GENETIC TRANSFORMATION OF SWITCHGRASS (*Panicum virgatum* L.) WITH ENDOGLUCANASE GENE AND CHARACTERIZATION OF PLANTS WITH ENDOGLUCANASE TRANSGENE

Madhavi Dere

Thesis submitted to the faculty of Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science

In

Horticulture

Bingyu Zhao

Richard Veilleux

Saghai Maroof

Boris Vinatzer

June 22, 2012

Blacksburg, VA

Key words: Switchgrass, *Panicum virgatum*, genetic transformation, hygromycin, GUS assay, chloroplast, RuBisCO, signal peptide, *Hypocrea jecorina*, endoglucanase, TAIL PCR, activation tagging, GFP, pSQ5

GENETIC TRANSFORMATION OF SWITCHGRASS (*Panicum virgatum* L.) WITH ENDOGLUCANASE GENE AND CHARACTERIZATION OF PLANTS WITH ENDOGLUCANASE TRANSGENE

Madhavi Dere

ABSTRACT

As a warm season grass native to the North American continent, switchgrass is considered as one of the most promising biofuel crops in the USA. It is a C4 plant that makes it energy efficient. Switchgrass has a deep root system that allows it to grow on marginal land with low water and nutrient input. Switchgrass has been used as a forage crop and its use for biofuel will not affect food security. Biofuels are more environment-friendly than fossil fuels as they do not produce net greenhouse gases. However, the problem of high cost of production per unit for biofuel has to be overcome if we want to replace fossil fuels with biofuels. One of the major factors related to the high cost of biofuel are the expensive cellulase enzymes used in the pretreatment of feedstock. Endoglucanase is the key enzyme used for breaking down cellulose before fermentation. Currently, endoglucanase is produced from engineered *E. coli* or yeast strains, which is still expensive for enzyme production and purification of industrial scales. Expression of endoglucanase in plants has been previously reported. However, there are no reports of transgenic switchgrass producing cellulase enzyme. In this study, the catalytic domain of beta-endoglucanase gene was codon-optimized and synthesized based on the cDNA cloned from *Hypocrea jecorina*. Rice RuBisCO small subunit targeting signal peptide was fused to the N-terminus of the beta-endoglucanase gene, which was expected to target the fusion protein to chloroplast. This subcellular compartment targeting could minimize negative effects on cell function and plant development. The endoglucanase gene was cloned with maize ubiquitin promoter in a modified binary vector pCambia 1305-2 and transformed into switchgrass genotype HR8 by using *Agrobacterium tumefaciens*. In this study, I generated five independent transgenic switchgrass lines and they were confirmed by growing on the selection agent hygromycin, GUS assay, PCR amplification, and southern blotting hybridization, for the presence of hygromycin and endoglucanase genes. However, based on RT-PCR analysis, only

two transgenic lines were confirmed to produce mRNAs of the endoglucanase gene. These two transgenic lines were further characterized for their agronomic traits and the chlorophyll contents. Our results suggested that expression of endoglucanase in switchgrass could reduce chlorophyll content and affect plant development. Nevertheless, in this study, we demonstrated that a fungal endoglucanase gene could be expressed in switchgrass transgenic plants, though the gene expression level and the subcellular localization need to be carefully regulated in order to minimize the toxic effect of endoglucanase on plant cells.

ACKNOWLEDGMENT

First of all, I would like to acknowledge Dr Zhao who is the chair of my thesis committee and who served as principal investigator in all projects. I learnt many things from Dr Zhao including all techniques and how to work simultaneously on many projects. I always feel inspired from him about his industry and perseverance that is required to be a successful researcher.

I am thankful to Dr Richard Veilleux for allowing me to use resources in his lab and his technical guidance. I am thankful to Dr Saghai Maroof for encouraging me to work always positively and making me assertive to reach my project goals. I am grateful to Dr Boris Vinatzer for his encouragement in my research work.

I would like to express gratitude to the head of department Dr Roger Harris for his trust in my abilities and helping me in reaching my career goals. I am also thankful to Joyce Shelton and other staff in the Department of Horticulture for their help in completing all paper work.

I am also acknowledging my first advisor Dr Rumen Conev and my friends Dr Bhadra Gunesekara, Dr Juan Ruiz-Rojas, Dr Qiang Cheng, Sy Traore, Kerri Mills, Nan Lu, Shuchi Wu and all friends for their help in learning new techniques and being there whenever I needed their help.

I am most indebted to my family. My father is the person who has influenced me most because of his convictions and his generosity as human being towards others throughout his life. I am also thankful to my mother Padmavati Dere and Bhushan Dere for their affection.

DEDICATION

This thesis is dedicated to my father Suresh Dere and mother Padmavati Dere for their love and support and trust in me. Also I will like to mention my brother Bhushan Dere who is also my most trusted friend.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES:

CHAPTER 1. LITERATURE REVIEW

1.1 Switchgrass as a biofuel plant:

Switchgrass (*Panicum virgatum*) is a native grass in North America [1]. Switchgrass along with big bluestem (*Andropogon gerardii*) and Indiangrass (*Sorghastrum nutans*) dominate in prairies of America. It has been grown as forage plant. It has a deep rooting system hence it requires less water and nutrient input. It can grow on marginal land. It is also rotated with other crops as switchgrass prevents soil erosion. The switchgrass root biomass can reach 8 dry Mg/ha, which is an excellent carbon sink hence it is used for sequestration of land. Switchgrass is a C4 plant with Kranz anatomy. Hence it is energy efficient and fixes carbon more efficiently to produce biomass. Its yield of lignocellulosic biomass was estimated to be 14.2 tons per hectare [2]. It produces 540% more renewable energy than consumed. Hence, switchgrass is recognized as a major biofuel crop. First biofuel was obtained from corn. First generation biofuels are made from the sugars and vegetable oil found in arable crops, which can be easily extracted using conventional technology. In comparison, second generation biofuels are made from lignocellulosic biomass or woody crops, agricultural residues or waste, which makes it harder to extract the required fuel. Second generation of biofuel crops produce lignocellulosic biomass from non-arable crops like switchgrass, *Miscanthus*, poplar, and willow or from waste products like straws of cereals from farms or poultries. First generation biofuel are useful but limited in most cases: there is a threshold above which they cannot produce enough biofuel without threatening food supplies and biodiversity. Many first generation biofuels are dependent of subsidies and are not cost competitive with existing fossil fuels such as oil, and some of them produce only limited greenhouse gas emissions savings. Switchgrass is most preferred plant for biofuel production. *Miscanthus* is also developed as second-generation biofuel crop. *Miscanthus* can grow better than switchgrass and produce more vegetative biomass in northern part of the US. However, *Miscanthus* is a non-native and temperate plant with narrow genetic background [3].

1.2 Genetics, breeding and functional genomics resources of switchgrass:

The estimated tetraploid switchgrass (2n=4x=36) genome size is 3.32 pg /cell and represents two heterozygous genomes, it is approximately 10 times larger than the genome of *Ciona intestinalis (*160 Mbp) *[4].* Ploidy size variations in switchgrass (*Panicum virgatum*) have been revealed by flow cytometry [5]. Switchgrass is multigene self-incompatible with the S-Z type of incompatibility, which results in pre-fertilization abortion of self-crosses. Post fertilization abortion results from crosses of ecotypes differing in ploidy level [6]. Because of the self-incompatibility, switchgrass is highly heterozygous, and it is difficult to generate inbred parent lines for hybrid breeding [7, 8]. However, even though parents are heterozygous, it was demonstrated that progeny obtained from switchgrass crosses shows heterosis [9].

