et al*., 1998) and cancer treatment (Estrela *et al*., 2006).

Cysteine and methionine are important in the integration of sulfur in human diet. Sulfur constitutes 0.25% of human body weight (very similar to potassium), being the third most abundant mineral in the body and the sixth most abundant macro-mineral in breast milk (reviewed in Parcell, 2002). It is an essential element because of its incorporation in proteins and other bio-molecules (Atmaca, 2004). Sulfur is also important in many enzymatic reactions, as well as in the formation of collagen and keratin. Sulfur is involved in cellular respiration and is present in insulin and the anticoagulant heparin. In fact, sulfur containing compounds can be found in all body cell types and are indispensable for life. Unlike humans, plants are able to produce sulfur-containing amino acids by using inorganic sulfur, thus plants can be an important source of sulfur for humans and animals (Atmaca, 2004).

1.7 Previous work

Recently, Marrero (2010) isolated two *Bsas* genes from cassava, *MANes;BsasA* and *MANes;BsasB*. Based on phylogenetic analysis, and by comparing them with other *Bsas* genes from Arabidopsis, spinach and potato, the genes were predicted to be mitochondrial. The genes were expressed in *Escherichia coli* NK3 mutant strain, lacking

the activity of any of the *Bsas* proteins. *MANes;BsasB* showed the highest β-CAS activity (1.6333 mmol H_2S min^{-1} mg^{-1} protein), but *MANes;BsasA* also showed a high β-CAS activity (1.2256 mmol H_2S min^{-1} mg^{-1} protein) (Figure 8). In fact, both enzymes presented a higher activity than the only mitochondrial β-CAS in *Arabidopsis* (*Bsas3;1*), that showed 1.0233 mmol H_2S min^{-1} mg^{-1} protein. *MANes;BsasA* showed the highest CS activity (0.1105 mmol H_2S min^{-1} mg^{-1} protein), while CS activity of *MANes;BsasB* was only 0.0227 mmol H_2S min^{-1} mg^{-1} protein. The CAS/CS activity ratio was 11 for *MANes;BsasA* and 72 for *MANes;BsasB*. The CS/CAS ratio was 0.090 for *MANes;BsasA* and 0.014 for *MANes;BsasB* (Figure 8). Results suggest that *MANes;BsasB* might have a main role in cyanide detoxification, encoding a true β-CAS enzyme, while *MANes;BsasA* has a main role in cysteine biosynthesis. Both genes have a similar sequence, but are not identical. The size of *MANes;BsasA* is 1497 bp, while *MANes;BsasB* is 1352 bp. Multiple sequence alignments with other members of the *Bsas* family showed the presence of highly conserved regions in the novel cassava proteins (Figure 9). That region is shared between all the members of the *Bsas* family.

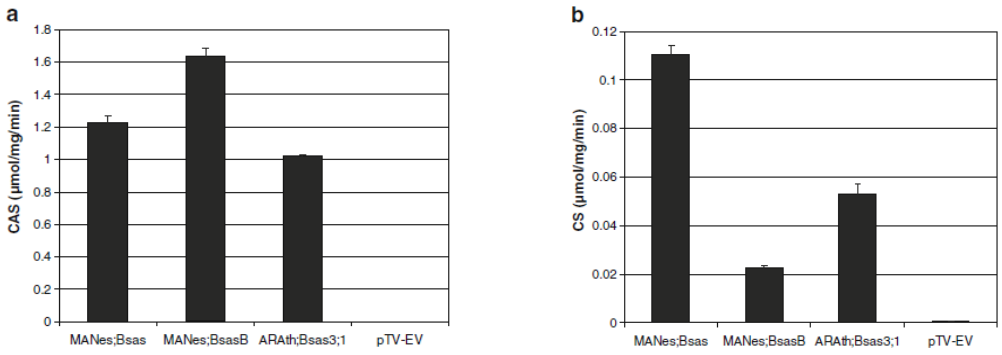

Figure 8. Comparison of CAS to CS activity levels of *MANes;BsasA* and *MANes;BsasB*. *ARAth;Bsas3;1* and pTV118N empty vector (pTV-EV) were used as positive and negative controls, respectively. A) CAS activity levels were measured in µmol sulfide/min/mg protein. B) CS activity levels were measured in µmol cysteine/min/mg protein. (Source: Marrero *et al*., 2010)

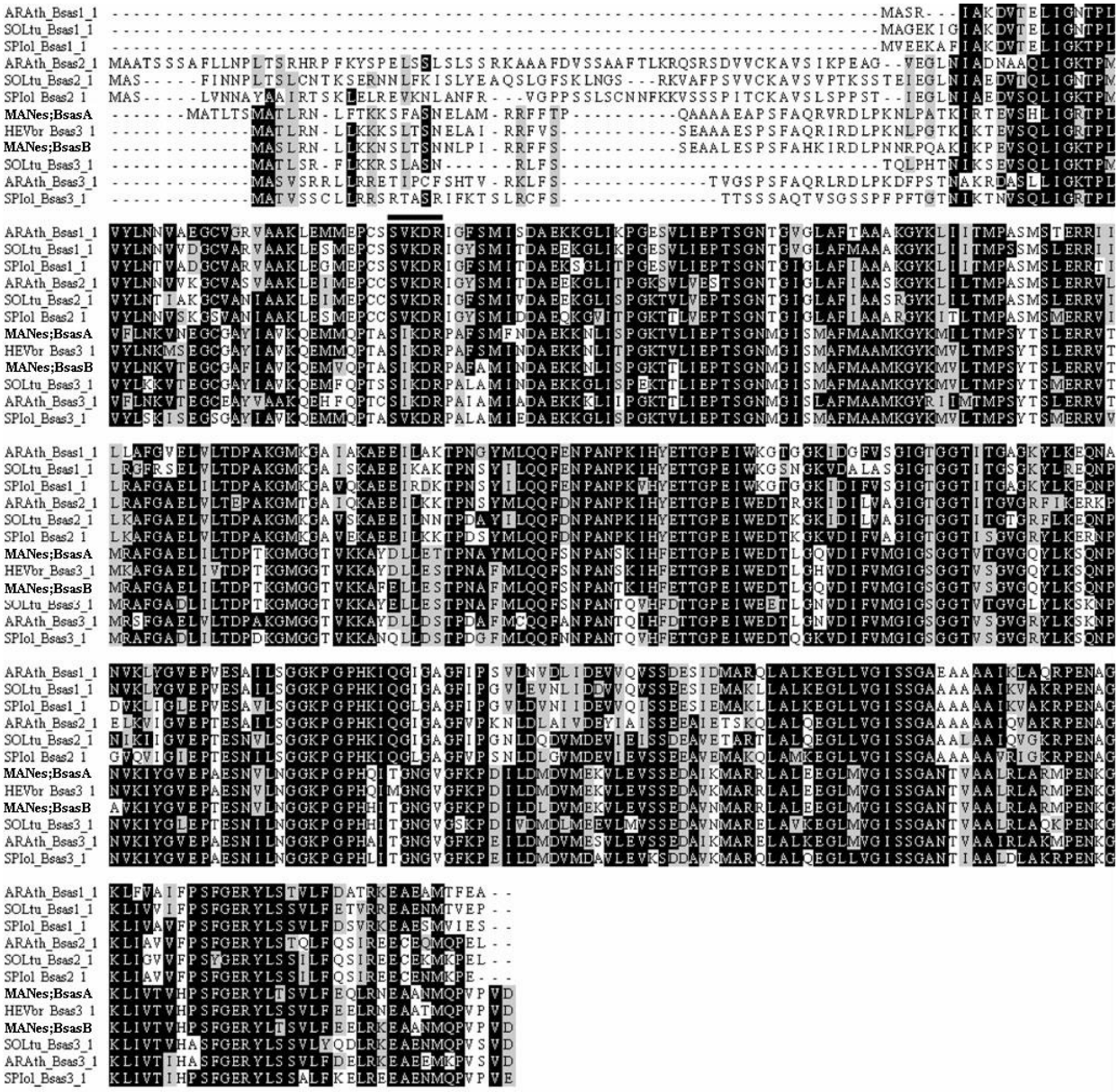

Figure 9. Protein sequence alignments of *MANes;BsasA* and *MANes;BsasB* with other members of the *Bsas* family. (Source: Marrero *et al*., 2010)

1.8 Arabidopsis thaliana

1.8.1 The biology of Arabidopsis

Arabidopsis thaliana is a member of the Brassicaceae (mustard) family and has been accepted as a model plant system for molecular genetic studies attempting to

identify plant genes and determine their functions. Arabidopsis grows as a small ground hugging rosette of about 2-5 cm diameter (leaves may vary from 5 to more than 30 depending on the growth conditions and plant genotype), from which several flowering stem are produced, and they can extend until a height of about 70 cm. The flowers can be 3 mm long and produce four petals, six stamens and a single ovary consisting of two fused carpels (Wilson, 2000). Arabidopsis is self-fertile (autogamous) and can produce hundreds of pods or siliques on fertilization with each pod containing about 50 seeds. Arabidopsis doesn't have any nectarines to attract insects for pollination, but naturally outcrossing by insect can occur at low frequencies (Wilson, 2000).

1.8.2 Characteristics that make Arabidopsis suitable for laboratory experiments

Arabidopsis has several characteristics that facilitate the rapid genetic analysis. For example, it has a relatively small genome (only 15 times that of *Escherichia coli* and eight times that of *Saccharomyces cerevisiae*) when comparing it to other angiosperms. The genome range from 100-120 $X10^6$ base pairs and the reason for its small genome size is due to the presence of little repetitive DNA, something very important when doing mutagenesis or map-based cloning strategies (Wilson, 2000). Furthermore, the plant has a short life cycle which can be completed in six to eight weeks and the growth of the plant is non-seasonal, leading to the possibility of producing several generations in one year. In addition, its genome have been completed sequenced (Sanderfoot and Raikhel, 2001) and is very manageable inside the laboratory because it is a small plant. It is diploid which offers an advantage over other experimental model plants that are polyploid. In addition, the plant is easily transformed by *Agrobacterium tumefaciens* and

thousands of seeds can be produced from one single plant, which is perfect for mutagenesis experiments. As mentioned before Arabidopsis is self-fertile, which is also beneficial to maintain homozygous lines.

Lastly, there are many Arabidopsis genetic resources centers with wild type populations and many mutants generated by mutagenesis. The major diversity of Arabidopsis is maintained in two resource centers: Nottingham Arabidopsis Stock Centre (NASC) in UK, and Arabidopsis Biological Resource Centre (ABRC) at the Ohio State University in Ohio (USA). Over 50,000 accessions are available with majority of the lines available as seeds.

1.8.3 Arabidopsis *Bsas*

In the genome of *A. thaliana* nine putative *Bsas* genes have been identified and until now six of them have been well characterized with respect to their enzymatic properties and sub cellular localization (Watanabe *et al*., 2008). The nine genes of this family in *A. thaliana* are *Bsas1;1, Bsas1;2, Bsas2;1, Bsas2;2, Bsas3;1, Bsas4;1, Bsas4;2, Bsas4;3* and *Bsas5;1*. The isoforms *Bsas4;1* and *Bsas4;2*, which are localized in the cytosol, have very low CS or CAS activity (Hatzfeld *et al*., 2000). The isoforms *Bsas1;2, Bsas4;3* and *Bsas5;1* are much less characterized (Watanabe *et al*., 2008). Of the remaining four isoforms, three of them are CSs, which are cytosolic, plastidic and mitochondrial, and the other one (*Bsas3;1*) is the only β-CAS, being located in the mitochondria (Hell *et al*., 1994; Wirtz *et al*., 2004; Yamaguchi *et al*., 2000). For example, Watanabe *et al*. (2008) found that mitochondrial *Bsas3;1* played a major role as CAS based on enzyme activity and judging by the substrate specificity of the protein. Comparatively, they found that cytosolic *Bsas1;1*, plastidic *Bsas2;1*, and mitochondrial

Bsas2;2 are the genuine CSs in Arabidopsis based on enzyme activities. In the study they found that *Bsas1;1* is primarily responsible for CS activity in leaf and root, thus is the true CS in Arabidopsis. This conclusion was reached by using Arabidopsis mutants, lacking one of the *Bsas* genes, and looking for the decrease of CS activity or CAS activity on leaf and roots. In another study, Hatzfeld *et al.* (2000) found that *Bsas3;1* proteins in Arabidopsis are β-CAS due to its higher affinity for cyanide.

CS activity in Arabidopsis is increased during sulfur, nitrogen and carbon starvation conditions and during abiotic stresses such as salt and heavy metal exposure (Takahashi and Saito, 1996; Dominguez-Solís *et al*., 2001). It has been shown that over-expression of different CS isoforms from different plant species in transgenic tobacco and Arabidopsis plants improves heavy metal tolerance and provides protection against oxidative stress (Dominguez-Solís *et al*., 2001; Youssefian *et al*., 2001; Noji *et al*., 2001). Furthermore, CS over-expression increases cysteine biosynthesis under sulfur stress (Saito *et al*., 1994). Besides the fact that CS exist in different isoforms located in different cell compartments, recent studies indicate that in higher plants the main CS is the cytosolic (Haas *et al*., 2008). In addition, sub-cellular fractionation studies in leaves of *Datura innoxia* (Kuske *et al*., 1996), spinach (Lunn *et al*., 1990) and pea (Droux, 2003) indicate that the major part of CS activity is in the cytoplasm. In fact, cytosolic and plastidic CS contributes to 95% of CS activity in Arabidopsis leaves, while mitochondrial CS contributes to only 5% (Figure 7b) (Haas *et al*., 2008).

Table 3. Most important *Bsas* genes in Arabidopsis and the mutant ID

Isoform	Subcellular location	In Vitro Activity	Mutant ID
Bsas1;1	Cytosol	CSase	SALK_072213
Bsas2;1	Plastids	CSase	SALK_021183
Bsas2;2	Mitochondria	CSase	SALK_000860
Bsas3;1	Mitochondria	CASase	SALK_022479

The predominant CS is *Bsas1;1* and the true CASase is *Bsas3;1*. The subcellular locations were predicted by TargetP and WoLF PSORT (Watanabe et al., 2008). The mutants can be obtained from Ohio State University.

1.9 T-DNA mutation

T-DNA can cause gene aberrations through insertion, so it acts as an insertional mutagen. Those types of mutations can lead to the identification genes involved in a specific biological process since the genes get “molecularly tagged” by the presence of the T-DNA. The most common method of T-DNA transformation is using *Agrobacterium tumefaciens*. This type of tagging is based on a unique DNA transfer system of *A. tumefaciens*, a soil plant pathogen that naturally infect mostly dicotyledonous plants causing grown galls tumors (de la Riva *et al*., 1998). Agrobacterium is capable of transferring a segment of their Tumor inducing (Ti) plasmid, known as the T-DNA. It is flanked by 25 base pair repetitions and is stably integrated into the plant nuclear genome. Genes inside the T-DNA region are expressed in the plant and encode functions for the synthesis of plant growth factors and specific metabolites like opines, which can be used as a source of carbon for the *Agrobacterium* (Zupan and Zambryski, 1995). Deletion of oncogenes located between the 25 bp boundaries of the T-DNA does not affect the process of T-DNA transfer because it is regulated by other virulence genes in the plasmid but outside the T-DNA region (de la Riva *et al*., 1998; Gelvin,

2000; Zupan and Zambryski, 1995). Therefore, any foreign DNA in the T-DNA region can be transferred into the plant genome.

The whole process can be divided in to four steps, where the first step is the bacterial colonization which occurs when the Agrobacterium is attached to the plant cell surface. The second step is the induction of bacterial virulence system which occurs when trans-membrane receptors recognize phenolic compounds released from wound sites of the plant (de la Riva *et al*., 1998). In Arabidopsis it has been found that Agrobacterium can infect the floral structures or ovules and no wounds are required. Once the signal molecules are detected then the third step occurs, which is the activation of virulence genes VirD1 and VirD2 on the Ti plasmid. Those genes are the responsible of cleaving the T-DNA region on the 25 bp repeats, creating a single stranded DNA that represent the copy of the bottom T-DNA strand (de la Riva *et al*., 1998). The next step is the T-DNA transfer, which is mediated by the virulence gene VirE. The DNA segment has to be translocated to the plant nucleus passing through three membranes, the plant cell wall and cellular spaces. For this, virulence proteins VirB serve as protein channels (Dang and Christie, 1997) and protein complexes VirE surround the T-DNA and prevents the attack of nucleases thus facilitating the transfer (Zupan and Zambryski, 1995; de la Riva *et al*., 1998). When the complex crosses the plant nuclear membrane the last step occurs, which is the integration of the T-DNA into the plant genome. This is carried on by illegitimate recombination, where the 5'-end of the T-DNA joins a nick in the plant DNA, forming a gap, and then the 3'-end pairs with another region of the genomic DNA close-by (Zupan and Zambryski, 1995). Plant repair enzymes then join the 3'-end to plant DNA, repair the gaps and DNA synthesis is

achieved using the T-strand as template (Zupan and Zambryski, 1995). One advantage is that the T-DNA inserts are stable, meaning that once integrated, it remains at their original position. With this method and by self-fertilization large numbers of transformed seeds can be obtained with stable T-DNA inserts.

1.10 Forward and Reverse genetics

The function of a gene or the identity of genes responsible for a trait can be determined by two approaches: forward genetics and reverse genetics. Forward genetic refers to the classical identification of genes, that are responsible for a particular phenotype in an organism, through perturbation of their function using mutagenesis and then cloning and sequencing the gene (Wilson, 2000). In this type of genetics ethyl methane sulfonate (EMS) mutagenesis is preferred to cause point mutations, where a lot of hits per genome are obtained. Therefore the number of individuals required to represent saturation mutagenesis can be less than the required for T-DNA insertions. Once a mutant has been identified and mapped the next step is the cloning of sequences linked to the particular phenotype (Peters *et al*., 2003).

Reverse genetics is a method used to ascertain the function of a gene whose function is previously unknown (Sessions *et al*., 2002). For this, T-DNA tagged populations are essential for the screening of mutants based on sequence analysis of potential coding regions. With PCR analysis the plant lines containing the insert positioned in the gene of interest can be identified and analyzed for phenotypic aberrations. For this type of genetics, stock centers distribute DNA and seeds from available T-DNA tagged populations. The difference is that in forward genetics the genetic basis of a phenotype or trait is to be found, while in reverse genetic the

phenotypes as a result of particular genes are studied (Wilson, 2000). Using reverse genetics we can investigate the function of all genes in a family, something that cannot be achieved easily with forward genetics. Also, it is possible to study the function of a gene that is involved in the process of interest on other organism, but for which no forward genetic mutants have been identified. For example, many genes have not been yet mutated in organisms, but reverse genetics allow their study. With this approach every single gene from genome sequences can be studied.