Switchgrass, along with maize, belongs to the tribe Paniceae of the subfamily Panicoidae, and it diverged about 23 million years ago. Based on their natural habitats, switchgrass plants are divided into lowland and upland ecotypes. Most lowland cultivars are tetraploid (2n=4X=36), whereas most of the upland cultivars are octoploid (2n=8x=72) with few being hexaploid. Lowland cultivars are disease resistant whereas upland cultivars are drought/ cold resistant. Upland and lowland ecotypes have been generally differentiated on the basis of plant phenotype: lowland plants are taller, have fewer and larger tillers, longer and wider leaves, thicker stems, and flower later than upland plants. Most lowland ecotypes also have a distinct blue coloring on stems and leaves due to a waxy bloom on the epidermis [5, 9].

The draft genome sequence of switchgrass was released recently (http://switchgrassgenomics.org/). Many functional genomics sources are being developed for switchgrass. Comprehensive Expressed Sequence Tags (ESTs) are currently available [10] with approximately 720,590 ESTs having been deposited in GenBank [10]. An intensive analysis of 61,585 ESTs identified 12,829 clusters; 7,532 singletons; and 27,329 different consensus sequences [10]. More than 2,817 microsatellite (or Simple Sequence Repeat, SSR) sequences have been identified from the EST database, which led to the design of 1,780 SSR primer pairs

for genetic linkage mapping[11]. As another effort, over 11.5 million high quality ESTs were recently generated by 454 pyrosequencing, three full-length enriched cDNA libraries were constructed with RNA from multiple AP13 tissues grown under optimal and stress conditions. About 100,000 clones were sequenced from both ends with the Sanger method and over 69,000 high quality reads were produced [12]. The Affymetrix cDNA microarray chip (Pvi_cDNAa520831) based on PviUT1.2 contains ~122,400 probe sets. This chip has a 11mm feature size, with 11 probes for each transcript without mismatch probes. This chip was developed at the Noble Foundation [13]. Recently, micro RNAs have been used as tools to control gene expression, many microRNA and their targets have been identified in switchgrass (*Panicum virgatum*) [14-17].

The genetic complexity of switchgrass largely contributes to the difficulty of direct application of genetics and genomic tools developed for model plants. Therefore, developments of suitable reverse functional genetics tools are important for annotating the biological functions of switchgrass genes controlling the biomass traits. Intensive switchgrass mutant collections are critical for functional analysis of plant genes but are currently missing in the switchgrass research community. Various approaches, including (1) chemical-induced mutagenesis, (2) radiation-mediated mutagenesis, (3) T-DNA/transposon insertion mutagenesis, and (4) T-DNA activation tagging mutagenesis, have been successfully employed for generating large mutant collections in model plant species [18-22]. The first three methods could generate “loss-of-function” mutations that disrupt a genetic pathway, while T-DNA activation tagging is based on the overexpression of a gene that may result in “gain-of-function” mutations [23]. Switchgrass is a tetraploid, and most of the genes are expected to act redundantly, and therefore loss-of-function mutant screening will be inefficient to identify genes. Switchgrass is self-incompatible, making it difficult to generate homozygous lines required for identifying mutations, which are often recessive genes. However, T-DNA activation tagging mutagenesis could result in dominant phenotypes and overcome the gene redundancy and heterozygosity issues in switchgrass. Therefore, T-DNA activation tagging should be the most efficient mutagenesis approach in switchgrass. The T-DNA activation tagging system was originally developed by designing a T-

DNA vector with four copies of an enhancer element from the cauliflower mosaic virus (CaMV) 35S promoter [24]. These enhancers could cause transcriptional activation of the nearby genes and result in dominant "gain-of-function" phenotypes. However, T-DNA activation system requires the *Agrobacterium*-mediated transformation process to generate a large T-DNA mutant population, with each mutant plant containing a T-DNA inserted in a different chromosomal location. For most plant species except Arabidopsis, a tedious tissue culture process is required for generating T-DNA transgenic lines. Therefore, it is difficult to generate large T-DNA mutant populations in plant species where tissue culture is inefficient. To harness the benefits of T-DNA insertion while circumventing the difficulties presented by tissue culture, the transposon-based activation tagging system was developed [21, 25-27]. Transposon tagging was originally developed based on the phenomenon of "jumping genes" in maize that was studied by Barbara McClintock in the late 1940s. In maize, the movement of a Dissociator (*Ds)* transposable element is controlled by the action of a companion Activator (*Ac*) element that encodes the transposase enzyme. When the *Ds* element is present in the genome without the *Ac* element, it is unable to move; this results in a stabilized insertion in the genome, allowing the analysis of changed gene expression. The *Ac*/*Ds* system has been intensively exploited for gene cloning and functional genomics studies in maize. The same mutagenesis approach was also implemented in plant species other than maize, where large mutant collections have been successfully developed [21, 25, 28-31]. Transposon activation tagging vectors were designed by cloning the *Ds* element along with a tetramer of CaMV 35S enhancer [21] or an outward strong promoter [25, 29] into a T-DNA construct. The T-DNA containing *Ds* element is then transformed into the plant genome. The *Ac* element can be cloned in the same T-DNA construct with *Ds* element (one-vector system) or in different T-DNAs (two-vector system) that requires genetic crossing to bring *Ac* and *Ds* elements into the same genome. The beauty of the transposon activation tagging system is that it only requires a limited number of primary transformants; more *Ds* insertions could be generated through propagation of the primary transgenic lines.

Model *plants* can be useful for functional genomics to help understand the biological functions of genes in crop plants with complicated genomes such as switchgrass [32, 33]. Arabidopsis has been used as model plant for three decades. However; Arabidopsis is a dicot whereas grasses are monocots hence Arabidopsis is not the perfect model to study genes in grass including switchgrass. Rice is another model plant that has been intensively studied for two decades. However; rice is a tropical crop plant that has adopted unique growth conditions therefore it may not be perfect for studying gene functions in other temperate grass species. To this end, two new plant species *Brachypodium distachyon* and *Setaria viridis* has been proposed as the model plants for grass species used in bioenergy research. *Brachypodium distachyon* has a small genome size of 355 Mbp, diploid accessions, small physical stature, self-fertility, short life cycle, and simple growth requirements. Its genome sequence was completed in 2010 [34]. All these characters make *Brachypodium distachyon* an attractive model plant for bioenergy research. However, *Brachypodium distachyon* is a C3 plant that may not be useful to study C4 plants including switchgrass. Hence*, Setaria viridis* is now developed as an model plant for C4 photosynthesis since it is closely related to switchgrass phylogenetically [35].

1.3 Switchgrass molecular markers and genetic linkage maps:

Detailed understanding of genome structure enables efficient marker-assisted selection (MAS) and can speed up this process considerably. Complete linkage maps are therefore required to both understand chromosome pairing and allow MAS. Switchgrass has extensive genetic diversity and potential for genetic improvements, but each cycle of phenotypic selection can take several years [2, 36].

Various attempts towards genetic mapping were performed with SDRF (single dose restriction fragments) [37]: RAPD (Random Amplified Polymorphic DNA) markers [38], RFLP (Restriction Fragment Length Polymorphism) markers [11], AFLP (Amplified Fragment Length Polymorphism) markers [39], SSR (Simple Sequence Repeat) makers [40, 41] , EST-SSR markers and SRAP (Sequence Related Amplified Polymorphism)[42].