1.11 Gateway Cloning Technology

The Gateway® Technology is a cloning method based on the site-specific recombination properties of the bacteriophage lambda (Landy, 1989). This technique provides a rapid and efficient way to move your gene of interest into multiple vector systems by a single reaction. The Gateway® Technology can be divided in 2 steps:
Cloning of your PCR product into a pENTR™TOPO® vector to create an entry clone, and a LR recombination reaction between the entry clone and a Gateway® destination vector to generate the expression construct (Invitrogen user manual). In the first step, the PCR product is made by adding four bases to the forward primer (CACC). A complementary overhang in the entry vector (GTGG) invades the 5' end of the PCR product and stabilizes it in the correct orientation with the help of Topoisomerase I in the vector (Figure 10). The PCR product is then inserted and ligated in the correct orientation, creating an entry clone.

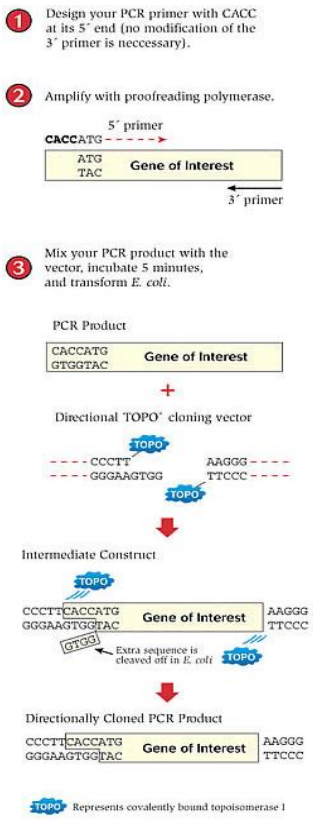

Figure 10. Directional TOPO® cloning. (Source: tools.lifetechnologies.com)

In the second step, the Gateway® technology use the signal DNA sequence and enzymatic machinery of the bacteriophage lambda recombination system to facilitate the transfer of the desire DNA (Hartley *et al*., 2000). The gene of interest in the entry clone is flanked by the signal sequences attL1 and attL2, which are the sites where recombination occurs (Figure 11). Similarly, the destination vector contains the signal sequences attR1 and attR2. The recombination reaction of an attL substrate with an attR substrate is called the LR reaction, which is catalyzed by LR Clonase™ enzyme mix (Invitrogen). The LR reaction is the *in vitro* version of the phage lambda excision reaction, where the transfer of the target DNA fragment in the entry clone to a destination vector generates an expression clone (Sasaki *et al*., 2004). The advantage of this technique is that once you have your gene of interest in an entry vector you can

move it to many different Gateway expression vectors. In addition, the whole process can only take one week of work, which is a huge advantage over tradition cloning with restriction enzymes.

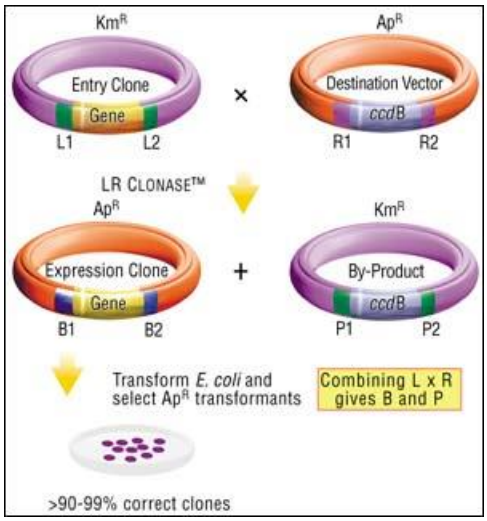

Figure 11. Gateway Cloning Technology overview. The gene of interest in the entry clone is flanked by the signal sequences attL1 and attL2, which are the sites where recombination occurs (Figure 11). Similarly, the destination vector contains the signal sequences attR1 and attR2. The recombination reaction of an attL substrate with an attR substrate is called the LR reaction, which is catalyzed by LR Clonase™ enzyme mix. (Source: www.bio.davidson.edu)

When the LR reaction occurs, a gene (ccdB) that is present in the destination vector is replaced by the DNA of interest (Figure 11). The ccdB gene is a lethal gene that interferes with the DNA gyrase. The idea of having this gene is due to the presence of both the plasmid with the integration of the insert DNA and without it after the recombination event. When this happens, only a fraction of the transformed bacterial cells acquires a plasmid recombined with the DNA of interest, meaning that it is necessary to separate the bacteria carrying the insert DNA from the non-recombinants ones (Bernard *et al*., 1996). In the gateway cloning reaction, the cells that acquire the destination vector without the insert will die because of the lethal gene ccdB. However,

if the recombination reaction was successful, the bacteria will receive the destination vector containing the insert and not the ccdB gene.

1.12 Objectives

Although *MANes;BasA* and *MANes;BsasB* were characterized in bacteria, their characterization in plants has not be accomplished. As a general objective, this study aim to characterize β-cyanoalanine synthase and cysteine synthase from cassava by using mutants of the model plant *Arabidopsis thaliana.* The mutants used have target genes in the cyanide detoxification pathway silenced thus enabling the functional complementation with the novel cassava genes. The finding in this study will decide if cassava plants will be transformed with the novel genes for genetic improvement.

MATERIALS AND METHODS

2 Materials and methods

2.1 MANes;*BsasA* and *MANes;BsasB* isolation and cloning

2.1.1 β-CAS and CS cloning into pKYLX

MANes;BsasA and *MANes;BsasB* genes were amplified from cDNA of an in-vitro cassava variety, Mcol 2215, that was synthesized by Marrero (2010). Primers were designed to amplify engineered truncated versions of the genes, lacking the signal peptide (Marrero *et al*., 2010). The restriction sites in the primers allowed the cloning of the genes into pKYLX71:35S^2 binary vector (Schardl *et al*., 1987) under the control of the double cauliflower mosaic virus (CaMV) 35S constitutive promoter. Ligation was carried on with T4 DNA ligase and incubation at 4°C overnight.

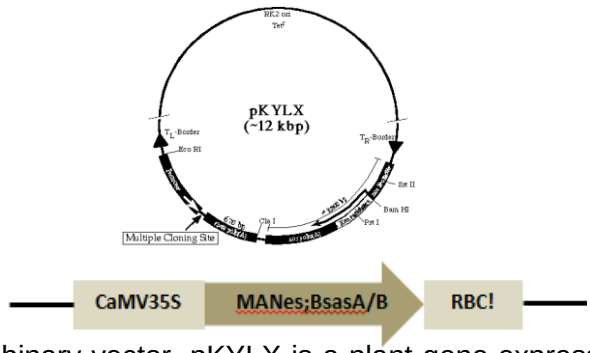

Figure 12. pKYLX71:35S^2 binary vector. pKYLX is a plant gene expression vector for use with Agrobacterium-mediated gene transfer in plants. Genes are under the control of the double cauliflower mosaic virus (CaMV) 35S constitutive promoter with a Rubisco promoter (RBC!). The vector contains tetracycline resistance in bacterium and kanamycin resistance in plants. (Source: Schardl *et al*., 1987)

2.1.2 β-CAS and CS cloning into pB2GW7

The engineered truncated versions of *MANes;BsasA* and *MANes;BsasB* were cloned into pB2GW7 expression vector according to the Gateway® Cloning

Technology. The first step involved the preparation of an entry clone, where the addition of four nucleotides (CACC) in the 5' end of the forward primer allowed the insertion of the genes into the entry vector pENTR/SD (Invitrogen). The ligation reaction of the PCR products into pENTR/SD was prepared as the following: 1 µL of fresh PCR product + 0.5 µL of salt solution + 0.5 µL of sterile water + 1µL of TOPO® vector (pENTR/SD). The ligation was carried on at room temperature for 15 min followed by transformation of *E. coli*.

The cloning of the genes into pB2GW7 was carried out according to the Gateway® LR Clonase™ II Enzyme Mix kit (Invitrogen). The following components were added to a microcentrifuge tube at room temperature and mixed: 50-150 ng of entry clone + 1 µL of 150 ng/µL of pB2GW7 + TE buffer, pH 8.0, to a final volume of 8 µL. To each reaction, 2 µL of LR Clonase™ II enzyme mix were added, followed by incubation of the samples for 1 hour at 25ºC. During this step a recombination reaction occurs between the entry clone and the destination vector resulting in the expression clone. Finally, 1µL of Proteinase K solution was added to terminate the reaction and the samples where incubated at 37 ºC for 10 min. The final product was then inserted into *E. coli* for the screening.

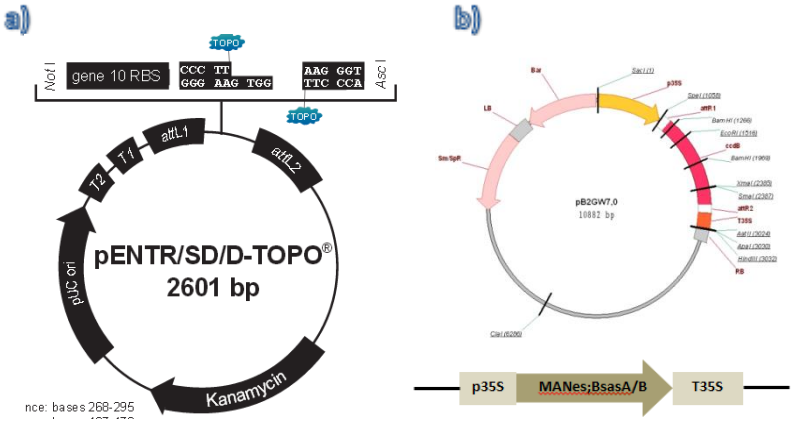

Figure 13. Gateway Cloning Vectors. A) Entry clone. The Sticky end between the AttL regions allows the insertion of the PCR product containing a CACC sequence in the 5' end. The Topoisomerases in the vector ligate the PCR product into the entry clone. The vector confers resistance to kanamycin in bacteria. B) Destination/expression vector. The AttR regions are the sites where recombination with the attL regions of the entry clone occurs. This way the gene of interest can be transferred from the entry clone to the destination vector in a recombination event. The genes are under the control of the cauliflower mosaic virus (CaMV) 35S constitutive promoter. This vector confers resistance to spectinomycin in bacteria and Basta (glufosinate) in plants. (Source: www.lifetechnologies.com)

2.1.3 Transformation of chemically competent *E. coli*

Expression vectors with either *MANes;BsasA* or *MANes;BsasB* were transformed into One Shot® TOP10 Chemically Competent *E. Coli* (Invitrogen) as described in the manual. The cells were thawed on ice and 50–100 ng of DNA was added to each vial. The mixture was incubated on ice for 30 minutes followed by heat-shock for 30 sec at 42ºC. The vials were placed on ice for 2 min after heat-shock and then 250 µL of pre-warmed S.O.C. medium were added. Vials were incubated at 37ºC for 1 hour at 225 rpm and 50-150 µL of culture were spread on pre-warmed selective plates (Kanamycin for pENTR/SD; Tetracycline for pKYLX; Spectinomycin for pB2GW7) and incubated overnight at 37ºC. Colony PCR was carried on for the confirmation of the constructs.

2.1.4 Agrobacterium tumefaciens transformation

Agrobacterium tumefaciens LBA4404 strain ElectroMAX cells (Invitrogen) were transformed by electroporation. About 20 µL of the cells were thawed on ice, mixed by tapping gently and transferred to a microcentrifuge tube containing 1 µL of 100 ng/µL of the plasmid resuspended in 10mM Tris-HCl (pH 7.5) and 1 mM EDTA. The cell/DNA mixture was transferred into a chilled 0.1 cm electroporation cuvette and electroporation was performed at 2.0kV in a MicroPulser electroporator (Bio-Rad, Hercules, CA, USA). One mL of YM medium (0.4 g/L yeast extract, 10 g/L mannitol, 0.1 g/L NaCl, 0.8 mM $MgSO_4 \cdot 7H_2O$ and 2.2 mM $K_2HPO4 \cdot 3H_2O$; pH 7.0) was added to the cuvette immediately after electroporation and the resulting mixture was transferred to a 15 mL snap-cap tube followed by incubation for 3 hours at 30ºC and 225 rpm. Approximately 100 µL of the culture were plated in YM plates containing the corresponding antibiotics. Colony PCR was performed for the confirmation of the construct.

2.1.5 Sequencing

The constructs were confirmed by sending the samples to Nevada Genomic Center (NGC) for sequencing. Primers amplifying from the promoter region to the terminator allowed the sequencing of the genes. The plant selection marker, kanamycin in pKYLX and BASTA® in pBG2W7, were also partially sequenced for their presence.

2.2 Arabidopsis thaliana handling and transformation

2.2.1 Plant material

A. thaliana seeds were obtained from the Arabidopsis Biological Resource Centre (ABRC) in Ohio State University (Columbus, OH). Wild type seeds and 3

different *Bsas* mutants were used: mitochondrial β-CAS mutants, cytosolic CS mutants and mitochondrial CS mutants. Arabidopsis, as previously mentioned, has three predominant CS, which are mitochondrial, plastidic and citosolic. Recently, Watanabe *et al*. (2008) used Arabidopsis mutants for each of the nine *Bsas* genes in the plant to characterize them and they found that the predominant CS in Arabidopsis is the cytosolic one, or *Bsas1;1*. That gene showed the highest CS activity in both roots and leaves and for that reason it will be used as the true CS for this study. The ID numbers for the true CS, β-CAS and mitochondrial CS mutants are SALK_072213, SALK_022479 and SALK_000860, respectively. Mutations are homozygous and the type of mutation is T-DNA insertion. The advantage of this mutation technique is that it is specific and will not produce several random mutations along the genome like other types of mutagenesis.

2.2.2 Seed germination and plant care

Because of the weak growth habit that some of the mutants showed it was necessary to germinate the seeds in tissue culture and once the plants reach the four leave stage they were transplanted to soil (Wilson, 2000). The seed surface was sterilized to kill any bacteria or fungi that can be prejudicial to plant growth. To accomplish this the seeds were suspended in 70% (v/v) ethanol for about five minutes with gentle agitation and then the ethanol was replaced by a 50% (v/v) bleach solution for about 10 minutes. The seeds were washed 3-4 times with sterile water and placed on solid Murashige and Skoog (MS) medium for optimal germination.

Once the plants reached the small rosette stage they were transplanted to Miracle Gro soil (Home Depot), which is a good soil for a wide range of plants grown in

smaller container types. Plants were covered with a plastic cylinder to avoid cross-fertilization between them. Plant growing conditions were 16 hours of photoperiod and a temperature of 24-25°C. After approximately one week, the initial bolts were cut to induce many secondary bolts which is conducive to obtain more seeds after transformation.

2.2.3 Plant transformation: Floral Dip method

The Floral Dip Method used for Arabidopsis transformation has been shown to be successful in transforming the ovules of the plant using Agrobacterium resulting in many transformants from one single plant. Once the floral buds are evident, the Agrobacterium solution is applied to them. Clough and Bent (1998) were investigating the more efficient way to transform the plants. They applied the Agrobacterium solution to un-open flowers, open flowers and siliques (bags with seeds). After several experiments it was found that the highest rate of transformation was obtained when using un-open flowers while the plants least susceptible for transformation are the ones with open flowers and mature siliques (Figure 14).

Figure 14. *Arabidopsis thaliana* flowers

The protocol utilized in this study was according to Clough and Bent (1998), Bechtold *et al*. (1993) and Zhang *et al*. (2006). Healthy Arabidopsis plants were grown under continuous light at a temperature of about 23ºC. First bolts were clipped to induce

the proliferation of many secondary bolts and after approximately one week the plants were ready for the dipping.

Agrobacterium tumefaciens carrying the gene of interest was cultured for 2 days at 28ºC in 10-20 mL of 3 different media to determine which one is the optimal for plant transformation: YM, LB (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, pH 7.5), and YEBS (1 g/L yeast extract, 5 g/L beef extract, 5 g/L sucrose, 5 g/L bacto-peptone, 0.5 g/L magnesium sulphate, pH 7). After 2 days the culture was used to inoculate 250 mL of media and after one more day the cells were centrifuged and resuspended in a 5-10% sucrose solution (Clough and Bent, 1998). The surfactant silwet L-77 was added to a concentration of 0.05% (500µL/L). Silwet L-77 decreases the superficial tension and allows the Agrobacterium to penetrate the plant tissue and reach the ovules. The transformation event was carried on three different ways: The Agrobacterium solution was applied to the flower buds with a pipette, the plants were sprayed with the Agrobacterium solution, and the plants were submerged in the Agrobacterium solution from 5 sec to 1 min. The plants were then covered with a plastic dome to maintain high humidity and placed in the dark overnight (Silwet L-77 is photosensitive). The next day the plants were put back in the incubator with the normal conditions as previously discussed for approximately 5 weeks. Once the plants started drying, further watering was stopped until the dry mature siliques were ready to be harvested.

2.2.4 Screening of the transgenics

Seeds obtained (M1) from first generation plants (M0) were sterilized and plated in MS Agar Medium with kanamycin (50 µg/mL) or basta for selection of the transgenic putatives. Basta was also sprayed over plants germinated in soil. Plants that survived

the treatment were kept in soil for DNA extraction and confirmation of the construct through PCR.

2.2.5 Arabidopsis DNA isolation

One larger leaf of each plant was collected and placed in a microcentrifuge tube containing 150 µL of 2X CTAB buffer (1.4M NaCl, 2% CTAB (cetyl-trimethyl-ammonium bromide), 20 mM EDTA and 100 mM Tris-HCl pH 8). The tissue was ground with a pestle and incubated at 65 ºC for 10 min, followed by the addition of 150 µL of chloroform. Samples were vortexed thoroughly and spun down for 5 min in a microcentrifuge at maximum speed. The upper layer was transferred to a new tube and 150 µL of 2-propanol were added and mixed by vortexing to precipitate the DNA. Samples were centrifuged again for 5 min and the resulting pellet was washed with 300 µL of 70% ethanol. The ethanol was discarded and the pellet was dried at room temperature to then be dissolved in water and incubated at 65ºC for 10 min.