Okada et al (2010) constructed the most comprehensive map for switchgrass so far. The linkage maps provide genetic evidence for disomic inheritance in lowland, tetraploid switchgrass. Gene-derived markers enabled a comparative analysis to sorghum, revealing syntenic relationships between the diploid sorghum genome and the tetraploid switchgrass subgenomes. Transmission ratio distortion and multilocus interactions were analyzed in detail to document their potential influence on map accuracy and map-based studies in switchgrass. The most recent data show that the linkage map size is 1645cM in the male parental line (cv. Alamo) and 1376 cM in the female parental line (cv. Kanlow), with an average marker spacing of 10 cM in both maps [43]. However, further mapping in other populations and more saturation with molecular markers are essential for effective marker-aided selection in switchgrass. Interestingly, it was reported that a few self - compatible switchgrass lines were identified by using SSR marker[7]. These self -compatible lines were further used for creating mapping populations by selfing. A genetic map was developed by using SSR markers for this population. It revealed a syntenic relationship between switchgrass and *Setaria viridis* [44].

1.4 Genetic improvement of monocots through transgenes

Agrobacterium has wide host range including many dicots and some monocots. After the first transgenic plants of dicotyledonous species were obtained by *Agrobacterium* mediated transformation in the early 1980s much effort was applied to extend the host range of *Agrobacterium* to monocotyledonous species [45]. A significant breakthrough occurred in 1993-1994 when highly regenerable explants of rice, mature embryos, or calluses derived from mature seeds were inoculated with disarmed *Agrobacterium* with selectable marker [46, 47]. This success was achieved by using a super binary vector that has extra copies of vir B, vir C, vir G .

Most of the economically important monocots have been relatively recalcitrant to genetic transformation compared to some monocots [48]. Transient gene expression [49] and stable genetic transformation of embryogenic calli in switchgrass have been reported [50-53] and stable genetic transformation have been reported [54, 55]. *Agrobacterium*-mediated transformation of plants genome has the advantage over biolistic method of transformation that it inserts few copies of transgene and thus it has less probability of causing transgene silencing [56, 57]. Great progress has been made for *Agrobacterium* mediated transformation of grass species, once considered recalcitrant such as rice [47], maize [58, 59], wheat [60], sorghum [61] and creeping bentgrass [62]. Relatively high transformation frequencies were reported in tall fescue [63, 64], *Brachypodium distachyon* [65] and maize [59]. However, most of the transformation efforts of monocots still suffer from efficiency, which is also a major obstacle for switchgrass transformation. The transformation efficiency varies a great deal depending on the explants and genotypes used in the experiment. Certain genotypes from Alamo were reportedly more competent for *Agrobacterium* mediated transformation. However, these materials seem to have gone to private sector and are no longer available to the public [55].

Optimizing callus induction and regeneration conditions are important for tissue culture of crop. Denchev et al 1996 [4]used 3% maltose MS media with 2-4 Dichlorophenoxy acetic acid (2-4 D) and Benzyl Adenine (Benzyl Adenine) for callus induction and regeneration in switchgrass by organogenesis and somatic embryogenesis. Alexandrovo et al 1996, [66] used BA (Benzyl Adenine) with 2,4 dichloro- phenoxy acetic acid for generating sterile explant with the same genotype from inflorescence.

Optimizing transformation and transient gene expression system is essential for genetic transformation of plant. Transformation was achieved in switchgrass by particle bombardment of somatic embryos with a GFP-BAR construct [67]. *Agrobacterium*-mediated transformation of the AGL1 strain with a binary vector coding for BAR and GUS was used for stable transformation

in switchgrass [52]. Xi et al, 2009 did transformation by using *Agrobacterium* strain EHA 105 with hygromycin and GUS marker[68]. They could achieve stable transformation as confirmed by southern blot and PCR and T1 progeny analysis. Recently, transformation is achieved by using base of seedling as ex-plant, which reduced time by minimizing tissue culture [69]. Also, transient gene expression systems and protoplast transient assay systems have been developed [49], [50]. Switchgrass polyubiquitine promoters and gateway compatible vectors have recently been developed for high throughput switchgrass transformation research [39, 70, 71].

Many transgenes are identified which can facilitate use of switchgrass for biofuel production. Bin et al (2011) identified 4-coumarate-coenzyme A ligase (4CL) as the key enzyme involved in lignin biosynthesis. Silencing 4CL could reduce lignin content in switchgrass biomass by 27% and make it more efficient in biofuel production [72]. Research done at the Samuel Noble Research Foundation and Oak Ridge National Lab found that COMT (caffeic acid 3-O methyl trasnferase) gene, CAD (cinnamyl alcohol dehydrogenase) gene expression alteration resulted in low lignin [73, 74]. Switchgrass contains two CAD genes which are involved in lignin formation [75] [76]. R2R3-MYB transcription factor PvMYB4 for improvement of sachharification capacity by decreasing recalcitrance of cell wall was also identified in switchgrass [77]. To enhance the profitability of switchgrass biomass, researchers also attempt to express value added products in switchgrass. For example, transgenic switchgrass expressing polyhydroxybutyrate for use as biodegradable plastic was recently reported [55].

1.5 Importance of biofuel:

Fossil fuel, which is obtained from the earth crust, causes methane and other greenhouse gas emission. While bioethanol produced via biological methods is renewable, eco-friendly and less toxic to humans, therefore it has garnished significant attention recently [78-81]. The USA consumes 25% of world energy and is largely dependent on the importation of oil from the gulf countries. The large import of fossil fuels puts a heavy burden on the US economy and the

fluctuation of the energy supply could seriously threaten national security[82]. In 2006 President Bush stated his Advanced Energy Initiative which seeks to reduce America's dependence on foreign sources of oil by replacing more than 75% of oil imports by 2025 [83]. The biomass derived biodiesel and other forms of biofuel are a major component to meet the challenge of the Advanced Energy Initiative [84]. In the future, most of the ethanol will come from cellulosic ethanol, which delivers up to ten times more energy than is required for its production. (*Source: RFA Resource Center, Ethanol Facts- Food vs. Fuel). The U.S. Department of Agriculture and Energy's *Billion Ton Study* found that we could grow adequate biomass feedstock to displace approximately 30% of the current gasoline consumption by 2030 on a sustainable basis with no conversion of U.S. croplands. It determined that 1.3 billion tons of U.S. biomass feedstock is potentially available for the production of biofuels. This is more than enough biomass to produce the mandated 3 billion gallons of cellulosic ethanol in 2015. The Energy Independence and Security Act that corresponded to the President's call to reduce U.S. gas consumption by 20% in 10 years require use of 36 billion gallons of biofuels by 2022. To reach that goal, we must use 15 billion gallons of cellulosic ethanol.

1.6 Cell wall and cellulose:

Most of the energy contained in biomass is stored in plant cell walls. Cellulose along with hemicellulose and lignin form the major components of the cell wall [85]. Cellulose is not digestible by animals and only some fungi and microbes and termites have the enzymes for digestion of it. Plant primary cell wall have cellulose microfibrils along with cross-linking hemicellulose and pectin. In secondary cell walls, cellulose constitutes 50% of the wall. Polysaccharides are synthesized in the golgi apparatus and are deposited to the wall surface by vesicles. Large hexameric complexes present in the plasma membrane synthesize cellulose microfibrils. The cellulose strands are extruded by a set of six cellulose synthase complexes, each responsible for six stands of glucan therefore the presence of 36 glucan chains in one microfibril. The repeating unit in cellulose is not glucose but the beta linked dimer of D-

glucopyranose, cellobiose. The common hexoses and pentoses in the plant cell wall are glucose, mannose, galactose, xylose, arabinose, and fructose.

1.7 Feedstock process and cellulase enzymes:

Extensive research has been completed on conversion of lignocellulosic material to ethanol in the last two decades. The bioconversion of lignocellulosic feedstock into ethanol has three major steps as pretreatment, hydrolysis and fermentation (fig.2). The first step pretreatment is required to remove lignin. Chemical or physical methods are applied for this purpose [82, 86, 87]. The second step is usually characterized by use of cellulase (endoglucanase, exoglucanase, cellulase) enzymes. The third step of fermentation is carried by use of yeast or bacteria. It results in production of ethanol ([86];[82, 87]).

Endoglucanase, exoglucanase, cellobiosidase are three enzymes that act together to degrade cellulose as shown in Figure 1 [88] [89]. Endoglucanase breaks internal bonds in cellulose microfibril. It cleaves the 1,4 beta bond between cellobiose units. Glucoses are connected to each other by 1,4 beta glucose bonding. Exoglucanase cleaves cellulose microfibril ends. Cellobiosidase cleaves the bond between two glucose units. Thus synchronization of all the three enzymes results in the breaking down of cellulose microfibrils. Endoglucanase enzyme is the key enzyme in this process and its cost is one of the barriers in the usage of fermentation in the industry for biofuel.

Plant bioenergy research has been focusing on reducing the cost of cellulase enzymes and increasing biomass yield by plant molecular farming [83, 90]. To make the biorefinery industry more profitable, future biorefinery operations need to first extract high-value chemicals already present in the biomass, such as fragrances, flavoring agents, food-related products, and high-value pharmaceuticals that provide health and medical benefits, the pretreated feedstock will then be processed for biofuel production. Biofuel crops will only be cost effective in the long

run if they are further developed transgenically to be more productive and have the right properties as fuels as well as residues that have value [91].

Endoglucanase used in this project is from *Hypocrea jecorina* (*Trichoderma ressi*), which is one of the most commonly used types of fungus for the endoglucanase enzyme[92] [89]. Logen Inc., which is word leader in technology of cellulosic ethanol production, uses *Hypocrea jecorina* derived enzyme in feedstock pretreatment. It has high temperature stability as well as it is one of the most widely studied enzymes. It is GH 12 type of enzyme. Glycoside Hydrolase family of enzyme catalyse hydrolysis of the glycoside linkage to release smaller sugars.

1.8 Aim :

To genetically modify switchgrass producing transgenic plants with the endoglucanase gene and test endoglucanase gene expression.

Long term goal of genetic transformation of switchgrass with endoglucanase gene is to prove that it is possible to co-produce proteins in plants which might be of pharmaceutical importance or that can be used for other industrial applications. Currently, plants are being used for production of many pharmaceutical products, an application called as molecular farming [93].

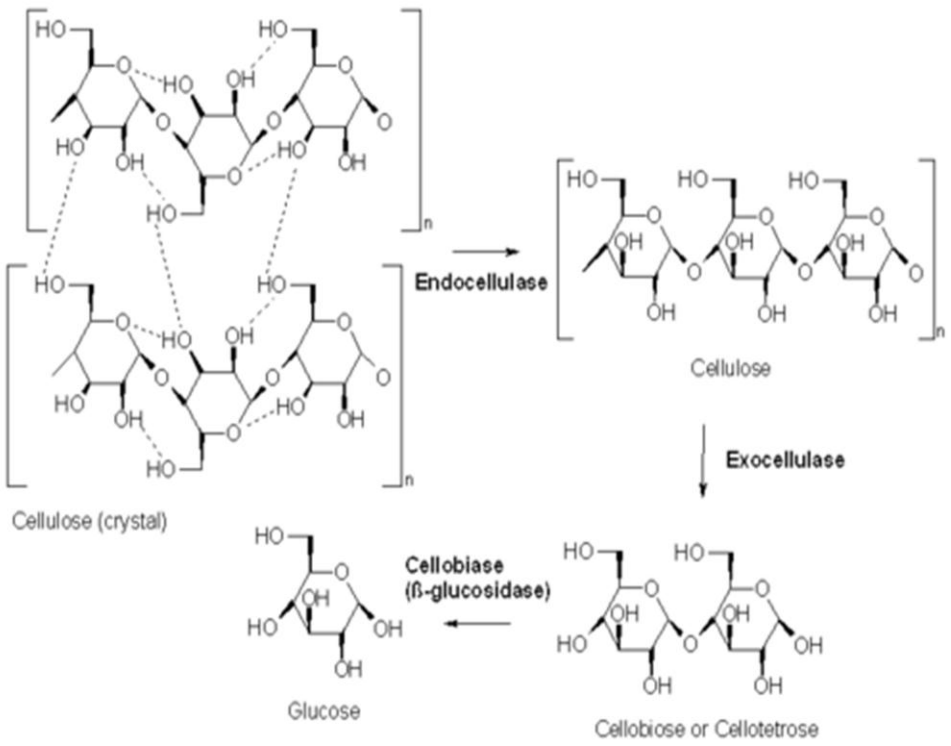

Figure 1 Cellulase enzymes: Endoglucanase, exocellulase and cellobiose are involved in enzyme hydrolysis of cellulose to glucose (http://upload.wikimedia.org/wikipedia/commons/b/be/Types_of_Cellulase2.png)