2.3 Polymerase Chain Reaction (PCR)

PCR was performed in 25 uL reaction mixtures containing the following: 50–100 ng of template DNA, 1X *Taq* buffer (10 mM Tris-HCl pH 9.0, 50 mM KCl and 1.5 mM MgCl2), 0.2 mM dNTP mix, 0.2 uM of each primer and 1 U *Taq* DNA polymerase (Amresco® or GoTaq®). For PCR confirmation of the transgenics different concentrations of $MgCl_2$ were used. For the cloning plant and selection of transgenics Platinum ® *Taq* DNA polymerase high fidelity (Invitrogen) was used, with 2 mM of $MgSO_4$ instead of $MgCl_2$. The PCR program for the cloning part and the selection of transgenics using the Platinum® *Taq* was as follows:

Temperature	Time	Cycles
94ºC	2 min	1
94 ºC	30 sec	35
55 ºC	30 sec	
68 ºC	1 min	
68 ºC	5 min	

The PCR program for amplification of the genes from bacteria was as follows:

Temperature	Time	Cycles
94ºC	3 min	1
94 ºC	30 sec	35
55 ºC	1 min	
72 ºC	1 min	
72 ºC	5 min	

In addition, touchdown PCR was performed when working with the putatives transgenics. The program was run as the following:

Temperature	Time	Cycles
94ºC	3 min	1
94ºC	30 sec	20 – each cycle with a 0.5ºC decrease in annealing temp.
65ºC	1 min	
72ºC	1:15 min	
94ºC	30 sec	20
55ºC	1 min	
72ºC	1:15 min	
72ºC	5 min	

2.4 Protein extraction from Arabidopsis plants

Protein extraction was performed according to Harada *et al*. (2001). Leaf tissues were ground with a pestle in a microcentrifuge tube containing 100 µL of 50 mmol/L phosphate buffer (pH 7.5), 5 mmol/L $MgCl_2$, 2 mmol/L DTT, 1 mmol/L EDTA, 0.5 mmol/L phenylmethylsulfonylfluoride (PMSF) and 0.1% Triton X-100. The samples were then centrifuged at 16,000xg for 15 min at 4ºC. The resulting supernatant was used for the enzymatic assays.

2.5 Enzyme activity assays

Cysteine synthase activity was measured according to Maruyama *et al*. (2001) and Lunn *et al*. (1990). The reaction mixture contained 2.5 mM Na_2S, 5 nM O-acetyl-l-serine and 5 mM dithiothreitol (DDT) in 50 mM sodium phosphate buffer pH 7.5. Crude protein extract (10 uL) was added until a final volume of 100 µL. The reaction was terminated by the addition of trichloroacetic acid (TCA) and the cysteine formed was quantified spectrophotometry at 600 nm using the method reported by Gaitonde (1967). The enzymatic activity of β-CAS was determined according to Hasegawa *et al*. (1994) and Elias *et al*. (1997). The reaction mixture contained 25 mM KCN and 5 mM L-cysteine in 50 mM Tris-HCl pH 8.5. The reaction was initiated with the crude protein extract and stopped with ferric chloride and N-dimethyl-phenylenediamine sulfate (DPDS). The amount of methylene blue formation was measured spectrophotometrically at 600 nm. The standard curve in both assays was created with known concentrations of bovine serum albumin.

RESULTS

3 Results

3.1 *MANes;BsasA* and *MANes;BsasB* isolation and cloning

3.1.1 β-CAS and CS cloning into pKYLX

Both *MANes;BsasA* and *MANes;BsasB* were amplified from cassava cDNA with primers containing restriction sites for the cloning into pKYLX binary vector (Figure 15a). In order to differentiate between the two genes through PCR a set of primer that hits both genes but in different positions was used (Figure 15b).

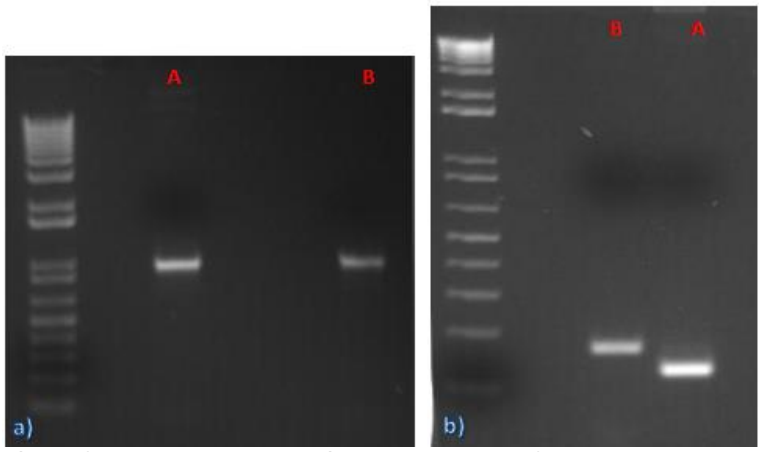

Figure 15. Amplification of *MANes;BsasA* and *MANes;BsasB* from cassava cDNA. A) Primers used contain restriction sites for the cloning into pKYLX. B) Primers used hit both genes in different positions, allowing the differentiation between the two.

After the cloning of the genes into pKYLX vector, transformation of *E. coli* was achieved through heat shock. One colony for each construct (Empty vector, *MANes;BsasA* and *MANes;BsasB*) was analyzed by PCR with 3 different primer sets to prove that it contains the gene of interest in the correct binary vector (figure 16a). The first primer combination amplified from the promoter region to the gene, while the second combination amplified from the gene to the terminator (RBC!). Similarly, the third primer combination amplified from the promoter of the selectable marker

kanamycin (NOS) to kanamycin gene (nptII). Because of the high similarity between *MANes;BsasA* and *MANes;BsasB* the bands appears to be in ther same position.

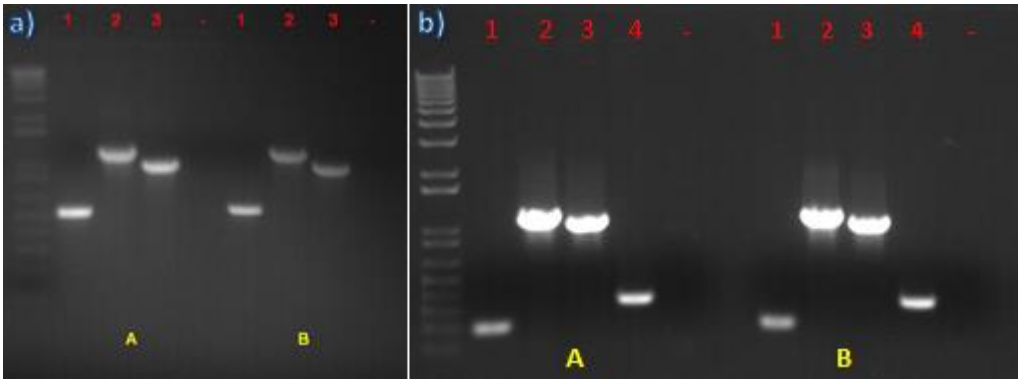

Figure 16. Confirmation of pKYLX construct in bacteria. A) *E. coli* PCR with three primer combinations for the comfirmation of the construct. Both colonies (with *MANes;BsasA* and *MANes;BsasB*) were amplified with the following primeer sets: Promoter to gene, gene to terminator, kanamycin (selectable marker). B) Agrobacterium PCR with the following 4 primer combinations: Promoter to gene, gene to terminator, kanamycin, VirD region. This PCR allows the confirmation of the gene in the correct vector.

3.1.2 Transformation of *Agrobacterium tumefaciens*

Agrobacterium tumefaciens LBA4404 was transformed by electroporation with pKYLX vector containing either *MANes;BsasA* or *MANes;BsasB*. Similarly, 4 different primer combinations allowed the identification of the construct in the binary vector (Figure 16b). The first combination amplified from the promoter to the beginning of the gene, similar to the one used for *E. coli.* Combinations 2 and 3 are the same as described in Figure 16a, while combination #4 amplify the virD region of Agrobacterium. Samples were then sequenced to confirm the constructs (Figure 17).

Figure 17. Part of the sequencing results obtained from Nevada Genomics Center (NGC). A) Sequencing of *MANes;BsasA* compared to the *MANes;BsasA* sequence in NCBI (accession numer EU350583). B) Sequencing of *MANes;BsasB* compared to the *MANES;BsasB* sequence (Marrero *et al*. (2010)).

3.2 β-CAS and CS cloning into pB2GW7

3.2.1 Creating an entry clone: Cloning of the genes into pENTR/SD

Both genes were amplified from cassava cDNA with a modification in the forward primer (CACC in the 5' end) for the entry reaction (Figure 18). No modification in the reverse primer was neccesary.

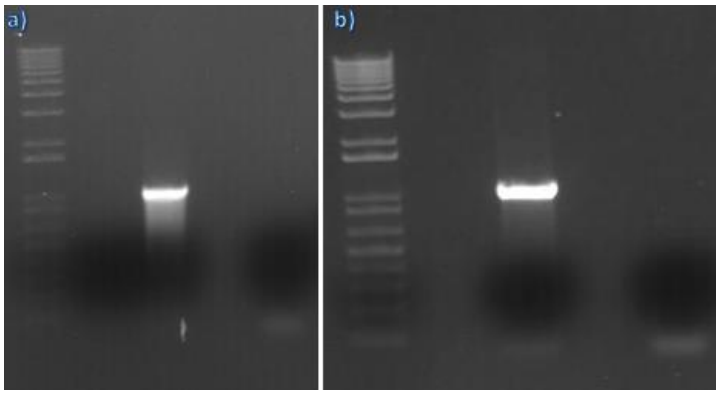

Figure 18. Amplification of cassava *Bsas* genes from cDNA for the entry reaction. A) *MANes;BsasA* amplification with an overhang CACC in the 5' end. B) *MANes;BsasB* amplification with an overhang CACC in the 5' end.

The modified genes were extracted from the gel and used for the entry clone reaction. The final product was inserted into *E. coli* through heat shock and the culture was plated in solid LB media with Kanamycin. Several colonies were obtained (Figure 19c) and 10 where chosen for colony PCR using primers specific for the genes (Figure 19a,b).

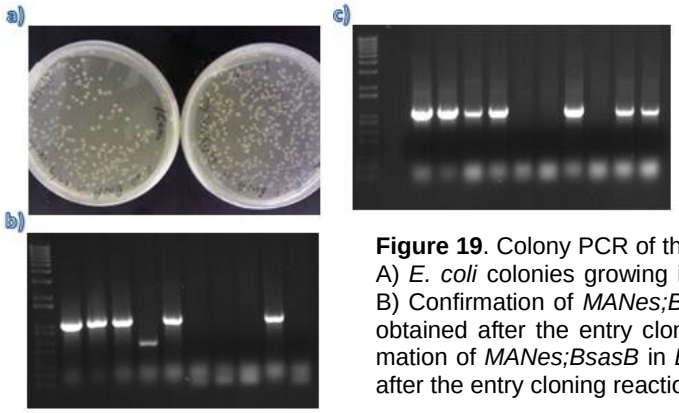

Figure 19. Colony PCR of the entry cloning reaction. A) *E. coli* colonies growing in the selectable media. B) Confirmation of *MANes;BsasA* in *E. coli* colonies obtained after the entry cloning reaction. C) Confirmation of *MANes;BsasB* in *E. coli* colonies obtained after the entry cloning reaction.

After chosing one colony for each construct a plasmid prep was carried on and the products were run in a gel as follows: entry vector with *MANes;BsasA* (EV-A), empty entry vector as a control (EV) and entry vector with *MANes;BsasB* (EV-B). It can be clearly seen that the genes were inserted into the vector (Figure 20), producing an entry clone to be used in the recombination reaction.

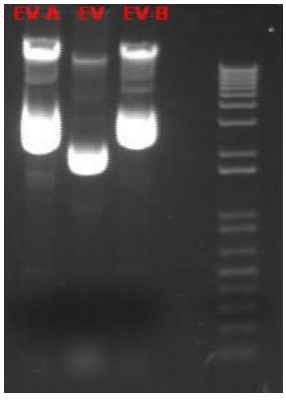

Figure 20. Size of the entry clones. (EV-A, entry vector with *MANes;BsasA*; EV, entry vector; EV-B, entry vector with *MANes;BsasB*)

3.2.2 Cloning of the genes into pB2GW7 by Gateway® Cloning Technology

The LR reaction, or recombination reaction, was carried on with the product of the entry clone reaction. Again, several clonies for each construct were obtained and 5 where chosen for each construct for the colony PCR (Figure 21a). When running the plasmid the difference is not clear between the empty vector and the vector with *MANes;BsasA* (A) or *MANes;BsasB* (B) (Figure 21b). A PCR was conducted using primers that amplify from the promoter region to the terminator region to compare between the three samples. We can see that the genes were introduced in the vector because the amplification resulted in a shorter fragment that in empty vector (Figure 21c). The reason is because pB2GW7 contains the ccdB gene (suicide gene) between the attR1 and arrR2 regions, which is larger than the *Bsas* genes. This gene allows the negative election of the donor vectors in *E. coli* following recombination (Bernard *et al*., 1996). The ccdB protein interferes with *E. coli* DNA gyrase, inhibiting the grow of the common *E. coli* strains. Cells that take up unreacted vectors carrying the ccdB gene will die. This technique allows a high efficient recovery of the desired clones.

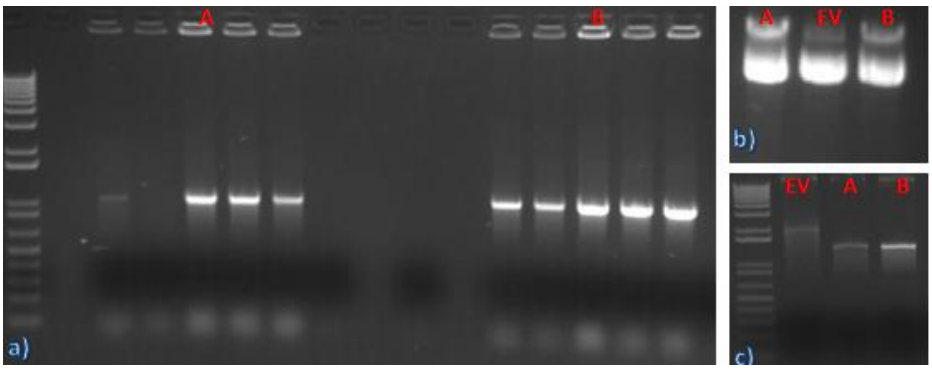

Figure 21. Colony PCR of the LR reaction and confirmation of the constructs. A) Colony PCR. Five colonies for each construct were chosen for the PCR. B) Comparison between the destination vector with and without any gene. C) Comparison of the construct by PCR. Primers that amplify from the promoter to the terminator region allow the identification of *MANes;BsasA/B* in pB2GW7. The larger fragment that resulted in the empty vector is a product of the ccdB gene.

3.2.3 Transformation of *Agrobacterium tumefaciens*

Once the expression clone was created it was inserted in *Agrobacterium tumefaciens* LBA4404 through electroporation and many colonies were obtained (Figure 22). Again, after colony PCR one colony for each construct (empty vector, vector with *MANes;BsasA* and vector with *MANes;BsasB*) was analyzed by PCR with primers from promoter to terminator (Figure 23b) and compared to the samples from *E. coli* (Figure 23a). The Basta gene (Bar) was also amplified from Agrobacterium for the confirmation of pB2GW7 (Figure 23b). Samples were sequenced prior to transformation of Arabidopsis.

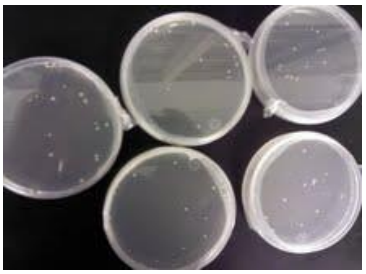

Figure 22. Colonies obtained after the electroporation of *Agrobacterium tumefaciens* LBA4404.

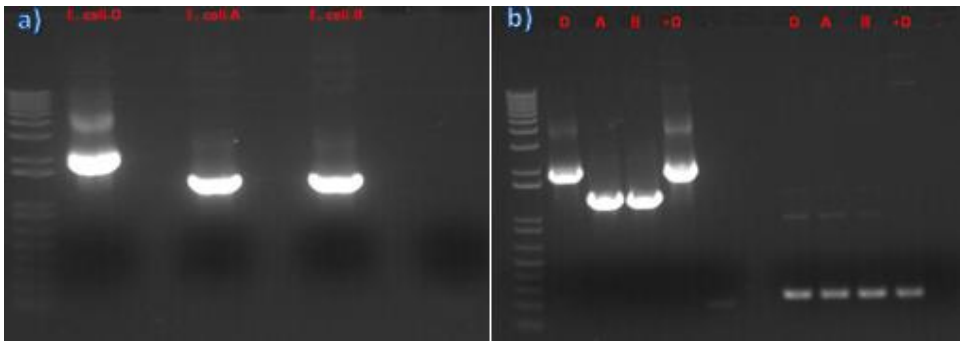

Figure 23. Confirmation of pB2GW7 constructs in bacteria using primers from promoter to terminator. Agrobacterium samples (b) were compared to *E. coli* samples (a). Amplification of Bar (Basta) was also included in part (b).

3.3 Arabidopsis transformation

Arabidopsis transformation was carried on with pKYLX, and after obtaining several false putatives transgenics, the genes were switched to pB2GW7 binary vector. Because of the false putatives that we obtained from the first transformations, different media were used to grow Agrobacterium, as discussed earlier (Table 4).

Table 4. Summary of the transformation events and modifications to the protocol.