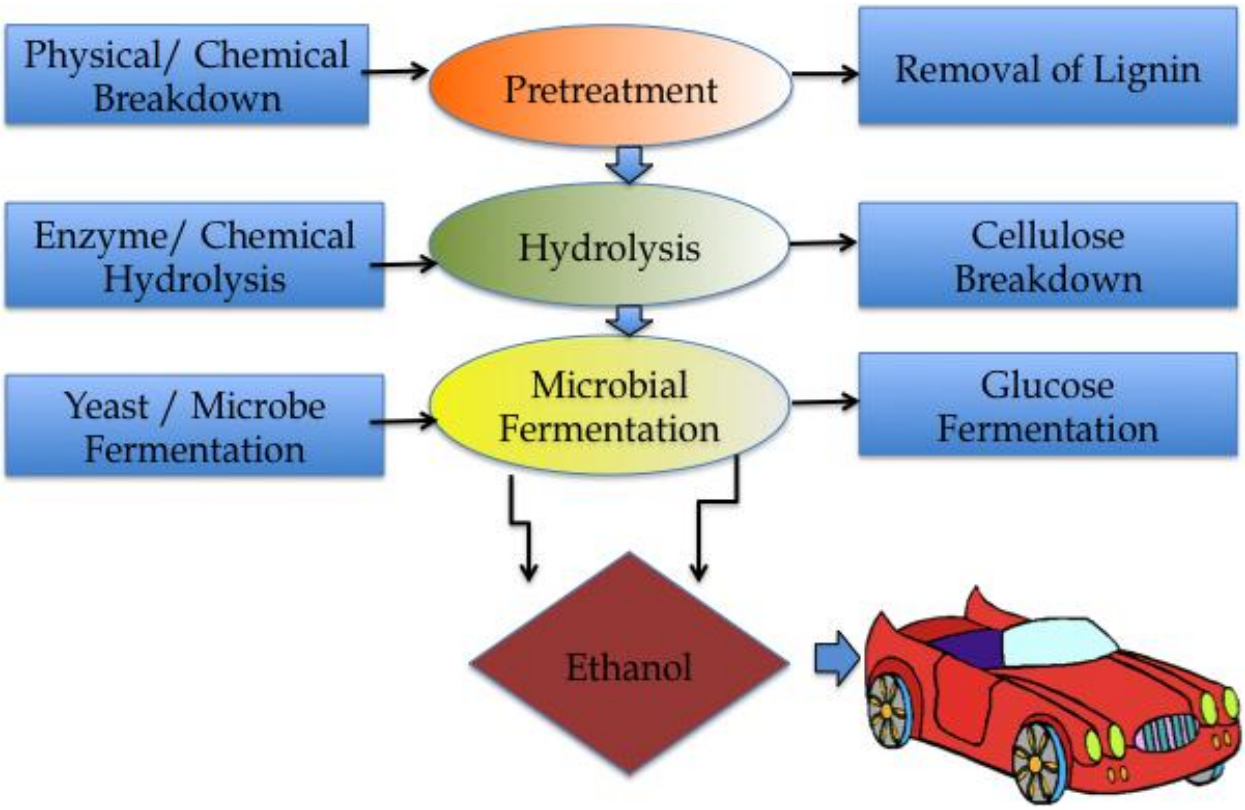

Figure 2 Overview of biorefinery processing: Switchgrass is processed in three steps as pretreatment, hydrolysis and microbial fermentation for production of ethanol.

CHAPTER 2: GENERATE TRANSGENIC SWITCHGRASS PLANTS EXPRESSING THE ENDO GLUCANASE GENE

2.1 Introduction

Switchgrass was recently recognized as a major biofuel crop because of its high yield of lignocellulosic biomass which was estimated to produces 540% more renewable energy than consumed [2]. Current plant bioenergy research has focused on reducing the cost of cellulase enzymes and increasing biomass yield by molecular plant breeding. The bioconversion of lignocellulosic feedstock into biofuels has three major steps including feedstock pretreatment, enzymatic hydrolysis and sugar fermentation [83, 90]. The enzymatic hydrolysis of cellulose using enzymes such as cellulase and cellobiohydrolase creates the fermentable sugar, glucose that is fermented to produce ethanol ([82, 86, 87]. It is possible to co-produce cellulase enzyme or other value added proteins in switchgrass plants which might reduce the cost of biofuel production and increase the profitability of the bioenergy industry [93].

Several reports have been published describing expression of the cellulases in transgenic maize, rice and tobacco etc. They could express proteins of cellulase[94, 95]. However, there are no reports about switchgrass transformation with cellulase genes. Previously, the cellulases have been targeted to various cellular compartments such as apoplast [95-98], cell wall[94], cytosol [82, 98, 99], endoplasmic reticulum[100], and chloroplast[96, 98, 100-104]. It has been found that targeting cellulase to the chloroplast is a better option, as protein accumulation in this compartment causes no effect on cell until feedstock treatment [94-98, 101, 105-107].

In this report, I constructed a vector for generating transgenic switchgrass plants expressing the endoglucanase from *Hypocrea jecorina* targeting to the chloroplast. Five independent transgenic switchgrass lines expressing the endoglucanase gene were developed. Our preliminary result suggests that the endoglucanase gene could be functionally expressed in

transgenic switchgrass plants. However, the gene expression level and the subcellular localization of endoglucanase need to be carefully controlled in order to minimize the toxic effect to plant growth.

2.2. Materials and Methods

2.2.1 Cloning of the endoglucanase gene

I used *Hypocrea jecorina*, which is the anamorph of mesophilic, filamentous fungus *Trichoderma ressi*. It can secrete large amount of cellulase enzymes. Largest ethanol producing biorefinerie in Canada used enzyme from *Hypocrea jecorina* for enzyme hydrolysis of feedstock. The endoglucanase gene sequence of *Hypocrea jecorina* was obtained from Genbank (AB003694). The codon preferences for switchgrass were characterized through analyzing the currently available ESTs by Genscript (>500k ESTs were deposited in GenBank). The codons usage of endoglucanase gene was optimized for expression based on the codon preferences of switchgrass genes. The optimized open reading frame (ORF) of the endoglucanase gene was synthesized by Genscript Inc (Piscataway, NJ). The N-terminal signal peptide of RuBisCO small subunit from rice was custom synthesized and fused to the N-terminal of the endoglucanase gene. A 9xHis epitope tag was added for protein purification on the C-terminal of the endoglucanase gene, which is intended for simplified protein purification to obtain pure protein for characterization The synthesized endoglucanase gene was subcloned into the Gateway compatible binary vector pVT1629 [108] that resulted in the transformation vector named as pEnG (Figure 3), in which the endoglucanase gene is under the regulation of the maize ubiquitin promoter.

2.2.2 Switchgrass tissue culture:

Mature seeds of switchgrass genotype HR8 (selected from cv. Alamo) were used for tissue culture as described previously [109]. In brief, the switchgrass seeds were rinsed with 60% Sulphuric acid for 30 minute for removing seed coats and with 100% bleach for 30 minute to kill any bacterial and fungal contamination. Afterwards, seeds were washed several times with sterilized water to remove traces of bleach. The sterilized seeds were dried in the tissue culture hood and plated on callus induction medium. The components of all tissue culture media are listed in Appendix 1. After two months incubation on callus induction medium, most of switchgrass seeds formed calli. Switchgrass has type II embryogenic callus with regeneration capacity. This callus was selected for subculturing. All chemicals used in tissue culture were purchased from Phytotechnology laboratories Inc (Kansas City, KS).

Hygromycin (Sigma Aldirich Inc.) was used as the selection reagent during tissue culture selection. Three concentrations: 50 mg/L, 60mg/L, 70 mg/L were tested for their effectiveness of selection. Non-transgenic calli were plated on medium supplemented with different concentrations of hygromycin for one month. Three replicas with 100 calli on each plate were kept at 29 °C in the dark in a growth chamber.

2.2.3 *Agrobacterium-mediated* switchgrass transformation:

Agrobacterium tumefaciens strain C58C1 was used for transformation. A glycerol stock of C58C1 containing the pEng. was streaked on a plate with 50mg/l kanamycin. Spatula with *Agrobacterium* was inoculated into 20 ml LB liquid culture with kanamycin antibiotic. After 24 h. *Agrobacterium* was centrifuged and resuspended in Infection medium to obtain OD600 of 0.6. After adding 200 µm acetosyringone, it was placed in dark for around 1-3 hours. *Agrobacterium* was poured into a magenta box. The magenta box cap was put onto the box. The box was put into a vacuum container. It was vacuumed at low vacuum pressure for 10 minutes. Vacuum container was brought to the sterile hood. Vacuum was released and

magenta box was taken out. Bacteria were dumped out and callus was placed onto a sterile paper towel. Callus was dried on sterile paper towel. Calli were transferred onto co-cultivation medium (appendix 1). It was sealed with 3M tape and placed in dark at RT (22-23^{0}C) for four days. Calli were transferred onto callus selection medium (Appendix 1) with appropriate antibiotics or selection agents. *Agrobacterium* transformation was repeated five times.

Selection of transformed calli was done on selection medium supplemented with 50 mg/l of Hygromycin B. (Sigma, Inc.) as the selection agent and augmentin 375 mg/l to inhibit *Agrobacterium*. Selection was carried out for a month. At this stage as there were problems with bacterial overgrowth, subculturing was done repeatedly and calli which used to show bacterial overgrowth were removed continuously at interval of half a day. Continuous subculturing was needed sometimes even at the interval of half a day. During this period, calli that were showing bacterial overgrowth were left out in a plate while others were transferred to the new medium. During selection non-transgenic calli were used as control to confirm the effectiveness of hygromycin in killing of non-transgenic calli.

Regeneration was done for a month on the regeneration medium with augmentin 375 mg/l and 50 mg/l hygromycin. During regeneration, subculturing was done after every two weeks. After a month calli started forming roots. At this stage only real transgenic plants survived. During regeneration non-transgenic calli were used as control again. True transgenics survived at this stage and they developed roots.

The transgenic plants verified by GUS assay were grown in the horticulture glasshouse at Virginia Tech, with temperatures set at 22-28^{0}C, night: day with a 12–14 h light regime. The plants were grown in Miracle-Gro Potting Mix (Miracle-Gro Lawn Products, Inc., Marysville, OH, USA) in pots and watered about twice a week. Wild type (WT) plants regenerated from non-transformed calluses were also grown in the same glasshouse under the same conditions for further use in crosses. Each transgenic line was multiplied by splitting tillers and maintained in

the glasshouse. The putative transgenic plants were crossed with wild type HR8 by bagging the flowering tillers in the greenhouse to form hybrid seeds.

2.2.4 GUS assay:

Leaves and roots of putative transgenic plants were used for a GUS assay to confirm the presence of the transformation construct in the putative transgenic plants [110]. In brief, leaves and roots were kept in 1mM GUS substrate along with phosphate buffer (50mM NaH2PO4, 1mMEDTA, 0.1% Triton, 0.5mM K3Fe (CN) 6 at pH 7) over-night at 37^0C. Leaves were treated with 70% ethanol first and then with 100% ethanol for two days so the chloroplasts were removed and then leaves could be observed under the microscope (Zeiss Axio Observer.A1 from Carl Zeiss, Thornwood, NY).

2.2.5 PCR analysis of transgenic plants:

Molecular analysis was carried out after genomic DNA extraction from transgenic plants [11]. Primers for the hygromycin gene were HYGRO_FOR 5 'CAAACTGTGATGGACGACACCG 3', HYGRO_REV 5'TATATGCTCAACACATGAGCG 3'. Primers for the endoglucanase gene along with signal peptide were Endoglucan SP For: 5'-caccATGGCCCCCTCCGTGATGGC-3’ and Endoglucan_ORF_Rev: 5’-CAGCTGGTTGATAGAGGCGGTC-3’ were used. Primers for endoglucanase gene amplification without signal peptide were Endoglucan_ORF_For: 5’-caccATGAAGTTCCTCCAGGTGCTCCCAGCT-3’ and Endoglucan_ORF_Rev: 5’-CAGCTGGTTGATAGAGGCGGTC. PCR reactions were prepared with EconoTaq 2X master mix + 100 pmol /µl primer forward + 100 pmol/µl primer reverse + DNA template 10 ng/µl + DD water 11 µl = Total 25 µl. The PCR program used had an initial denaturation step at 94^0 C for 2 minutes followed by 1 cycle of denaturation at 94^0 C for 30 sec, annealing of 56^0 C for 30 sec, extension 72^0 C for 1 min: 32 cycles, final extension 10 min. The PCR fragments were separated

in 1% agarose gel, stained in ethidium bromide staining solution, and visualized under UV light using the Gel-Document system (Bio-Rad, Hercules, CA).

2.2.6 Southern Blot

Genomic DNA was extracted from four endoglucanase plants for southern blot [70] analysis using a DNA fragment of the *HPTII* gene as a probe (Xu *et al*., 2011a). In brief, 10 µg of switchgrass genomic DNA was digested with restriction enzyme *Hin*dIII, which is absent in the *HPTII* gene. The digested DNA samples were electrophoresed in 0.8% agarose gel, and then transferred to a nylon membrane (Whatman Schleicher and Schuell, Keene, NH). After pre-hybridization in hybridization buffer [6xSSPE pH 7.0, 5x Denhardt Solution, and 0.5% sodium dodecyl sulfate (SDS)] containing 200 mg/ml of denatured Herring sperm DNA at 65^0 C for 3 h, hybridization with the probe was carried out at 68^0 C for 24 hrs. After hybridization, the filters were washed three times for 90 min each in a buffer (2x SSC, 0.1% SDS), at 68^0C, and exposed to X-ray film at -80^0C for 3 days.

2.2.7 RT-PCR analysis

RNA was extracted using Trizol reagent (Life Technologies Inc.) and DEPC treated pipettes, microcentrifuges and water. Briefly, 1 gm leaf tissue was homogenized in liquid nitrogen. It was incubated with Trizol reagent for 5 minutes. Then 200 µl chloroform was added. It was incubated at -20^0 C for 20 min. Then it was centrifuged at 12,000 rpm for 10 min. The aqueous phase was transferred to a new tube and precipitated with isopropyl alcohol. Precipitate was washed with 70% ethanol and air-dried. Then it was resuspended in DEPC treated water. Then RNA concentration was measured with a nanodrop (NanoDrop Technologies, Willington, DE) and 2 µl (500ng) was run on a gel for 10 minutes at 100V and stained for two minutes in EtBr. It was checked for presence of ribosomal RNA by observing separate three bands.

The first strand of cDNA was synthesized using random primer for RT from FINNZYMES DyNAmo cDNA Synthesis kit. Then the second strand was synthesized by using endoglucan SP For and endoglucan Rev. primer as described previously. RT-PCR products were run on gel along with g-DNA for comparison of product size. Due to the lack of intron the expected size for RT PCR was less (~ 750 bp) as compared to genomic DNA (~ 1000 bp).

2.2.8 Protein extraction:

Total soluble proteins (TSP) were extracted from leaf tissues as described by [95]. Briefly, 100 mg fresh leaf tissue was ground in the sodium acetate-grinding buffer (50 mM sodium acetate pH 5.5, 100 mM NaCl, 10% v/v glycerol, 0.5 mM ethylenediaminetetraacetic acid (disodium salt), 1 mM phenylmethylsulfonyl fluoride, 1mg/l aprotinin, 1 mg/l leupeptin, 1 mg/l pepstatin) at a ratio of 5 µl per mg of sample. Soluble extract was recovered from insoluble debris after centrifugation for 5 min. A solution of saturated ammonium sulfate was added to the extracts to achieve a final concentration of 2.7 M ammonium sulfate. After incubation on ice for 30 min, the resulting precipitate was recovered by centrifugation for 5 min. The ammonium sulfate pellet was resuspended in 5 µl of grinding buffer for each 2 µl of starting crude extract.

Amount of total protein in extract was measured on nanodrop (Thermo Scientific Inc.) at 280 nm as well as by Bradford assay (Quick Start Bradford Dyeing Agent, QuickStart BSA protein standard).