		WT	Cyto. CS	Mito. CS	Mito. β-CAS
Vector	pKYLX	X	X	X	X
	pYKLX-*BsasA*	X	X	X	X
	pKYLX-*BsasB*	X	X	X	X
	pB2GW7	X	-	-	-
	pB2GW7-*BsasA*	X	-	-	-
	pB2GW7-*BsasB*	X	-	-	-
***A. tumefaciens* growth media**	YM	X	X	X	X
	LB	X	X	X	X
	YEBS	X	X	X	X
Sucrose %	5%	X	X	X	X
	7.5%	X	X	X	X
	10%	X	X	X	X
# of plants dipped	30-50			X	
	50-100	X	X	-	X
# of dippings per plant	3	X	X	X	X
	2	X	X	X	X
	1	X	X	X	X
Dipping time	<5 sec	X	X	X	X
	5 – 30 sec	X	X	X	X
	1 min	X	-	-	-
Method of application	Spraying	X	X	X	X
	With pipette	X	X	X	X
	Dipping	X	X	X	X
Overnight incubation	yes	x	X	X	X
	no	X	X	-	X
Plants per pot	1	X	X	X	X
	2-5	X	X	X	X
Plant age at dipping	4 weeks	X	X	X	X
	5 weeks	X	X	X	X
	6 weeks	X	X	X	X
Seeds harvested (M1)	<500	-	-	X	-
	500-5,000	-	X	X	X
	>5,000	X	-	-	-
Putative plants obtained	<5	-	-	X	-
	5-10	-	X	-	X
	>10	X	-	-	-

3.3.1 Floral dipping method using pKYLX as the plant expression vector

Plants were germinated in MS media and after approximately 2 weeks they were transferred to soil. The dipping of the plants were carried out after several bolts with close flower were obtained (Figure 24). Some plants were dipped once while others where dipped 2 or 3 times with an interval of about 1 week between each dipping.

Figure 24. Germination and growing of *A. thaliana* for transformation with *A. tumefaciens*.

After the first dipping and incubations overnight it was observed that the plants got very weak and the majority died, probably due to too much humidity and agrobacterium infection (Figure 25). However, a small amount of seeds was harvested and screened in MS media with Kanamycin (50 mg/L) and Carbenicillin (100mg/L). Unfortunately, no putatives transgenics were obtained.

Figure 25. Plants several days after the dipping. The next day after dipping the plants started looking very weak. Eventually, the above parts of the plants, specially the parts that were dipped in the agrobacterium solution, dried and died.

The transformation protocol was then modified and no incubation was carried on after the dipping. Plants were kept in normal light conditions the whole time and they were able to normally grow and produce many seeds (Figure 26a). After harvesting

those seeds, several putatives were obtained for each mutant and WT, with both *MANes;BsasA* and *MANes;BsasB* (Figure 26b). However, the efficiency of the transformations was about 5%, which is too high when compared to what is reported in the literature (about 1%).

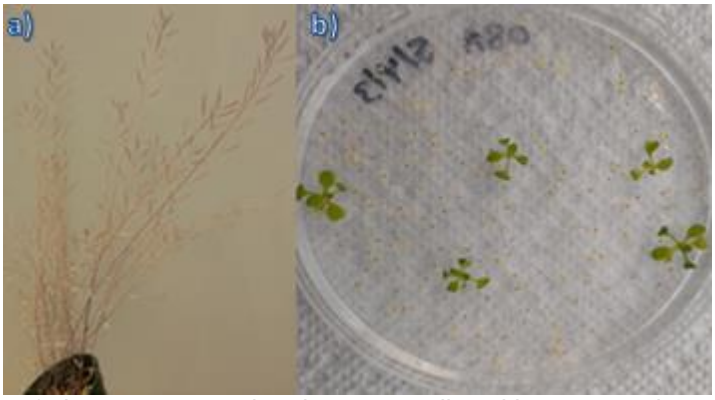

Figure 26. Screening of the M1 generation in MS media with Kanamycin. A) Seeds ready to be harvested. B) Putatives transgenics growing in kanamycin.

3.3.1.2 PCR screening of the putatives transgenics

Despite the fact that good putatives were obtained, no confirmation through PCR was achieved (Figure 27). Amplification of actin was used as a control (Figure 27a), but when trying to amplify *Manes;BsasA/B* no amplification besides the positive was obtained (Figure 27b). Different primers were used for the amplification of the inserts (Figure 27c,d), but no bands were obtained.

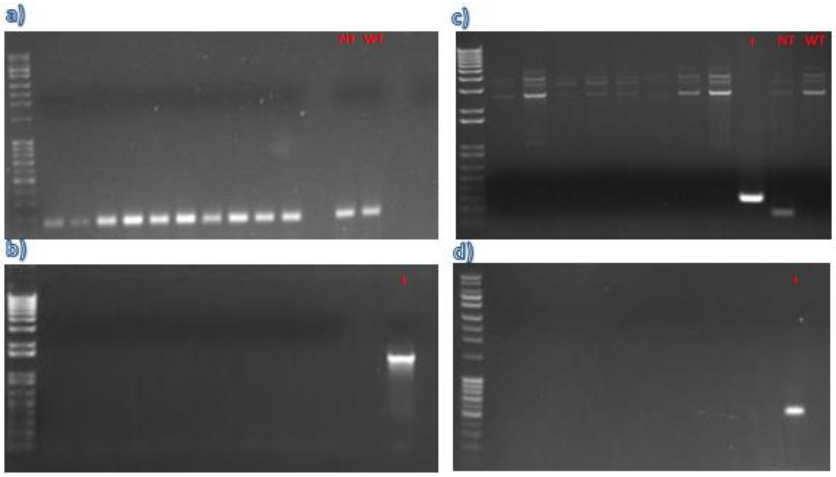

Figure 27. PCR of putatives transgenics. A) Amplification of actin as a control. B) Primers specific for the genes. C) Primers from the promoter to the beginning of the gene. D) Primers from the promoter to the middle of the gene. The positives used are from bacteria.

Several PCRs were done with more putatives that were obtained from about 6 transformation events, but it was unable to confirm presence of T-DNA construct (Figure 28). Primers that amplify from the promoter region to the genes (Figure 28b,d) and from the gene to the terminator (Figure 28c) were used.

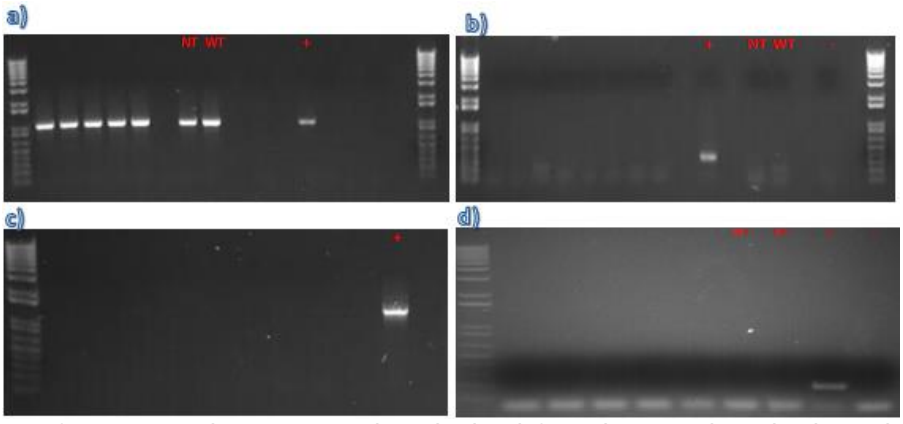

Figure 28. PCR of more putatives transgenics obtained from kanamycin selection. A) Rubisco as a control. B) Promoter to gene amplification. C) Gene to terminator (RBC!) amplification, d) promoter to gene amplification.

When trying to amplify the kanamycin gene to verify if the putative transgenics only had the empty construct we noticed that our control, the non-transgenic mutant (NT), also shows amplification (Figure 29a). Those mutants were created by T-DNA

insertion and kanamycin was used as a selectable marker. Beside the fact that the kanamycin gene should be silenced, as described in their web page, the codification sequence is present. For this reason amplification of the kanamycin gene is not a good choice for this mutants to look for the empty construct. Again, more PCR were carried one using different conditions such as lower or higher annealing temperature, more magnesium, more cycles, etc (Figure 29c,d), resulting in no amplification at all.

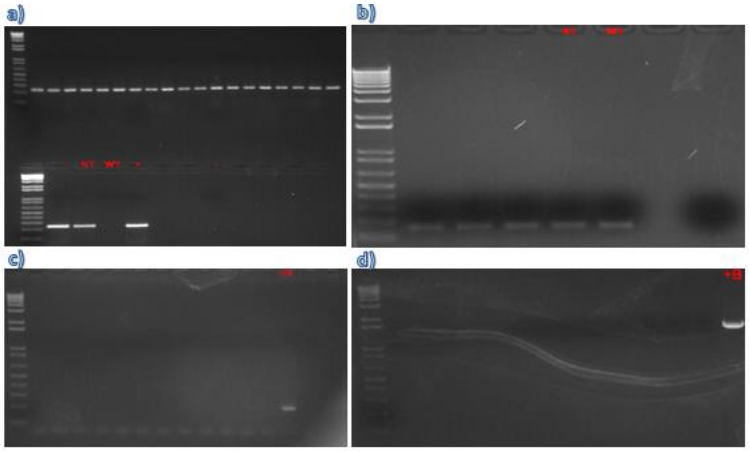

Figure 29. Putatives transgenics analyzed by PCR. A) Amplification of the kanamycin gene in the putatives. As can be seen, there is also amplification in the non-transgenic mutant (NT), meaning that this PCR cannot be used to determine the presence of the empty vector. The reason of the amplification in the NT is due to the kanamycin sequence that was used in the creation of the mutants by T-DNA mutagenesis. B) Amplification of the 18s subunit as a control. C) Promoter to terminator, using more magnesium. D) Gene to terminator with more cycles and a lower annealing temperature.

3.3.2 Floral dipping method using pB2GW7 as the plant expression vector

Plants care and conditions were as described in the previous section. Before every transformation event a miniprep of Agrobacterium was carried on in order to confirm the constructs (Figure 30).

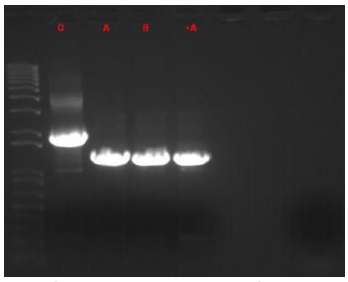

Figure 30. Confirmation of Agrobacterium constructs by PCR. Plasmids of Agrobacterium with *MANes;BsasA* (A), *MANes;BsasB* (B) and empty vector pB2GW7 (O) were amplify with primers from promoter to the terminator to verify the insert. The band in the empty vector is higher due to the presence of the ccdB.

The floral dipping method was performed as described in the previous part, without overnight incubation of the plants after the dipping. Only wild type plants (Col-0) were dipped with Agrobacterium carrying this new binary vector. Similarly, putatives transgenics were obtained for each construct when germinating seeds in MS media with Basta at 5-10 mg/L (Figure 31). However, no putative were obtained when germinating seeds in soil and spraying the plants with Basta herbicide. PCR was carried on for the confirmation of the constructs but no amplification was observed in the samples (Figure 32). Although neither of the plants got amplification the ones that had some type of amplification were used in the following PCRs.

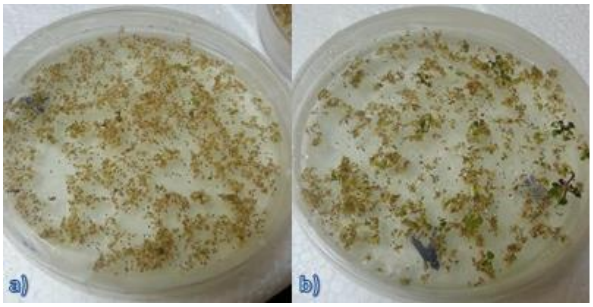

Figure 31. Screening of the M1 seeds in MS with Basta. A) No putatives were obtained in this plate. B) A few putatives were obtained when transforming WT plants with *MANes;BsasA.*

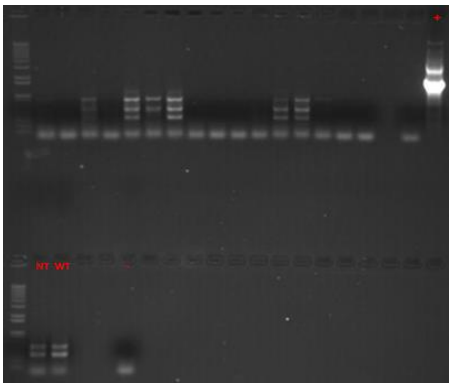

Figure 32. PCR of putatives transgenics surviving the Basta treatment in MS media. Primers from the promoter to the terminator area were used. No amplification in the samples was obtained, meaning that they were false putatives.

Three putatives for each construct (O; Empty Vector, A; *MANes;BsasA*, B; *MANes;BsasB*) were chosen for more analysis. A touchdown PCR was performed with the 9 sample with primers specific for the gene (Figure 33, upper gel) and primers from promoter to terminator (Figure 33, lower gel). A faint band was observed in some of the samples but it is also present in the samples from the plants transformed with empty vector, meaning that this faint band was not associated with the transgene because a higher band corresponding to the ccdB gene should be present in those plants transformed with the empty vector.

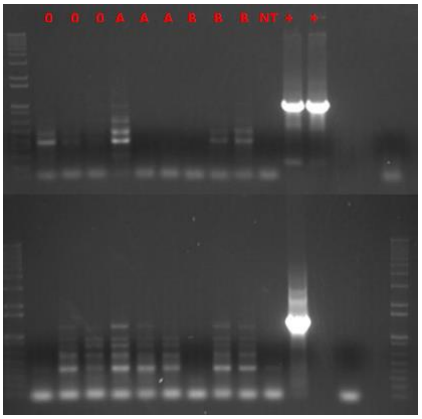

Figure 33. Analysis of putatives transgenics for the presence of the constructs. In the upper part of the gel primers specific for the genes were used. Only a faint is visible but it doesn't correspond to the genes. In the lower part of the gel primers from promoter to terminator was performed. Again, the faint bands in the samples don't correspond to the genes.

The samples were analyzed for the presence of the Basta (Bar) resistance gene (Figure 34). Beside the fact that some of the samples show amplification, there is also amplification in the NT. Two different sets of primers for basta were used but both of them amplify in the NT.

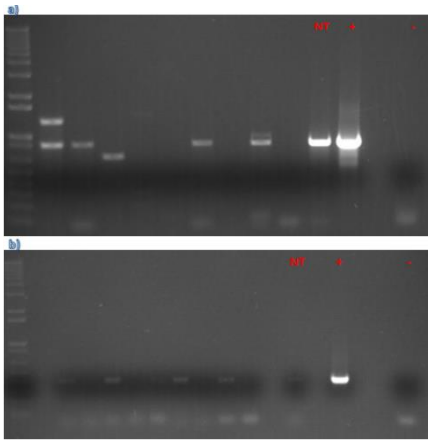

Figure 34. Analysis of putatives transgenics for Basta resistance gene (Bar). A) Touchdown PCR with primers for basta. The strong band in the non-transgenic mutant (NT) make it impossible to determine if the putatives have the insert. B) Regular PCR with a different set of primers for Basta. Again, a faint band in the NT make it impossible for the confirmation of the transgenics.

Finally, a touchdown PCR for Bar and for the genes was performed with a high annealing temperature to make it very specific for the inserts (Figure 35). Amplification in the NT was obtained again for Basta, and no amplification with the primers specific for the genes.

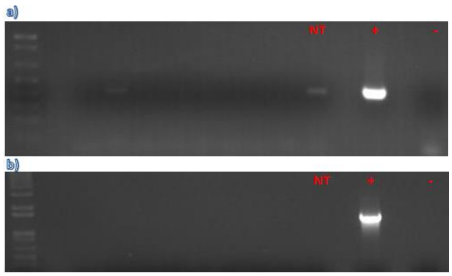

Figure 35. Screening of the transgenics by PCR. A) Amplification of Bar. Amplification was observed in one of the samples corresponding to empty vector and in the NT. B) Primers used amplifies from the promoter to the terminator. No amplification was observed. In both gels a touchdown PCR was performed with a high annealing temperature for a higher specify.

The last putatives transgenics obtained through Basta selection were also analyzed by touchdown PCR, but no amplification was observed (Figure 36).

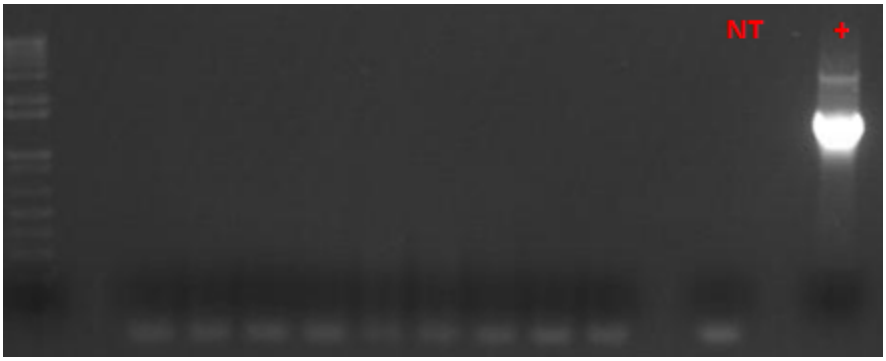

Figure 36. Analysis of more putatives transgenics.

Finally, the promoter and terminator of the selectable marker Basta (Bar) were amplified (Figure 37) to verify that those important components of the expression cassette were intact and no rearrangements occurred. A deletion or mutation in the promoter or terminator could result in a lack of transgenics because neither of the plants will grow under Basta herbicide if translation does not occurs.

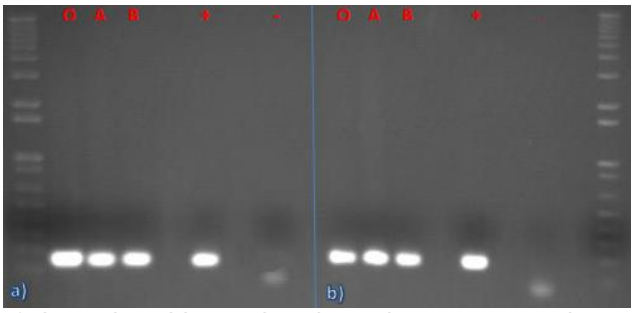

Figure 37. Amplification of the selectable marker (Basta) promoter and terminator. A) NOS promoter amplification. B) NOS terminator amplification. Results indicate that both the promoter and terminator are present. However, it is impossible to determine if any of them contain point mutations.