2.2.9 Western blot:

Western blot was done by using anti-His antibodies. We used a 12 % SDS PAGE gel as we expected endoglucanase protein size of 40 kDa and GUS protein of size 60 kDa. We loaded 30 µl of all five samples of putative transgenic along with a protein marker in the first lane. As

positive control protein extracts of GUS positive plant developed in our lab were loaded on the gel. The gel was run at 150 volts for 1.15 hr. Prehybridization was done with milk powder 2.5 % concentration in TBST buffer overnight at 4^{0} C. Primary hybridization was done with mouse 6-his monoclonal antibody from Covance Inc. 1 mg/ml (1000 or 15000 dilution) by keeping it on a shaker for two hours in TBST buffer. Afterwards antibody was washed with TBST buffer 4 times. Secondary hybridization was done with HRP conjugated goat anti-mouse secondary antibody from Covance Inc. at a 1:20,000 dilution in TBST buffer for 2 hr. After two hours, it was washed two or three times with TBST buffer for 10-15 minutes each time. SuperSignal West Pico Chemiluminiscnent substrate from Thermo scientific was used for hydrogen peroxide assay. Substrate and developer were mixed in 1:1 ratio to prepare a working solution. The blot was incubated for five minutes with working solution. Excess reagent was drained off and blot was exposed to X-ray film.

2.2.10 Chlorophyll content:

The chlorophyll content of transgenic plants along with WT controls was analyzed as described previously [111, 112]. In brief, 1 g leaf tissue was homogenized in 10 ml of 80 % acetone by using mortar and pestle. Centrifuging at 8000 rpm for 10 minutes cleared the homogenized samples. The supernatant was collected and was diluted six times to be in the detection range of spectrophotometer. It was measured for total chlorophyll content at 652 nm, chlorophyll A at 645 nm, chlorophyll B at 663 nm, a Beckman DU spectrophotometer was used. The chlorophyll content was calculated as Total chlorophyll= A652/34.5, Chlorophyll A= 0.02*A64, Chlorophyll B= 0.008 *A663. Three replicates were taken for each sample.

2.2.11 Phenotype evaluation of the putative transgenic plants:

Transgenic plants MD1, MD2 MD3, MD4, and MD5 were vegetatively propagated by tillers. Transgenic plants were measured for plant height, stem diameter (at the fourth internode of stem), tiller number and plant biomass. Three replicas were taken for each observation.

2.2.12 Statistical analysis

Data for all samples with at least three replicates were analyzed using the one way ANOVA and Tukey HSD test with software JMP version 9 (SAS Institute, Inc., Cary NC) for detecting significant difference between the transgenic and control plants.

2.2.13 Enzyme assay

A series of soluble dilutions ranging from 10^{1} 10^{6} were prepared and in a 96 well plate, 10 μl samples (representing 1-100 ng total soluble protein) were mixed with 100 μl reaction buffer (50mM Sodium acetate pH 5.5, 100mM NaCl, 0.5mM 4-methylumbelliferyl β-D Cellobioside. Plates were covered with adhesive lid and incubated at 65 c for 30 min. The reaction was stopped with the addition of the stop buffer (0.15M Glycine pH 10.0), and the fluorescence were read at 465 nm using spectrophotometer device at excitation wavelength of 360 nm. Background inflorescence contributed by switch –grass was subtracted and activity of each sample was calculated using a standard curve representing 4-160 pmol MU and compared to the activity of pure endoglucanase enzyme.

2.3. Results

2.3.1 Optimization of the protocol for switchgrass tissue culture and transformation:

Switchgrass transformation is still challenging for most laboratories. In this study, I tried to further optimize the switchgrass tissue culture and transformation protocol. Mature switchgrass seeds of cv. Alamo (HR8) were sterilized by using two different concentrations of bleach (Clorox). Our results suggested that the seed sterilization with 50% bleach was not able to give infection free callus. Seeds treated with 100 % bleach for 30 minutes were free of

microbial contamination. Seed germination on callus induction medium was not significantly affected by 100% bleach treatment (data not shown).

The sterilized switchgrass seeds were placed on callus induction medium supplemented with phytohormone 2,4- dichlorophenoxy acetic acid and benzyl adenine (Appendix 1: list of tissue culture medium used in this study). Embryogenic calli (type II) were observed within one month after placing the seeds on callus induction medium. As shown in figure 6 A, the type II callus is white, friable and compact. I also observed that continuous selection and subculture of type II callus was essential for successful switchgrass transformation, since only the white, hard and non-transcluscent embryogenic calli could be used for transformation.

2.3.2 Optimization of hygromycin concentrations:

Previous reports described different concentrations of hygromycin for selection of monocot transformation events [113]. In this study, three concentrations of hygromycin: 50 mg/L, 60mg/L, 70 mg/L, were tested for their effectiveness of selection. As shown in figure 4, it was observed that all three concentrations even after four weeks. Therefore I tested if the hygromycin selection was more effective during the callus regeneration stage. As shown in Figure 4, it was observed that all three concentrations of hygromycin could significantly inhibit calli regeneration, while 50 mg/l was the optimum. Hence we decided to use 50 mg/l in further transformations. As shown in figure 5, calli on control plates without hygromycin showed 100 % regeneration with small green leaves emerging from them.

2.3.3 Transformation with *Agrobacterium*:

For transformation, a modified procedure of Somleva et al (2002) was used. *Agrobacterium* transformation was done five times by using 100 calli each time. Three transformation experiments were successful. From these three successful transformations, the transformed calli were obtained at a frequency 1%, 2% and 1%, respectively. These transformed calli were selected on selection medium as well as regeneration medium. They had formed roots and

developed into plants. Two transformation experiments were unsuccessful due to bacterial overgrowth. Even repeated subculturring was not effective in controlling bacterial overgrowth.

2.3.4 GUS assay:

The GUS assay was done on leaves and roots for testing if the putative transgenic plants carried the transgene construct. Root and leaf samples required three days to get stained. This might have been due to weak expression of the GUS gene. The GUS gene that we had used has an intron, hence it is likely that the blue color was produced due to activity of the gene inserted in plants and not from exogenous bacteria. Roots give clearer staining as compared to leaves due to lack of pigmentation. As shown in Figure 7 transgenic plants gave distinct blue color staining, confirming integration of endoglucanase at the DNA level.

2.3.5 PCR analysis of transgenic plants:

DNA was extracted from plants growing in green house by using CTAB method. Transgenic plants with positive Gus staining were subject to PCR analysis by using primers specific to hygromycin resistance gene and the endoglucanase gene. The vector was used as positive control. As shown in figure 8, four transgenic lines showed presence of hygromycin resistance gene. Hygromycin gene in plasmid construct has given strong bands. Endoglucanase gene was further tested for four transgenic lines with wild type as negative control. As shown in figure. 9, the amplification product for endoglucanase gene was 1 kb, which we had predicted as the size of endoglucanase is 1 Kb. Therefore, results obtained from PCR analysis confirmed the presence of the transgene in four transgenic lines. It indicates that DNA is inserted in plant genome.

2.3.6 Southern blot analysis of transgenic plants:

To further confirm the integration of the transgene into the switchgrass genome, southern blotting was performed by using a hygromycin gene fragment as probe. The pEnG vector plasmid DNA was used as positive control. As shown in figure 10, we tested four transgenic plants and all plants showed positive hybridization with the hygromycin DNA probe, which suggests they all carry the transgene construct. We observed that the transgenic plants have one or two hybridization bands suggesting the presence of low numbers of transgene. MD2 shows a big band due to the incomplete digestion by the restriction enzyme.