3.4 Enzymatic assays

The enzymatic assays were performed for the 3 different mutants (mitochondrial CS, mitochondrial β-CAS, and cytosolic CS) and Wild type. The enzymatic activity of β-CAS was analyzed twice (Figure 38), obtaining similar values both times. The mitochondrial β-CAS mutant presented the lowest β-CAS activity, which makes sense because it lacks the only true β-CAS responsible for the majority of the CAS activity in the cell. The second lowest activity corresponds to the cytosolic CS mutant. The cytosolic CS is the most important *Bsas* gene due to its major contribution to both CS and CAS activity in the cell as demonstrated by Watanabe *et al*. (2008). Lastly, the activity of the mitochondrial CS mutant is very similar to the activity of the wild type plants.

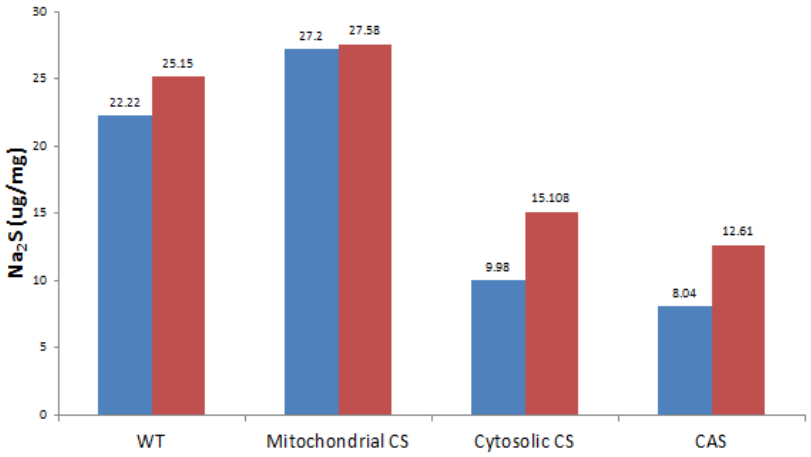

Figure 38. β-CAS activity in *Arabidopsis thaliana* mutants and wild type. The lowest activity corresponds to the β-CAS mutant who lacks the only true β-CAS in Arabidopsis. The second lowest activity correspond to the cytosolic CS mutant which lacks the most important *Bsas* gene in *Arabidopsis* as demonstrated in Watanabe *et al*. (2008). This gene is the major contributor to CS activity and the second most contributor of β-CAS activity. Finally, the mitochondrial CS mutant, who lacks the only mitochondrial CS in Arabidopsis, has a similar activity that the Wild Type.

The activity of CS was measured two times for each mutant and wild type (Figure 39). The lowest activity corresponds to the cytosolic CS mutant, which is what it was

predicted. The second lowest correspond to the β-CAS mutant as expected and finally the mitochondrial CS mutant has a very similar activity to the wild type. As shown in the literature, the mitochondrial CS don't contribute much to the overall CS and β-CAS activity in the cell and the lack of this enzyme can be compensated by the other *Bsas* proteins.

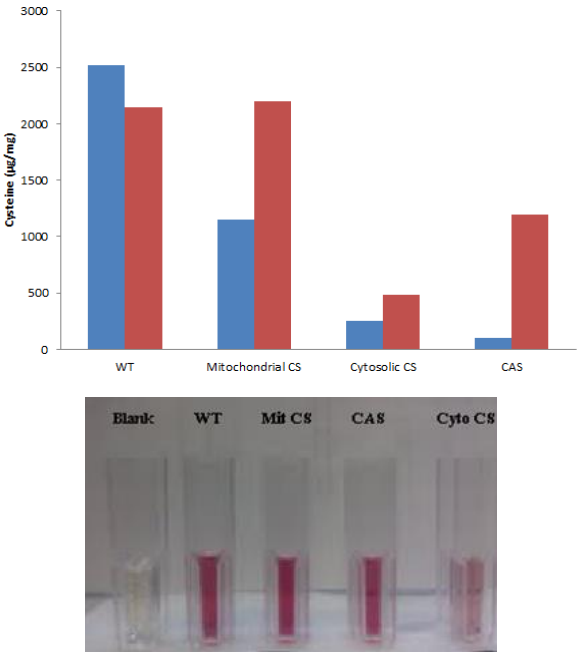

Figure 39. CS activity of Arabidopsis mutants and wild type. The lowest activity correspond to the cytosolic CS mutant, which lacks the most important CS in Arabidopsis, followed by the mitochondrial CAS mutant. The mitochondrial CS mutant has a similar activity to the wild type. The lower picture corresponds to the second enzymatic assay (red columns).

For comparison with the literature only the Wild Type was taken into consideration. For β-CAS activity a third reaction was prepared. The average of the 3 samples (26.06 µg/mg) is higher than the one reported in the literature (5.11 µg/mg) (Figure 40). This discrepancy can be due to differences in the preparation of the reagents and the equipment used for the detection of the coloration.

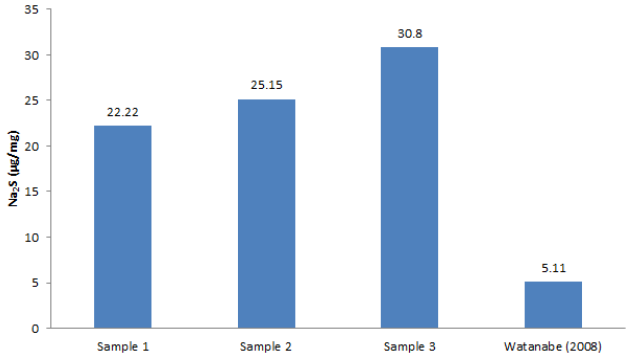

Figure 40. Comparison of β-CAS activity in wild type plants.

Similarly, the CS activity of the WT was compared to three different sources from the literature. A third reaction was carried on for CS activity, but when comparing it can be seen that the average obtained is higher than the numbers reported in the literature (Figure 41). Even there are differences between the three different publications used for the comparison. The CS enzymatic assay is very sensitive and variations in the amount of the crude extract added can significantly affect the reading of the spectrophotometer, which could be a possible explanation of the discrepancies.

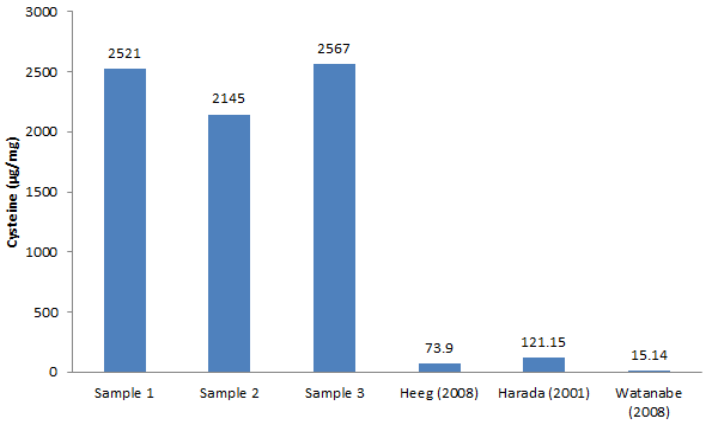

Figure 41. Comparison of CS activity in wild type plants. The activity obtained is much higher than the ones reported in the literature.

DISCUSSION

4 Discussions

A key part of the transformation of Arabidopsis is to grow healthy plants and keep them healthy until they begin to dry mature. This includes having the correct temperature, photoperiod, soil type and moisture. In addition, plants need space to grow optimally and fertilizer is needed periodically. A typical good healthy Arabidopsis plant may produce thousands of seeds with a high rate of germination. For this study, the necessary care was given to the plants, starting from the seeds to the mature plants. Seeds where carefully sterilized and germinated in media with the nutrients necessary to have optimal germination. A good rate of germination (>90%) was always observed, meaning that the sterilization procedure and the germination conditions were exceptional. Seedlings were grown in a plant incubator with optimal temperature and photoperiod, and it was constantly checked for the correct conditions. Bottom-watering was performed as indicated in the literature, about twice a week, and good quality soil was used. It is clear that healthy Arabidopsis plants were obtained in our laboratory, each one producing thousands of seeds with a high rate of germination. However, the selection protocol employed didn't make it possible to confirm transgenic lines with the T-DNA integrated into its genome.

When performing the floral dipping, as stated in the literature, it was observed that the plants died days after, especially in the above parts that came in contact with the Agrobacterium solution. When doing literature review it was seen that this problem has occurred in other laboratories as well. For example, one student wrote that he was having the same problem after the dipping, where the plants were dying after the treatment (check on www.researchgate.net , Question: Does someone know what could

have caused my Arabidopsis plant to die the next day after floral dip with Agrobacterium in 5% sucrose and 0.05% Silwet L-77 solution?, on May 18, 2013). This was probably caused by the addition of too much Silwet L-77 in the Agrobacterium solution, thus affecting the plant tissues. The recommended use of Silwet is 0.05%, however, Zhang *et al*. (2006) do not recommend a higher concentration than 0.02% because it might be toxic. The combination of the incubation overnight, providing a high humidity chamber to the plants, plus a high concentration of Silwet did probably caused the death of the plant tissues. However, after eliminating the incubation step the plants were able to grow normally, but the Silwet concentration could have interfered anyways in the transformation. Silwet is a surfactant that removes superficial tension, allowing the solution to penetrate deep into the plant tissues. Too much silwet could cause too much infection by the Agrobacterium, resulting in dead plant cells or a high copy number of the T-DNA which can also be toxic to the seeds. In future experiments, a concentration of 0.02% or less should be tried.

Another potential area of improvement is the removal of excess liquid after the dipping by draining the plants for a few seconds because the flower buds can be damaged if soaked in the solution for too long (Zhang et al., 2006). Again, this can result in too much Agrobacterium infection, or the Silwet itself can be very harmful to the plants because it is a detergent. It is possible that the drainage method employed in our laboratory was not efficient, resulting in damage inside the flower buds. It is important to remember that although we can see that the plants look healthy we do not completely understand what is ongoing inside the tissues.

When looking at scientific forums we can see that there are many other examples of people having the same challenges with obtaining successful transgenic plants. For example, another student asked if there is a fast and efficient method of transferring a gene into Arabidopsis after many attempts over a year (check on www.researchgate.net, January 28, 2014). Many scientist responded to this question and there are a few aspect that all of them mentioned, such as the stage of the plant at the transformation stage. It is important to have the plant at the right stage of flowering to have an efficient transformation event. It is possible that the stage at which the plants in our study was dipped was not the optimal for them to be successfully transformed. Despite the fact that the plants were healthy that is a critical step in the floral dipping method.

Furthermore, a mutation in one of the T-DNA borders could result in no integration of the T-DNA into the plant genome. Similarly, a mutation in the selectable marker or in the promoter of it could also result in negative results. Furthermore, the helper plasmid that comes within the Agrobacterium strain contains all the virulence genes in charge of the transfer of the T-DNA into the plant genome. There are many virulence genes involved in this process, and some of them have different isoforms. For example, VirB gene has 11 isoforms, while VirD has 3. A mutation in one of them could alter the transfer mechanism of the T-DNA because all of them work together as a unit, each one with an important role in this process. Sequencing of the virulence genes, T-DNA borders and selectable markers would be the next step to make sure that everything is normal and that there are no mutations preventing the integration of the insert into the plant genome. It is important to mention that this is the first time in our

laboratory working with these vectors and trying to produce Arabidopsis transgenic plants. Because we don't have positive controls it might be necessary to do this sequencing step.

In addition, very similar to point mutations, gene rearrangement is an event that occurs and in bacteria it is observed more frequently. Occasional rearrangement includes deletions, tandem duplications and inversions (Roth *et al*., 1996). This type of mutations can occur either in the helper plasmid or in the expression vector while in Agrobacteria, especially when grown for a few generations. Although PCR was used for the confirmation of the constructs and the genes, it doesn't provide information regarding mutations or plasmid rearrangements.

Another possible explanation is the fact that Agrobacterium has the ability to insert transgenes in the genome as tandem arrays, which can trigger RNAi of the resistance gene. RNA-interference (RNAi) is a mechanism in which a gene is silenced at the expression level. Transgene silencing is common in plants when the transgene is locally repeated (Matzke *et al*., 1994). It was demonstrated that changing the number of copies within a small array at a single site causes repeat-induced gene silencing in *A. thaliana* (Reviewed in Henikoff, 1998). In plants, the randomly insertion of transgenes in the genome by illegitimate recombination results in different number of inserted copies, chromosomal location and local arrangement (tandem insertion) between one transformant and another. There is an inverse correlation between copy number and gene expression, meaning that a high number of copies of a particular gene can lead to gene silencing (Vaucheret and Fagard, 2001).

5 Recommendations

- When adding Silwet L-77, a lower concentration such as 0.02% should be used
- The plants needs to be well drained after the dipping
- For each transformation event different plants stages should be tested
- More than one strain of Agrobacterium should be used for the transformations
- Before transforming with the desire construct a control test can be done by using just the empty vector to see if the protocol is working
- If having problems with the overnight incubation of the plants, incubate them for just a few hours or eliminate that step
- If transgenic plants are not obtained, sequencing of the T-DNA borders of the vectors and virulence genes from Agrobacterium should be performed

6 References

Akopyan, T., Braunstein, A. and Goryachenkova, E. 1975. Beta-cyanoalanine synthase: purification and characterization. *Proceedings of the National Academy of Sciences of the United States of America* 72: 1617-1621.

Alvarez, C., Calo, L., Romero, L. C., García, I. and Gotor, C. 2009. An O-acetylserine (thiol) lyase homolog with L-cysteine desulfhydrase activity regulates cysteine homeostasis in Arabidopsis. *Plant Physiology* 152: 656-669.

Alves, A. 2002. Cassava botany and physiology. In *Cassava: biology, production and utilization*, ed. R. J. Hillocks and J. M. Thresh, 67-89. Wallingford: CABI.

Atmaca, G. 2004. Antioxidant effects of sulfur-containing amino acids. *Yonsei Medical Journal* 45:776-788.

Babilonia, K. and Siritunga, D. 2012. Cysteine Biosynthesis in Plants: Essential Step in Sulfur-Metabolism. In *Cysteine: Biosynthesis, Chemical Structure and Toxicity*.

Ballatori, N., Krance, S. M., Notenboom, S., Shi, S., Tieu, K. and Hammond, C. L. 2009. Glutathione dysregulation and the etiology and progression of human diseases. *Biological Chemistry* 390: 191-214.

Bechtold, N., Ellis, J. and Pelletier, G. 1993. In Planta Agrobacterium Mediated Gene Transfer by Infiltration of Adult Arabidopsis Thaliana Plants. *Methods in Molecular Biology* 82: 1194–1199.

Bernard, P., Gabarit, P., Bahassi, E. M. and Couturier, M. 1994. Positive-selection Vectors Using the F Plasmid ccdB Killer Gene. *Gene* 148: 71–74.

Bonner, E. R., Cahoon, R. E., Knapke, S. M. and Jez, J. M. 2005. Molecular basis of cysteine biosynthesis in plants: structural and functional analysis of O-acetylserine sulfhydrylase from *Arabidopsis thaliana*. *The Journal of Biological Chemistry* 280: 38803-38813.

Bork, C., Schwenn, J. D. and Hell, R. 1998. Isolation and characterization of a gene for assimilatory sulfite reductase from *Arabidopsis thaliana*. *Gene* 212: 147-153.

Bradbury, J., Cliff, J. and Denton, I. 2011. Uptake of wetting method in Africa to reduce cyanide poisoning and konzo from cassava. *Food and Chemical Toxicology: An International Journal Published for the British Industrial Biological Research Association* 49: 539-542.

Bradbury, J. 2009. Development of a sensitive picrate method to determine total cyanide and acetone cyanohydrin contents of gari from cassava. *Food Chemistry* 113: 1329-1333.

Bradbury, J. and Denton, I. 2010. Rapid wetting method to reduce cyanogen content of cassava flour. *Food Chemistry* 121: 591-594.

Bradbury, J. and Denton, I. 2011. Mild methods of processing cassava leaves to remove cyanogens and conserve key nutrients. *Food Chemistry* 127: 1755-1759.

Burandt, P., Schmidt, A. and Papenbrock, J. 2002. Three O -acetyl-L-serine(thiol)lyase isoenzymes from Arabidopsis catalyse cysteine synthesis and cysteine desulfuration at different pH values. *Journal of Plant Physiology* 159: 111-119.

Chavarriaga-Aguirre, P. and Halsey, M. 2005. Cassava (*Manihot esculenta* Crantz): Reproductive biology and practices for confinement of experimental field trials. Report prepared for the Program for Biosafety Systems. Washington, D.C.

Cliff, J., Muquingue, H., Nhassico, D., Nzwalo, H. and Bradbury, J. 2011. Konzo and continuing cyanide intoxication from cassava in Mozambique. *Food and Chemical Toxicology* 49: 631-635.

Clough, S. and Bent, A. 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *The Plant Journal: For Cell and Molecular Biology* 16: 735-743.

Cobbett, C. S. 2000. Phytochelatins and Their Roles in Heavy Metal Detoxification. *Plant Physiology* 123: 825 -832.

Dang, T. and Christie, P. 1997. The VirB4 ATPase of Agrobacterium tumefaciens is a cytoplasmic membrane protein exposed at the periplasmic surface. *Journal of Bacteriology* 179: 453-462.

Dominguez-Solís, J. R., Gutierrez-Alcalá, G., Vega, J. M., Romero, L. C. and Gotor, C. 2001. The cytosolic O-acetylserine (thiol) lyase gene is regulated by heavy metals and can function in cadmium tolerance. *The Journal of Biological Chemistry* 276: 9297-9302.

Dröge, W. 1999. Cysteine and glutathione in catabolic conditions and immunological dysfunction. *Current Opinion in Clinical Nutrition and Metabolic Care* 2: 227-233.

Droux, M., Ruffet, M. L., Douce, R. and Job, D. 1998. Interactions between serine acetyltransferase and O-acetylserine (thiol) lyase in higher plants--structural and kinetic properties of the free and bound enzymes. *European Journal of Biochemistry* 255: 235-245.

Droux, M. 2003. Plant serine acetyltransferase: New insights for regulation of sulphur metabolism in plant cells. *Plant Physiol*. *Biochem*. 41: 619–627.

Droux, M. 2004. Sulfur Assimilation and the Role of Sulfur in Plant Metabolism: A Survey. *Photosynthesis Research* 79: 331-348.

El Ghawabi, S., Gaafar, M., El-Saharti, A., Ahmed, S., Malash, K. and Fares, R. 1975. Chronic cyanide exposure: a clinical, radioisotope, and laboratory study. *British Journal of Industrial Medicine* 32: 215 -219.

Elias, M., Sudhakaran, P. and Nambisan, B. 1997. Purification and characterisation of [beta]-cyanoalanine synthase from cassava tissues. *Phytochemistry* 46: 469-472.

Estrela, J. M., Ortega, A. and Obrador, E. 2006. Glutathione in cancer biology and therapy. *Critical Reviews in Clinical Laboratory Sciences* 43: 143-181.

Food and Agriculture Organization of the United Nations, and International Fund for Agricultural Development. 2000. *The world cassava economy: facts, trends and outlook*.

Food and Agriculture Organization of the United Nations, and International Fund for Agricultural Development . 2004. *A cassava industrial revolution in Nigeria*.

Gaitonde, M. 1967. A spectrophotometric method for the direct determination of cysteine in the presence of other naturally occurring amino acids. *Biochem J* 104: 627-633.

Gallo, M. and Sayre, R. 2009. Removing allergens and reducing toxins from food crops. *Current Opinion in Biotechnology* 20: 191-196.

Gelvin, S. 2000. AGROBACTERIUM AND PLANT GENES INVOLVED IN T-DNA TRANSFER AND INTEGRATION. *Annual Review of Plant Physiology and Plant Molecular Biology* 51: 223-256.

Giovanelli, J., Mudd, S. H. and Datko, A. H. 1985. *In vivo* regulation of *denovo* methionine biosynthesis in a higher plant *(Lemna). Plant Physiol.* 77: 450-455.

Haas, F. H., Heeg, C., Queiroz, R., Bauer, A., Wirtz, M. and Hell, R. 2008. Mitochondrial serine acetyltransferase functions as a pacemaker of cysteine synthesis in plant cells. *Plant Physiology* 148: 1055-1067.

Harada, E., Choi, Y. E., Tsuchisaka, A., Obata, H. and Sano, H. 2001. Transgenic Tobacco Plants Expressing a Rice Cysteine Synthase Gene Are Tolerant to Toxic Levels of Cadmium. *Journal of Plant Physiology* 158: 655–661.

Hasegawa, R., Tada, T., Torii, Y. and Esashi, Y. 1994. Presence of β-cyanoalanine synthase in unimbibed dry seeds and its activation by ethylene during pre-germination. *Physiologia Plantarum* 91: 141-146.

Hatzfeld, Y., Maruyama, A., Schmidt, A., Noji, M., Ishizawa, K. and Saito, K. 2000. β-Cyanoalanine Synthase Is a Mitochondrial Cysteine Synthase-Like Protein in Spinach and Arabidopsis. *Plant Physiology* 123: 1163 -1172.

Heeg, C., Kruse, C., Jost, R., Gutensohn, M., Ruppert, T., Wirtz, M. and Hell, R. 2008. Analysis of the Arabidopsis O-acetylserine (thiol) lyase gene family demonstrates compartment-specific differences in the regulation of cysteine synthesis. *The Plant Cell* 20: 168-185.

Hell, R. and Bergmann, L. 1990. γ-Glutamylcysteine synthetase in higher plants: catalytic properties and subcellular localization. *Planta* 180: 603-612.

Hell, R., Bork, C., Bogdanova, N., Frolov, I. and Hauschild, R. 1994. Isolation and characterization of two cDNAs encoding for compartment specific isoforms of O-acetylserine (thiol) lyase from Arabidopsis thaliana. *FEBS Letters* 351: 257-262.

Henikoff, S. 1998. Conspiracy of Silence Among Repeated Transgenes. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 20: 532–535.

Hesse, H., Lipke, J., Altmann, T. and Höfgen, R. 1999. Molecular cloning and expression analyses of mitochondrial and plastidic isoforms of cysteine synthase (O-acetylserine (thiol) lyase) from *Arabidopsis thaliana*. *Amino Acids* 16: 113-131.

Hunter, E. A. and Grimble, R. F. 1997. Dietary sulfur amino acid adequacy influences glutathione synthesis and glutathione-dependent enzymes during the inflammatory response to endotoxin and tumour necrosis factor-alpha in rats. *Clinical Science* (London, England: 1979) 92: 297-305.

Jørgensen, K., Bak, S., Busk, P., Sorensen, C., Olsen, C., Puonti-Kaerlas, J. and Moller, B. 2005. Cassava Plants with a Depleted Cyanogenic Glucoside Content in Leaves and Tubers. Distribution of Cyanogenic Glucosides, Their Site of Synthesis and Transport, and Blockage of the Biosynthesis by RNA Interference Technology. *Plant Physiology* 139: 363 -374.

Jorgensen, K., Morant, A., Morant, M., Jensen, N., Olsen, C., Kannangara, R., Motawia, M., Moller, B. and Bak, S. 2010. Biosynthesis of the cyanogenic glucosides linamarin and lotaustralin in cassava (Manihot esculenta Crantz): Isolation, biochemical characterization and expression pattern of CYP71E7, the oxime-metabolizing cytochrome P450 enzyme. *Plant physiology*155: 282-292.

Jost, R., Berkowitz, O., Wirtz, M., Hopkins, L., Hawkesford, M. J. and Hell, R. 2000. Genomic and functional characterization of the oas gene family encoding O-acetylserine (thiol) lyases, enzymes catalyzing the final step in cysteine biosynthesis in *Arabidopsis thaliana*. *Gene* 253: 237-247.

Kuske, C. R., Hill, K. K., Guzman, E. and Jackson, P. J. 1996. Subcellular Location of O-Acetylserine Sulfhydrylase Isoenzymes in Cell Cultures and Plant Tissues of *Datura innoxia* Mill. *Plant Physiology* 112: 659-667.

Lai, K., Yau, C., Tse, Y., Jiang, L. and Yip, W. 2009. Heterologous expression analyses of rice OsCAS in Arabidopsis and in yeast provide evidence for its roles in cyanide detoxification rather than in cysteine synthesis in vivo. *Journal of Experimental Botany* 60: 993 -1008.

Landy, A. 1989. Dynamic, Structural, and Regulatory Aspects of Lambda Site-specific Recombination. *Annual Review of Biochemistry* 58: 913–949.

Leustek, T., Martin, M. N., Bick, J. A. and Davies, J. P. 2000. Pathways and regulation of sulfur metabolism revealed through molecular and genetic studies. *Annual Review of Plant Physiology and Plant Molecular Biology* 51: 141-165.

Linka, M. and Weber, A. P. 2005. Shuffling ammonia between mitochondria and plastids during photorespiration. *Trends in Plant Science* 10: 461-465.

Lunn, J., Droux, M., Martin, J. and Douce, R. 1990. Localization of ATP Sulfurylase and O-Acetylserine(thiol)lyase in Spinach Leaves. *Plant Physiology* 94: 1345-1352.

Marrero-Degro, J., Marcano-Velázquez, J. and Siritunga, D. 2010. Isolation and Characterization of Novel β-Cyanoalanine Synthase and Cysteine Synthase Genes from Cassava. *Plant Molecular Biology Reporter*.

Maruyama, A., Ishizawa, K., Takagi, T. and Esashi, Y. 1998. Cytosolic beta-cyanoalanine synthase activity attributed to cysteine synthases in cocklebur seeds. Purification and characterization of cytosolic cysteine synthases. *Plant and Cell Physiology* 39: 671-680.

Maruyama, A., Saito, K. and Ishizawa, K. 2001. Beta-cyanoalanine synthase and cysteine synthase from potato: molecular cloning, biochemical characterization, and spatial and hormonal regulation. *Plant Molecular Biology* 46: 749-760.

Matzke, M., Matzke, J. M. and Scheid, O. M. 1994. Inactivation of Repeated Genes — DNA-DNA Interaction? *Homologous Recombination and Gene Silencing in Plants:* 271–307.

Mendoza-Cózatl, D., Loza-Tavera, H., Hernández-Navarro, A. and Moreno-Sánchez, R. 2005. Sulfur assimilation and glutathione metabolism under cadmium stress in yeast, protists and plants. *FEMS Microbiology Reviews* 29: 653-671.

Meyer, T., Burow, M., Bauer, M. and Papenbrock J. 2003. Arabidopsis sulfurtransfe-rases: investigation of their function during senescence and in cyanide detoxification. *Planta* 217: 1 - 10.

Nambisan, B. 2011. Strategies for elimination of cyanogens from cassava for reducing toxicity and improving food safety. *Food and Chemical Toxicology* 49: 690-693.

Noctor, G., Gomez, L., Vanacker, H. and Foyer, C. H. 2002. Interactions between biosynthesis, compartmentation and transport in the control of glutathione homeostasis and signaling. *Journal of Experimental Botany* 53: 1283 -1304.

Noji, M., Inoue, K., Kimura, N., Gouda, A. and Saito, K. 1998. Isoform-dependent differences in feedback regulation and subcellular localization of serine acetyltransferase involved in cysteine biosynthesis from *Arabidopsis thaliana*. *The Journal of Biological Chemistry* 273: 32739-32745.

Noji, M., Takagi, Y., Kimura, N., Inoue, K., Saito, M., Horikoshi, M., Saito, F., Takahashi, H. and Saito, K. 2001. Serine acetyltransferase involved in cysteine biosynthesis from spinach: molecular cloning, characterization and expression analysis of cDNA encoding a plastidic isoform. *Plant and Cell Physiology* 42: 627-634.

O'Brien, G. and Ip, S. 2008. Initial evaluation of a field-friendly incubation procedure for the colorimetric assay of cassava linamarase. *Food Chemistry* 107: 576-579.

Oluwole, O., Onabolu, A., Link, H. and Rosling, H. 2000. Persistence of tropical ataxic neuropathy in a Nigerian community. *Journal of Neurology, Neurosurgery & Psychiatry* 69: 96 -101.

Parcell, S. 2002. Sulfur in human nutrition and applications in medicine. *Alternative Medicine Review: A Journal of Clinical Therapeutic* 7: 22-44.

Peters, J., Cnudde, F. and Gerats, T. 2003. Forward genetics and map-based cloning approaches. *Trends in Plant Science* 8: 484-491.

Ravanel, S., Droux, M. and Douce, R. 1995. Methionine biosynthesis in higher plants. I. Purification and characterization of cystathionine gamma-synthase from spinach chloroplasts. *Archives of Biochemistry and Biophysics* 316: 572-584.

Ravanel, S., Gakière, B., Job, D. and Douce, R. 1998. The specific features of methionine biosynthesis and metabolism in plants. *Proceedings of the National Academy of Sciences* 95: 7805 -7812.

Ristoff, E. and Larsson, A. 2007. Inborn errors in the metabolism of glutathione. *Orphanet Journal of Rare Diseases* 2: 16.

de la Riva, G., Gonzalez-Cabrera, J., Vazquez-Padron, R. and Ayra-Pardo, C. 1998. Agrobacterium tumefaciens: a natural tool for plant transformation. *Electronic Journal of Biotechnology* 1: 118-113.

Roth, J. R., Benson, N., Galitski, T., Haack, K., Lawrence, J. G. and Miesel, L. 1996. Rearrangements of the Bacterial Chromosome: Formation and Applications.

Saito, K., Kurosawa, M., Tatsuguchi, K., Takagi, Y. and Murakoshi, I. 1994. Modulation of cysteine biosynthesis in chloroplasts of transgenic tobacco overexpressing cysteine synthase [O-acetylserine (thiol) - lyase]. *Plant Physiology* 106: 887-895.

Samiec, P. S., Drews-Botsch, C., Flagg, E. W., Kurtz, J. C., Sternberg, P. Jr, Reed, R. L. and Jones, D. P. 1998. Glutathione in human plasma: decline in association with aging, age-related macular degeneration, and diabetes. *Free Radical Biology and Medicine* 24: 699-704.

Sanderfoot, A. and Raikhel, N. 2001. Arabidopsis could shed light on human genome. *Nature* 410: 299.

Sasaki, Y., Sone, T., Yoshida, S., Yahata, K., Hotta, J., Chesnut, J., Honda, T. and Imamoto, F. 2004. Evidence for High Specificity and Efficiency of Multiple Recombination Signals in Mixed DNA Cloning by the Multisite Gateway System. *Journal of Biotechnology* 107: 233–243.

Sayre, R., Beeching, J., Cahoon, E., Egesi, C., Fauquet, C., Fellman, J., Fregene, M., Gruissem, W., Mallowa, S., Manary, M., Maziya-Dixon, B., Mbanaso, A., Schachtman, D., Siritunga, D., Taylor, N., Vanderschuren, H. and Zhang, P. 2011. The BioCassava Plus Program: Biofortification of Cassava for Sub-Saharan Africa. *Annu Rev Plant Biol* 62: 251-272.

Schardl, C., Byrd, A., Benzion, G., Altschuler, M., Hildebrand, D. and Hunt, A. 1987. Design and construction of a versatile system for the expression of foreign genes in plants. *Gene* 61: 1-11.

Sessions, A., Burke, E., Presting, G., Aux, G., McElver, J., Patton, D., Dietrich, B., et al. 2002. A High-Throughput Arabidopsis Reverse Genetics System. *The Plant Cell Online* 14: 2985 -2994.

Shibagaki, N., Rose, A., McDermott, J. P., Fujiwara, T., Hayashi, H., Yoneyama, T. and Davies, J. P. 2002. Selenate-resistant mutants of *Arabidopsis thaliana* identify Sultr1;2, a sulfate transporter required for efficient transport of sulfate into roots. *The Plant Journal: For Cell and Molecular Biology* 29: 475-486.

Siritunga, D. and Sayre, R. 2004. Engineering cyanogen synthesis and turnover in cassava (Manihot esculenta). *Plant Mol Biol* 56:661–669.

Siritunga, D. and Sayre, R. 2007. Transgenic approaches for cyanogen reduction in cassava. *Journal of AOAC International* 90: 1450-1455.

Stipanuk, M. H., Dominy, J. E., Lee, J. I. and Coloso, R. M. 2006. Mammalian Cysteine Metabolism: New Insights into Regulation of Cysteine Metabolism. *The Journal of Nutrition* 136: 1652S -1659S.

Tai, C. H. and Cook, P. F. 2000. O-Acetylserine Sulfhydrylase. *Adv. Enzymol. Relat. Areas Mol. Biol*. 74: 185–234.

Takahashi, H. and Saito, K. 1996. Subcellular localization of spinach cysteine synthase isoforms and regulation of their gene expression by nitrogen and sulfur. *Plant Physiology* 112: 273-280.

Tattersall, D., Bak, S., Jones, P., Olsen, C. E., Nielsen, J., Hansen, M., Hoj, P. and Moller, B. 2001. Resistance to an Herbivore Through Engineered Cyanogenic Glucoside Synthesis. *Science* 293: 1826-1828.

Tonukari, N. 2004. Cassava and the future of starch. *Electronic Journal of Biotechnology* 7.

Vaucheret, H. and Fagard, M. 2001. Transcriptional Gene Silencing in Plants: Targets, Inducers and Regulators. *Trends in Genetics* 17*:*29–35.

Warrilow, A. and Hawkesford, M. 2000. Cysteine synthase (O-acetylserine (thiol) lyase) substrate specificities classify the mitochondrial isoform as a cyanoalanine synthase. *Journal of Experimental Botany* 51: 985 -993.

Warrilow, A. and Hawkesford, M. 1998. Separation, subcellular location and influence of sulphur nutrition on isoforms of cysteine synthase in spinach. *Journal of Experimental Botany* 49: 1625 -1636.

Watanabe, M., Kusano, M., Oikawa, A., Fukushima, A., Noji, M. and Saito, K. 2008. Physiological Roles of the β-Substituted Alanine Synthase Gene Family in Arabidopsis. *Plant Physiology* 146: 310 -320.

White, W., Arias-Garzon, D., McMahon, J. and Sayre, R. 1998. Cyanogenesis in Cassava. *Plant Physiology* 116: 1219 -1225.

White, W., McMahon, J. and Sayre, R. 1994. Regulation of cyanogenesis in cassava. *Acta Hortic* 375:69–77

Wilson, Z. 2000. *Arabidopsis: A Practical Approach*. 1st ed. Oxford University Press, USA, March 16.

Wirtz, M., Berkowitz, O., Droux, M. and Hell, R. 2001. The cysteine synthase complex from plants. Mitochondrial serine acetyltransferase from *Arabidopsis thaliana* carries a bifunctional domain for catalysis and protein-protein interaction. *European Journal of Biochemistry* 268: 686-693.

Wirtz, M. and Hell, R. 2006. Functional analysis of the cysteine synthase protein complex from plants: structural, biochemical and regulatory properties. *Journal of Plant Physiology* 163: 273-286.

Wirtz, M., Droux, M. and Hell, R. 2004. O-acetylserine (thiol) lyase: an enigmatic enzyme of plant cysteine biosynthesis revisited in Arabidopsis thaliana." *Journal of Experimental Botany* 55: 1785 -1798.

Wurtele, E., Nikolau, B. and Conn, E. 1985. Subcellular and Developmental Distribution of beta-Cyanoalanine Synthase in Barley Leaves." *Plant Physiology* 78: 285-290.

Yamaguchi, Y., Nakamura, T., Kusano, T. and Sano, H. 2000. Three Arabidopsis Genes Encoding Proteins with Differential Activities for Cysteine Synthase and β-Cyanoalanine Synthase. *Plant and Cell Physiology* 41: 465 -476.

Youssefian, S., Nakamura, M., Orudgev, E. and Kondo, N. 2001. Increased cysteine biosynthesis capacity of transgenic tobacco overexpressing an O-acetylserine(thiol) lyase modifies plant responses to oxidative stress. *Plant Physiology* 126: 1001-1011.

Zhang, C., Meng, Q., Zhang, M., Huang, F., Gai, J. and Yu, D. 2008. Characterization of O-Acetylserine (Thiol) Lyase-Encoding Genes Reveals Their Distinct but Cooperative Expression in Cysteine Synthesis of Soybean [Glycine max (L.) Merr.]. *Plant Molecular Biology Reporter* 26: 277-291.

Zhang, X., Henriques, R., Lin, S., Niu, Q. and Chua, N. 2006. *Agrobacterium*-mediated transformation of Arabidopsis thaliana using the floral dip method. *Nature protocols* 1: 641-646.

Zupan, J. and Zambryski, P. 1995. Transfer of T-DNA from Agrobacterium to the plant cell. *Plant Physiology* 107: 1041-1047.

APPENDIXES

7 APPENDIXES

APPENDIX A. List of primers used in this study

Table A.1: *Bsas* cloning primers used in this study

Primer name	Sequence used for design	Primer sequence
MAN-FCP-SacI	*M. esculenta(MANes;BsasA)*	ccgcagggagctcctatggaagctc
MAN-RCP-XbaI	*M. esculenta(MANes;BsasB)*	ggtctagattagtcaactggaactgg
Manes-CASB NcoI-SacI	*M. esculenta(MANes;BsasB)*	ccggagctcccatggaatctc
CACC-BsasA Fw	*M. esculenta(MANes;BsasA)*	caccccgctgctatggaagctcc
CACC-BsasB Fw	*M. esculenta(MANes;BsasB)*	caccaagctgcgatggaatctcc
BsasA/B Rv	*M. esculenta(MANes;BsasA)*	cttagtcaactggaactgg

Table A.2: Primers used for the screening of putatives transgenic plants

Primer name	Sequence used for design	Primer sequence
MANes-Lpz-FW1	*M. esculenta(MANes;BsasA*	Gacggtcttgatagagccaacttc
MANes-Lpz-FW2	*M. esculenta(MANes;BsasA*	Gtggagcttacatagccgtcaag
MANes-Lpz-RV1	*M. esculenta(MANes;BsasA*	Ccctcatagtcacccttctttcc
MANes-Lpz-RV2	*M. esculenta(MANes;BsasA*	Cttgacggctatgtaagctccac
35sFW1	pKYLX	aagttcgacggagaaggtgacg
35sFW5.1	pKYLX	tcaccagtctctctctaagc
35sFW5.2	pKYLX	ggatcctcgagctgcagg
RBC	pKYLX	Gttgtcgaaaccgatgatacgaacg
RBC-Rv3	pKYLX	Gattctggtgtgtgcgcaatg
RBC-RV3.2	pKYLX	Acttcggtcattagaggccacgat
nospromFW1	pKYLX	Ccgcaacgattgaaggagccac
nospromFW2	pKYLX	Caaaaatgctccactgacgttcc
p35sFW1	pB2GW7	tgctgacccacagatggttag
p35sFW2	pB2GW7	aaacctcctcggattccattg
p35sFW3	pB2GW7	ccactatccttcgcaagacc
t35sRV1	pB2GW7	ctggtgatttttgcggactc
t35sRV2	pB2GW7	gctcaacacatgagcgaaacc
BarFW	pB2GW7	gaagtccagctgccagaaac
BarRV	pB2GW7	agtcgaccgtgtacgtctcc
nptII FW1 dpsc	pKYLX	atgattgaacaagatggattgcac
nptII RV1 dpsc	pKYLX	tcagaagaactcgtcaagaaggcg
BarFW	pB2GW7	gaagtccagctgccagaaac
BarRV	pB2GW7	agtcgaccgtgtacgtctcc
BarFW-17	pB2GW7	Atgccggtcgatctagtaac
BarFW-701	pB2GW7	Gaggtcgtccgtccactcc
BarRV-1082	pB2GW7	Ggacaagccgttttacgtttg
BarRV-846	pB2GW7	Tcgaggggatctaccatgag

Table A.3: Other primers used in this study

Primer name	Sequence used for design	Primer sequence
VirD2 Fw	A. tumefaciens (VirD2)	atgcccgatcgcgctgaagt
VirD2 Rv	A. tumefaciens (VirD2)	cctgacccaaacatctcggct
ManActin11fw	M. esculenta	tgcaatgtatgttgccatccaggc
ManActin11Rv	M. esculenta	ttacaccgtcaccagaatccagca
pNOS-FW	pB2GW7	ggaactgacagaaccgcaacg
pNOS-RV	pB2GW7	tggaacgtcagtggagcattt
tNOS-FW	pB2GW7	gaatcctgttgccggtcttg
tNOS-RV	pB2GW7	ttatcctagtttgcgcgcta

APPENDIX B. Sequences of the vectors and genes used in this study.

Sequence of the expression cassette in pKYLX (Source: http://www.uky.edu/~aghunt00/kylx.html)

>pKYLX
GAATTCGCCCGGGGATCTCCTTTGCCCCAGAGATCACAATGGACGACTTCCTATATCTCTACGATCTAG
TCAGGAAGTTCGACGGAGAAGGTGACGATACCATGTTCACCACTGATAATGAGAAGATTAGCCTTTTC
AATTTCAGAAAGAATCCTAACCCACAGATGGTTAGAGACGCTTACGCAGCAGGTCTCATCAAGACGAT
CTACCCGAGCAATAATCTCCAGGAGATCAAATACCTTCCCAAGAAGGTTAAAGATGCAGTCAAAAGAT
TCAGGACTAACTGCATCAAGAACACAGAGAAAGATATATTTCTCAAGATCAGAAGTACTATTCCAGTA
TGGACGATTCAAGGCTTGCTTCACAAACCAAGGCAAGTAATAGAGATTGGAGTCTCTAAAAAGGTAGT
TCCCACTGAATCAAAGGCCATGGAGTCAAAGATTCAAATAGAGGACCTAACAGAACTCGCCGTAAAG
ACTGGCGAACAGTTCATACAGAGTCTCTTACGACTCAATGACAAGAAGAAAATCTTCGTC{AACATGG
TGGAGCACGACACGCTTGTCTACCTCCAAAAATATCAAAGATACAGTCTCAGAAGACCAAAGGGAATT
GAGACTTTTCAACAAAGGGTAATATCCGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTT
ATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTGCGATAAAGGAAAGGCCA
TCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAA
AAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGAT}AACATGGTGGAGCACGACA
CGCTTGTCTACCTCCAAAAATATCAAAGATACAGTCTCAGAAGACCAAAGGGAATTGAGACTTTTCAA
CAAAGGGTAATATCCGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTTATTGTGAAGATA
GTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTGCGATAAAGGAAAGGCCATCGTTGAAGATG
CCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGT
TCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATATCTCCACTGACGTAAGGGATGACGCACAAT
CCCACTATCCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTCATTTCATTTGGAGAGGACACGCTGAA
ATCACCAGTCTCTCTCTAAGCTTGGATCCTCGAGCTGCAGGAGCTCGAATTGATCCTCTAGAGCTTTCG
TTCGTATCATCGGTTTCGACAACGTTCGTCAAGTTCAATGCATCAGTTTCATTGCGCACACACCAGAAT
CCTACTGAGTTCGAGTATTATGGCATTGGGAAAACTGTTTTTCTTGTACCATTTGTTGTGCTTGTAATTT
ACTGTGTTTTTTATTCGGTTTTCGCTATCGAACTGTGAAATGGAAATGGATGGAGAAGAGTTAATGAAT
GATATGGTCCTTTTGTTCATTCTCAAATTAATATTATTTGTTTTTTCTCTTATTTGTTGTGTGTTGAATTT
GAAATTATAAGAGATATGCAAACATTTTGTTTTGAGTAAAAATGTGTCAAATCGTGGCCTCTAATGAC
CGAAGTTAATATGAGGAGTAAAACACTTGTAGTTGTACCATTATGCTTATTCACTAGGCAACAAATAT
ATTTTCAGACCTAGAAAAGCTGCAAATGTTACTGAATACAAGTATGTCCTCTTGTGTTTTAGACATTTA
TGAACTTTCCTTTATGTAATTTTCCAGAATCCTTGTCAGATTCTAATCATTGCTTTATAATTATAGTTAT
ACTCATGGATTTGTAGTTGAGTATGAAAATATTTTTTAATGCATTTTATGACTTGCCAATTGATTGACA
ACATGCATCAATCGAT

Sequence of pB2GW7 vector (Source: http://www.informaxinc.com/)

>pB2GW7

CTCCCATATGGTCGACTAGAGCCAAGCTGATCTCCTTTGCCCCGGAGATCACCATGGACGACTTTCTCT
ATCTCTACGATCTAGGAAGAAAGTTCGACGGAGAAGGTGACGATACCATGTTCACCACCGATAATGAG
AAGATTAGCCTCTTCAATTTCAGAAAGAATGCTGACCCACAGATGGTTAGAGAGGCCTACGCGGCAGG
TCTCATCAAGACGATCTACCCGAGTAATAATCTCCAGGAGATCAAATACCTTCCCAAGAAGGTTAAAG
ATGCAGTCAAAAGATTCAGGACTAACTGCATCAAGAACACAGAGAAAGATATATTTCTCAAGATCAG
AAGTACTATTCCAGTATGGACGATTCAAGGCTTGCTTCATAAACCAAGGCAAGTAATAGAGATTGGAG
TCTCTAAGAAAGTAGTTCCTACTGAATCAAAGGCCATGGAGTCAAAAATTCAGATCGAGGATCTAACA
GAACTCGCCGTGAAGACTGGCGAACAGTTCATACAGAGTCTTTTACGACTCAATGACAAGAAGAAAAT
CTTCGTCAACATGGTGGAGCACGACACTCTCGTCTACTCCAAGAATATCAAAGATACAGTCTCAGAAG
ACCAAAGGGCTATTGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCATTGCCCA
GCTATCTGTCACTTCATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTACAAATGCCATCATTGCGA
TAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGA
GGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATATCTCC
ACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTC
ATTTCATTTGGAGAGGACTCCGGTATTTTTACAACAATACCACAACAAAACAAACAACAAACAACATT
ACAATTTACTATTCTAGTCGACCTGCAGGCGGCCGCACTAGTGATATCACAAGTTTGTACAAAAAAGC
TGAACGAGAAACGTAAAATGATATAAATATCAATATATTAAATTAGATTTTGCATAAAAAACAGACTA
CATAATACTGTAAAACACAACATATCCAGTCACTATGGCGGCCGCATTAGGCACCCCAGGCTTTACAC
TTTATGCTTCCGGCTCGTATAATGTGTGGATTTTGAGTTAGGATCCGGCGAGATTTTCAGGAGCTAAGG
AAGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAAGA
ACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGC
CTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCT
GATGAATGCTCATCCGGAATTCCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTC
ACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACG
ATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCC
CTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTT
AAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCG
ACAAGGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTCTGTGATGGCTTCCATGTCGGCAGA
ATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAAACGCGTGGATCCGGCTT
ACTAAAAGCCAGATAACAGTATGCGTATTTGCGCGCTGATTTTTGCGGTATAAGAATATATACTGATA
TGTATACCCGAAGTATGTCAAAAAGAGGTGTGCTATGAAGCAGCGTATTACAGTGACAGTTGACAGCG
ACAGCTATCAGTTGCTCAAGGCATATATGATGTCAATATCTCCGGTCTGGTAAGCACAACCATGCAGA
ATGAAGCCCGTCGTCTGCGTGCCGAACGCTGGAAAGCGGAAAATCAGGAAGGGATGGCTGAGGTCGC
CCGGTTTATTGAAATGAACGGCTCTTTTGCTGACGAGAACAGGGACTGGTGAAATGCAGTTTAAGGTT
TACACCTATAAAAGAGAGAGCCGTTATCGTCTGTTTGTGGATGTACAGAGTGATATTATTGACACGCC
CGGGCGACGGATGGTGATCCCCCTGGCCAGTGCACGTCTGCTGTCAGATAAAGTCTCCCGTGAACTTT
ACCCGGTGGTGCATATCGGGGATGAAAGCTGGCGCATGATGACCACCGATATGGCCAGTGTGCCGGTC
TCCGTTATCGGGGAAGAAGTGGCTGATCTCAGCCACCGCGAAAATGACATCAAAAACGCCATTAACCT
GATGTTCTGGGGAATATAAATGTCAGGCTCCCTTATACACAGCCAGTCTGCAGGTCGACCATAGTGAC
TGGATATGTTGTGTTTTACAGTATTATGTAGTCTGTTTTTTATGCAAAATCTAATTTAATATATTGATAT
TTATATCATTTTACGTTTCTCGTTCAGCTTTCTTGTACAAAGTGGTGATATCCCGCGGCCATGCTAGAGT
CCGCAAAAATCACCAGTCTCTCTCTACAAATCTATCTCTCTCTATTTTTCTCCAGAATAATGTGTGAGT
AGTTCCCAGATAAGGGAATTAGGGTTCTTATAGGGTTTCGCTCATGTGTTGAGCATATAAGAAACCCTT
AGTATGTATTTGTATTTGTAAAATACTTCTATCAATAAAATTTCTAATTCCTAAAACCAAAATCCAGTG
ACCTGCAGGCATGCGACGTCGGGCCCAAGCTTAGCTTGAGCTTGGATCAGATTGTCGTTTCCCGCCTTC
AGTTTAAACTATCAGTGTTTGACAGGATATATTGGCGGGTAAACCTAAGAGAAAAGAGCGTTTATTAG
AATAACGGATATTTAAAAGGGCGTGAAAAGGTTTATCCGTTCGTCCATTTGTATGTGCATGCCAACCA
CAGGGTTCCCCTCGGGATCAAAGTACTTTGATCCAACCCCTCCGCTGCTATAGTGCAGTCGGCTTCTGA
CGTTCAGTGCAGCCGTCTTCTGAAAACGACATGTCGCACAAGTCCTAAGTTACGCGACAGGCTGCCGC
CCTGCCCTTTTCCTGGCGTTTTCTTGTCGCGTGTTTTAGTCGCATAAAGTAGAATACTTGCGACTAGAA
CCGGAGACATTACGCCATGAACAAGAGCGCCGCCGCTGGCCTGCTGGGCTATGCCCGCGTCAGCACCG
ACGACCAGGACTTGACCAACCAACGGGCCGAACTGCACGCGGCCGGCTGCACCAAGCTGTTTTCCGAG
AAGATCACCGGCACCAGGCGCGACCGCCCGGAGCTGGCCAGGATGCTTGACCACCTACGCCCTGGCG
ACGTTGTGACAGTGACCAGGCTAGACCGCCTGGCCCGCAGCACCCGCGACCTACTGGACATTGCCGAG

CGCATCCAGGAGGCCGGCGCGGGCCTGCGTAGCCTGGCAGAGCCGTGGGCCGACACCACCACGCCGG
CCGGCCGCATGGTGTTGACCGTGTTCGCCGGCATTGCCGAGTTCGAGCGTTCCCTAATCATCGACCGCA
CCCGGAGCGGGCGCGAGGCCGCCAAGGCCCGAGGCGTGAAGTTTGGCCCCCGCCCTACCCTCACCCCG
GCACAGATCGCGCACGCCCCGCGAGCTGATCGACCAGGAAGGCCGCACCGTGAAAGAGGCGGCTGCAC
TGCTTGGCGTGCATCGCTCGACCCTGTACCGCGCACTTGAGCGCAGCGAGGAAGTGACGCCCACCGAG
GCCAGGCGGCGCGGTGCCTTCCGTGAGGACGCATTGACCGAGGCCGACGCCCTGGCGGCCGCCGAGA
ATGAACGCCAAGAGGAACAAGCATGAAACCGCACCAGGACGGCCAGGACGAACCGTTTTTCATTACC
GAAGAGATCGAGGCGGAGATGATCGCGGCCGGGTACGTGTTCGAGCCGCCCGCGCACGTCTCAACCG
TGCGGCTGCATGAAATCCTGGCCGGTTTGTCTGATGCCAAGCTGGCGGCCTGGCCGGCCAGCTTGGCC
GCTGAAGAAACCGAGCGCCGCCGTCTAAAAAGGTGATGTGTATTTGAGTAAAACAGCTTGCGTCATGC
GGTCGCTGCGTATATGATGCGATGAGTAAATAAACAAATACGCAAGGGGAACGCATGAAGGTTATCG
CTGTACTTAACCAGAAAGGCGGGTCAGGCAAGACGACCATCGCAACCCATCTAGCCCGCGCCCTGCAA
CTCGCCGGGGCCGATGTTCTGTTAGTCGATTCCGATCCCCAGGGCAGTGCCCGCGATTGGGCGGCCGT
GCGGGAAGATCAACCGCTAACCGTTGTCGGCATCGACCGCCCGACGATTGACCGCGACGTGAAGGCC
ATCGGCCGGCGCGACTTCGTAGTGATCGACGGAGCGCCCCAGGCGGCGGACTTGGCTGTGTCCGCGAT
CAAGGCAGCCGACTTCGTGCTGATTCCGGTGCAGCCAAGCCCTTACGACATATGGGCCACCGCCGACC
TGGTGGAGCTGGTTAAGCAGCGCATTGAGGTCACGGATGGAAGGCTACAAGCGGCCTTTGTCGTGTCG
CGGGCGATCAAAGGCACGCGCATCGGCGGTGAGGTTGCCGAGGCGCTGGCCGGGTACGAGCTGCCCA
TTCTTGAGTCCCGTATCACGCAGCGCGTGAGCTACCCAGGCACTGCCGCCGCCGGCACAACCGTTCTT
GAATCAGAACCCGAGGGCGACGCTGCCCGCGAGGTCCAGGCGCTGGCCGCTGAAATTAAATCAAAAC
TCATTTGAGTTAATGAGGTAAAGAGAAAATGAGCAAAAGCACAAACACGCTAAGTGCCGGCCGTCCG
AGCGCACGCAGCAGCAAGGCTGCAACGTTGGCCAGCCTGGCAGACACGCCAGCCATGAAGCGGGTCA
ACTTTCAGTTGCCGGCGGAGGATCACACCAAGCTGAAGATGTACGCGGTACGCCAAGGCAAGACCATT
ACCGAGCTGCTATCTGAATACATCGCGCAGCTACCAGAGTAAATGAGCAAATGAATAAATGAGTAGA
TGAATTTTAGCGGCTAAAGGAGGCGGCATGGAAAATCAAGAACAACCAGGCACCGACGCCGTGGAAT
GCCCCATGTGTGGAGGAACGGGCGGTTGGCCAGGCGTAAGCGGCTGGGTTGTCTGCCGGCCCTGCAAT
GGCACTGGAACCCCCAAGCCCGAGGAATCGGCGTGACGGTCGCAAACCATCCGGCCCGGTACAAATC
GGCGCGGCGCTGGGTGATGACCTGGTGGAGAAGTTGAAGGCCGCGCAGGCCGCCCAGCGGCAACGCA
TCGAGGCAGAAGCACGCCCCGGTGAATCGTGGCAAGCGGCCGCTGATCGAATCCGCAAAGAATCCCG
GCAACCGCCGGCAGCCGGTGCGCCGTCGATTAGGAAGCCGCCCAAGGGCGACGAGCAACCAGATTTT
TTCGTTCCGATGCTCTATGACGTGGGCACCCGCGATAGTCGCAGCATCATGGACGTGGCCGTTTTCCGT
CTGTCGAAGCGTGACCGACGAGCTGGCGAGGTGATCCGCTACGAGCTTCCAGACGGGCACGTAGAGG
TTTCCGCAGGGCCGGCCGGCATGGCCAGTGTGTGGGATTACGACCTGGTACTGATGGCGGTTTCCCAT
CTAACCGAATCCATGAACCGATACCGGGAAGGGAAGGGAGACAAGCCCGGCCGCGTGTTCCGTCCAC
ACGTTGCGGACGTACTCAAGTTCTGCCGGCGAGCCGATGGCGGAAAGCAGAAAGACGACCTGGTAGA
AACCTGCATTCGGTTAAACACCACGCACGTTGCCATGCAGCGTACGAAGAAGGCCAAGAACGGCCGC
CTGGTGACGGTATCCGAGGGTGAAGCCTTGATTAGCCGCTACAAGATCGTAAAGAGCGAAACCGGGC
GGCCGGAGTACATCGAGATCGAGCTAGCTGATTGGATGTACCGCGAGATCACAGAAGGCAAGAACCC
GGACGTGCTGACGGTTCACCCCGATTACTTTTTGATCGATCCCGGCATCGGCCGTTTTCTCTACCGCCT
GGCACGCCGCGCCGCAGGCAAGGCAGAAGCCAGATGGTTGTTCAAGACGATCTACGAACGCAGTGGC
AGCGCCGGAGAGTTCAAGAAGTTCTGTTTCACCGTGCGCAAGCTGATCGGGTCAAATGACCTGCCGGA
GTACGATTTGAAGGAGGAGGCGGGGCAGGCTGGCCCGATCCTAGTCATGCGCTACCGCAACCTGATCG
AGGGCGAAGCATCCGCCGGTTCCTAATGTACGGAGCAGATGCTAGGGCAAATTGCCCTAGCAGGGGA
AAAAGGTCGAAAAGGTCTCTTTCCTGTGGATAGCACGTACATTGGGAACCCAAAGCCGTACATTGGGA
ACCGGAACCCGTACATTGGGAACCCAAAGCCGTACATTGGGAACCGGTCACACATGTAAGTGACTGAT
ATAAAAGAGAAAAAAGGCGATTTTTCCGCCTAAAACTCTTTAAAACTTATTAAAACTCTTAAAACCCG
CCTGGCCTGTGCATAACTGTCTGGCCAGCGCACAGCCGAAGAGCTGCAAAAAGCGCCTACCCTTCGGT
CGCTGCGCTCCCTACGCCCCGCCGCTTCGCGTCGGCCTATCGCGGCCGCTGGCCGCTCAAAAATGGCT
GGCCTACGGCCAGGCAATCTACCAGGGCGCGGACAAGCCGCGCCGTCGCCACTCGACCGCCGGCGCC
CACATCAAGGCACCCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCC
GGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCG
GGTGTTGGCGGGTGTCGGGGCGCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTATACTGGCTT
AACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATG
CGTAAGGAGAAAATACCGCATCAGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGT
TCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATA
ACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGC

TGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGG
CGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGT
TCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAG
CTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCC
CGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTT
ATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAG
TTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAA
GCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTG
GTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTT
CTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGCATGATATATCT
CCCAATTTGTGTAGGGCTTATTATGCACGCTTAAAAATAATAAAAGCAGACTTGACCTGATAGTTTGG
CTGTGAGCAATTATGTGCTTAGTGCATCTAATCGCTTGAGTTAACGCCGGCGAAGCGGCGTCGGCTTG
AACGAATTTCTAGCTAGACATTATTTGCCGACTACCTTGGTGATCTCGCCTTTCACGTAGTGGACAAAT
TCTTCCAACTGATCTGCGCGCGAGGCCAAGCGATCTTCTTCTTGTCCAAGATAAGCCTGTCTAGCTTCA
AGTATGACGGGCTGATACTGGGCCGGCAGGCGCTCCATTGCCCAGTCGGCAGCGACATCCTTCGGCGC
GATTTTGCCGGTTACTGCGCTGTACCAAATGCGGGACAACGTAAGCACTACATTTCGCTCATCGCCAG
CCCAGTCGGGCGGCGAGTTCCATAGCGTTAAGGTTTCATTTAGCGCCTCAAATAGATCCTGTTCAGGA
ACCGGATCAAAGAGTTCCTCCGCCGCTGGACCTACCAAGGCAACGCTATGTTCTCTTGCTTTTGTCAGC
AAGATAGCCAGATCAATGTCGATCGTGGCTGGCTCGAAGATACCTGCAAGAATGTCATTGCGCTGCCA
TTCTCCAAATTGCAGTTCGCGCTTAGCTGGATAACGCCACGGAATGATGTCGTCGTGCACAACAATGG
TGACTTCTACAGCGCGGAGAATCTCGCTCTCTCCAGGGGAAGCCGAAGTTTCCAAAAGGTCGTTGATC
AAAGCTCGCCGCGTTGTTTCATCAAGCCTTACGGTCACCGTAACCAGCAAATCAATATCACTGTGTGG
CTTCAGGCCGCCATCCACTGCGGAGCCGTACAAATGTACGGCCAGCAACGTCGGTTCGAGATGGCGCT
CGATGACGCCAACTACCTCTGATAGTTGAGTCGATACTTCGGCGATCACCGCTTCCCCCATGATGTTTA
ACTTTGTTTTAGGGCGACTGCCCTGCTGCGTAACATCGTTGCTGCTCCATAACATCAAACATCGACCCA
CGGCGTAACGCGCTTGCTGCTTGGATGCCCGAGGCATAGACTGTACCCCAAAAAAACATGTCATAACA
AGAAGCCATGAAAACCGCCACTGCGCCGTTACCACCGCTGCGTTCGGTCAAGGTTCTGGACCAGTTGC
GTGACGGCAGTTACGCTACTTGCATTACAGCTTACGAACCGAACGAGGCTTATGTCCACTGGGTTCGT
GCCCGAATTGATCACAGGCAGCAACGCTCTGTCATCGTTACAATCAACATGCTACCCTCCGCGAGATC
ATCCGTGTTTCAAACCCGGCAGCTTAGTTGCCGTTCTTCCGAATAGCATCGGTAACATGAGCAAAGTCT
GCCGCCTTACAACGGCTCTCCCGCTGACGCCGTCCCGGACTGATGGGCTGCCTGTATCGAGTGGTGATT
TTGTGCCGAGCTGCCGGTCGGGGAGCTGTTGGCTGGCTGGTGGCAGGATATATTGTGGTGTAAACAAA
TTGACGCTTAGACAACTTAATAACACATTGCGGACGTTTTTAATGTACTGAATTAACGCCGAATTGAAT
TATCAGCTTGCATGCCGGTCGATCTAGTAACATATAGATGACACCGCGCGCGATAATTTATCCTAGTTT
GCGCGCTATATTTTGTTTTCTATCGCGTATTAAATGTATAATTGCGGGACTCTAATCATAAAAACCCAT
CTCATAAATAACGTCATGCATTACATGTTAATTATTACATGCTTAACGTAATTCAACAGAAATTATATG
ATAATCATCGCAAGACCGGCAACAGGATTCAATCTTAAGAAACTTTATTGCCAAATGTTTGAACGATC
TGCTTGACTCTAGGGGTCATCAGATTTCGGTGACGGGCAGGACCGGACGGGGCGGCACCGGCAGGCT
GAAGTCCAGCTGCCAGAAACCCACGTCATGCCAGTTCCCGTGCTTGAAGCCGGCCGCCCGCAGCATGC
CGCGGGGGGCATATCCGAGCGCCTCGTGCATGCGCACGCTCGGGTCGTTGGGCAGCCCGATGACAGCG
ACCACGCTCTTGAAGCCCTGTGCCTCCAGGGACTTCAGCAGGTGGGTGTAGAGCGTGGAGCCCAGTCC
CGTCCGCTGGTGGCGGGGGGAGACGTACACGGTCGACTCGGCCGTCCAGTCGTAGGCGTTGCGTGCCT
TCCAGGGACCCGCGTAGGCGATGCCGGCGACCTCGCCGTCCACCTCGGCGACGAGCCAGGGATAGCG
CTCCCGCAGACGGACGAGGTCGTCCGTCCACTCCTGCGGTTCCTGCGGCTCGGTACGGAAGTTGACCG
TGCTTGTCTCGATGTAGTGGTTGACGATGGTGCAGACCGCCGGCATGTCCGCCTCGGTGGCACGGCGG
ATGTCGGCCGGGCGTCGTTCTGGGCTCATGGTAGATCCCCTCGATCGAGTTGAGAGTGAATATGAGAC
TCTAATTGGATACCGAGGGGAATTTATGGAACGTCAGTGGAGCATTTTTGACAAGAAATATTTGCTAG
CTGATAGTGACCTTAGGCGACTTTTGAACGCGCAATAATGGTTTCTGACGTATGTGCTTAGCTCATTAA
ACTCCAGAAACCCGCGGCTCAGTGGCTCCTTCAACGTTGCGGTTCTGTCAGTTCCAAACGTAAAACGG
CTTGTCCCGCGTCATCGGCGGGGGTCATAACGTGACTCCCTTAATTCTCATGTATGATAATTCGAGCT

Sequence of pENTR/SD/D-TOPO (Source: https://www.lablife.org/)

>pENTR/SD/D-TOPO
CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCG
CCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAA
ACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAG
CGGGCAGTGAGCGCAACGCAATTAATACGCGTACCGCTAGCCAGGAAGAGTTTGTAGAAACGCAAAA
AGGCCATCCGTCAGGATGGCCTTCTGCTTAGTTTGATGCCTGGCAGTTTATGGCGGGCGTCCTGCCCGC
CACCCTCCGGGCCGTTGCTTCACAACGTTCAAATCCGCTCCCGGCGGATTTGTCCTACTCAGGAGAGCG
TTCACCGACAAACAACAGATAAAACGAAAGGCCCAGTCTTCCGACTGAGCCTTTCGTTTTATTTGATG
CCTGGCAGTTCCCTACTCTCGCGTTAACGCTAGCATGGATGTTTTCCCAGTCACGACGTTGTAAAACGA
CGGCCAGTCTTAAGCTCGGGCCCCAAATAATGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAAC
AAATTGATGAGCAATGCTTTTTTATAATGCCAACTTTGTACAAAAAAGCAGGCTCCGCGGCCGCCTTGT
TTAACTTTAAGAAGGAGCCCTTCACCAAGGGTGGGCGCGCCGACCCAGCTTTCTTGTACAAAGTTGGC
ATTATAAGAAAGCATTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCAT
TATTTGCCATCCAGCTGATATCCCCTATAGTGAGTCGTATTACATGGTCATAGCTGTTTCCTGGCAGCT
CTGGCCCGTGTCTCAAAATCTCTGATGTTACATTGCACAAGATAAAAATATATCATCATGAACAATAA
AACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCGAGG
CCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCA
ATCAGGTGCGACAATCTATCGCTTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCA
AAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCT
CTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGA
AAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGT
GTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTC
GCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGG
CTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTTTTGCCATTCTCACCGGATTCAGTCGTCAC
TCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGG
ACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTC
CTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTC
ATTTGATGCTCGATGAGTTTTTCTAATCAGAATTGGTTAATTGGTTGTAACACTGGCAGAGCATTACGC
TGACTTGACGGGACGGCGCAAGCTCATGACCAAAATCCCTTAACGTGAGTTACGCGTCGTTCCACTGA
GCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGC
TTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTT
TCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAG
GCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTG
CTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAG
CGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGA
GATACCTACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCC
GGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTT
TATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGG
AGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCAC
ATGTT

Sequence of *MANes;BsasA* (accession number EU350583)

>MANes;BsasA

CACTGGAGTGTGTTGTCCACTGCGAGAGAGAGCCTAGAGGGTTGTAATAGTGATACCGTGGGTGAGGA
AAAGCTTCTAGCTTTCATCAATGGCTACTCTTACATCAATGGCTACTCTTAGGAACTTGTTCACGAAAA
AATCTTTTGCCTCAAACGAGCTTGCTATGCGGAGGTTCTTCACACCGCAGGCCGCTGCTGCTGAAGCTC
CTTCTTTTGCTCAAAGAGTCAGGGATCTGCCCAAGAATCTCCCTGCAACTAAGATCAGGACTGAAGTTT
CTCATCTCATTGGGAGAACTCCCCTTGTTTTTCTTAACAAAGTGAATGAAGGATGTGGAGCTTACATAG
CCGTCAAGCAAGAGATGATGCAACCAACTGCCAGCATCAAAGACAGGCCGGCTTTTTCAATGTTCAAT
GATGCAGAAAAGAAGAATTTAATCTCTCCTGGGAAGACGGTCTTGATAGAGCCAACTTCTGGTAATAT
GGGGATTAGTATGGCTTTTATGGCAGCCATGAAAGGATACAAGATGATTTTAACCATGCCCTCTTACA
CAAGCTTGGAAAGAAGGGTGACTATGAGGGCATTTGGAGCTGAGCTAATTCTCACTGATCCCACCAAG
GGGATGGGAGGAACAGTTAAGAAGGCCTATGATCTTTTGGAAACCACACCAAATGCTTACATGCTACA
ACAGTTTTCGAATCCTGCAAATAGTAAGATCCATTTTGAAACGACAGGGCCTGAAATATGGGAGGATA
CTCTTGGACAAGTTGATATCTTTGTAATGGGAATAGGCAGTGGAGGAACAGTCACTGGCGTTGGACAG
TACCTTAAATCCCAAAATCCTAATGTTAAGATATATGGAGTGGAGCCTGCTGAAAGTAATGTGCTGAA
TGGTGGTAAACCAGGTCCTCATCAAATTACTGGAAATGGAGTAGGATTTAAACCAGACATATTGGACA
TGGACGTTATGGAAAAAGTTCTTGAGGTTAGCAGTGAAGATGCGATAAAAATGGCTAGGAGGTTGGC
ATTGGAGGAGGGACTTATGGTGGGAATTTCATCTGGAGCCAACACAGTTGCTGCACTTAGACTTGCAA
GGATGCCAGAGAACAAAGGCAAACTCATCGTGACTGTTCATCCAAGTTTCGGGGAGCGATACTTGACT
TCTGTGCTGTTTGAACAACTGAGAAATGAGGCTGCAAACATGCAACCAGTTCCAGTTGACTAAGTGCT
GTCAAAAGGATAGTTCAATAAATAATATCAGAATCTATCGTGTGCAGCTTTTTGTTGACATCATGTATT
ACTTGTTAGCTCTCTTGGGTGAAAAACTACGTCTCCAGAATATTAGTTGTGATAAAGCTTTTGTAATCT
GGTGCAAGTTCTGCATTATGGCATCTGTGGAAAAAACTAGCTAGTTCAGCTTGAGTGTTGTTGTATAAG
CAAGCTTTATTAATCCAAAACAAAACAAAGTATCCGCTACTAAAAAAAAAAAAAAAAAAAAAA

Sequence of *MANes;BsasB*

>MANes;BsasB

ACTGGACTCTGTTATCCACTACTTGCGGTTGTTGCTTGTAATACTCTTGAGAGTTCAATGGCTTCCCTTA
GGAACTTGTTGAAGAAAAATTCTTTAACGTCCAACAATCTTCCTATAAGGAGGTTTTTCTCTTCCGAAG
CTGCGCTTGAATCTCCTTCATTTGCTCACAAAATCAGGGATCTGCCCAACAATCGTCCTCAAGCTAAAA
TCAAGCCTGAAGTTTCTCAGCTTATTGGGAAAACTCCTCTTGTTTACCTTAACAAAGTGACTGAAGGAT
GTGGAGCTTTCATTGCAGTCAAGCAAGAGATGGTGCAGCCAACTGCCAGTATCAAGGACAGGCCGGCT
TTTGCAATGATCAATGATGCAGAAAAGAAGAATTTAATTTCTCCTGGAAAGACGACCTTGATAGAGCC
AACTTCTGGAAATATGGGGATCAGTATGGCTTTTATGGCAGCCATGAAAGGATACAAGATGGTTCTAA
CCATGCCTTCTTACACCAGCTTGGAGAGGAGGGTAACTATGAGGGCATTTGGAGCTGAGCTAATTCTC
ACTGATCCCACCAAGGGAATGGGAGGAACAGTTAAGAAGGCTTTTGAGCTTTTGGAATCAACACCAA
ATGCTTTCATGCTGCAACAATTTTCAAATCCTGCAAATACTAAGATCCATTTTGAAACTACGGGTCCTG
AAATTTGGGAGGATACACTTGGACAAGTTGACATCTTTGTAATGGGAATAGGCAGCGGAGGAACAGT
CTCTGGTGTTGGACAGTACCTTAAATCTCAAAATCCTGCTGTTAAGATATATGGAGTGGAGCCTACTGA
AAGTAATGTGCTGAATGGTGGTAAACCAGGACCTCACCATATTACTGGAAACGGAGTTGGATTTAAAC
CAGATATATTGGACTTGGATGTAATGGAAAAAGTTCTTGAGGTTAGCAGCGAAGATGCAGTAAACATG
GCTCGGAGGTTGGCGTTGGAGGAGGGGCTTATGGTGGGAATATCATCCGGAGCCAACACAGTTGCTGC
GCTTAGACTTGCTAGAATGCCAGAGAACAAAGGCAAACTTATTGTGACTGTTCATCCAAGTTTCGGGG
AGCGATACTTGACTTCTGTCCTGTTTGAAGAACTGAGAAAAGAGGCTGCAAACATGCAACCAGTTCCA
GTTGACTAAGGCCTTCGAATGGAATGTTCTATGCAACCAATTCCTGTGCTGTGCAACCATTTGTTGAGA
TCGTATTACTTTTTAGCTCTGTTGAGTTTTAAACTGTGCCCAGAAAACTTATTGTGATAAAAGCTTATTT
GTCTAATAGAAATCCTATTTATGGCTTCAAAAAAAAAAAAAAAAAAAAAAA