2.3.7 RT-PCR:

To test if the endoglucanase gene is expressed in transgenic switchgrass plants, we did RT-PCR by using endoglucanase gene-specific primers. We successfully isolated RNA from all transgenic plants, which were growing in green house and before flowering stage. After confirming RNA integrity by three separate bands on a gel, it was used for RT-PCR analysis. As control PCR for checking successful RNA isolation, we performed RT-PCR with housekeeping gene i.e. actin gene primers (data not shown). Later on we did RT-PCR on all transgenic plants using endoglucanase primers. During this stage we had expected to observe cDNA products of size 750 bp. Endglucanase gene after removal of intron is expected to show band of 750 bp. To confirm the size difference on gel we had used genomic DNA along with RT PCR products for electrophoresis. As shown in figure 11, we had identified two RT PCR positive plants along with one RT-PCR negative plants. As can be seen in figure 11 there was a size difference in the genomic DNA and RT-PCR products. We tested the remaining two transgenic lines also by RT PCR. However, they were negative and hence the figure is not shown. As we had confirmed by RT-PCR that the transgenic lines MD2 and MD4 expressed the gene. We focused on these two lines and multiplied these two lines by splitting tillers for producing enough biomass.

2.3.8 Protein extraction:

To confirm the expression of endoglucanase in transgenic plants, we isolated total soluble protein by using the ammonium sulfate protein precipitation method. After doing Bradford test, protein content was analyzed by plotting on the graph of standard BSA. As shown in table 3, we could isolate 0.3 mg, 0.15 mg and 0.21 mg total soluble protein from 1 gram of fresh leaf tissue of wild type, MD2 and MD4 respectively. However, it was difficult to detect the endoglucanase enzyme in the total protein extract by western blot when using the anti-His antibody (data not shown), hence the yield of endoglucanase in the transgenic plants is not clear at this time.

2.3.9 Chlorophyll content:

In the vector construct that we used for transformation, the signal peptide of the RuBisCO small subunit was used for targeting endoglucanase to the plastid. Accumulation of heterologous protein in plastid could affect the biological function of the chloroplast. As observed in the figure 12, MD-2 and MD4 have light green leaves as compared to wild type leaves. Hence, we measured the chlorophyll content. We used HR8 (wild type) as control and we compared it with MD2 and MD4. As shown in table 2, transgenic MD-2 and MD-4 showed less than 50 % of total chlorophyll content as compared to wild type, which could explain the light green leaf color phenotype observed in figure 12. This would suggest that there is endoglucanase expression in chloroplasts. Graph in figure 13 shows that there is significant difference in the total chlorophyll content in MD2 and MD4 as compared to wild type plant.

2.3.10 Phenotype of transgenic plants:

Transgenic plants MD2 and MD4 show different phenotypes as compared to wild type. They have yellowish leaves as compared to wild type plant. These plants show delayed flowering and fewer branches on stem and dwarfness. MD2 has a more distinct phenotype with narrow

leaves and more branches, which gives plant bushy appearance. It was observed that stem length is showing significant difference between means of wild type and MD2, MD4 at $p<0.01$. Plant biomass for MD2 is significantly different than that of wild type at $p<0.01$.

2.3.11 Enzyme assay

We had done primary enzyme assay with crude protein extract and found that MD2 and MD4 are showing enzyme activity (data not shown). However, we are currently purifying protein and then we can calculate amount of endoglucanase protein expressed in total protein.

2.4.6 Discussion:

In this study, I generated five transgenic switchgrass plants expressing the fungal endoglucanase gene. The expression of the endoglucanase gene was confirmed in two lines by using GUS assay, RT-PCR, and phenotype observation. However, in this study, we failed to detect the endoglucanase protein by western blot, though we repeated western blot many times by using different concentrations of antibodies. However, none of the blotting experiments showed bands from blotting for either GUS or endoglucanase gene. In further improvement of the protocol we attempted increasing protein concentration by using ammonium sulfate for precipitation. Using ammonium sulfate precipitation for protein extraction was effective as we measured extracted protein by Bradford method. However we still could not detect any protein bands in the western blot. We were expecting two bands each for the GUS protein and the endoglucanase protein. We had previously confirmed by GUS assay that GUS is expressed in these plants. Hence, failure in detection of protein bands might be attributed to lack of specific antibody or lack of active his tag from the vector. To improve our ability to detect the endoglucanase protein, we are in process of expressing the endoglucanase in *E. coli* and the purified protein will be used as antigen to generate specific antibodies.

Previous reports showed the endoglucanase protein can be expressed as high as 0.1 %, 0.24%, 4.9 % or 5.2% of total soluble proteins [95, 99, 103, 114]. It was observed that targeting the protein to the chloroplast increases total protein amount according to Ziegelhoffer et al 2009 [101]. Most importantly, the active recombinant enzyme could be recovered very easily and efficiently from dried pant material. It was observed that RuBisCO small subunit signal peptide if used for chloroplast targeting, it can target the endoglucanase to leaves of tobacco plants[96, 115]. It was found that protein accumulated in chloroplast was more in amount as compared to protein targeted to vacuole or with native endoglucanase signal peptide. Similar results were obtained with sugarcane plants. It was found that different subcellular targeting signals have a substantial impact on the accumulation of leaves. Protein accumulation in chloroplast leaves was largest as compared to vacuolar or endoplasmic reticulum targeting [100] . As per experiments done in *Arabidopsis,* targeting endoglucanase to chloroplast by signal peptide for RuBisCO small subunit gives more endoglucanase protein than by using signal peptides of chlorophyll a/b binding or RuBisCO activase[103]. In this study, we were able to isolate the total soluble protein from the transgenic plants. Since we are not able to detect the endoglucanase protein by Western blot, we were not able to estimate the exact yield of endoglucanase in our transgenic plants. However, from crude protein used for enzyme assay we could conclude that protein is expressed in plant.

In this study, we targeted the endoglucanase to chloroplast by using the RuBisCO signal peptide (Figure 3). It was reported that expression and accumulation of high levels of endoglucanase in chloroplast has no toxic effect on plant growth [105, 115, 116]. However, in this study, we observed both transgenic plants to have yellowish color leaves with reduced chlorophyll content, which suggests that endoglucanase accumulation in switchgrass chloroplasts may still have toxic effect. It is not clear if switchgrass is specifically more sensitive to any fungal cellulase or because of the particular endoglucanase enzyme used in this study. In the future, we would like to use endoglucanase specific antibody to examine the subcellular localization of the transformed switchgrass plant cells.

We also observed that MD2 has a bushy phenotype with smaller stem size than the wild type controls (table 3); therefore, we suspect that in this transgenic plant, not all endoglucanase peptides have been targeted to the chloroplast and some might have leaked out in cytosol. As result of cellulase activity on these plants might have been negatively affected.

All transgenic plants have different expressions due to transgene this might be due to different insertion sites of transgene in genome. It is observed in southern blot that transgene insertion sites for each plant are different. MD2 had most distinct phenotype as compared to MD4 and other transgenics. It might be because as we observed in southern blot there are two copies of transgene in MD2 and other transgenic plants have only one.

In this study, we confirmed three putative transgenic plants at the DNA level but not at the mRNA level as shown by RT-PCR analysis (Figure 11). It is possible that the transgene was silenced during tissue culture and transformation selection. Gene silencing at the transcriptional level might be due to histone, which prevents access to genes by other transcriptional components. Hence multiple copy insertion causes silencing of integrated gene. It is also known that DNA methylation on the transgene has important roles in gene silencing. *Agrobacterium* could transfer T-DNAs into plant cells and get randomly integrated within the plant genome [117, 118]. When the transgene is inserted at certain locations with high heterochromatin, the transgene tend to be inactive. Therefore, it is also possible that the gene silencing was caused by the insertion location effect. It is also possible that the expression of endoglucanase is toxic to switchgrass development as shown in the two lines MD2 and MD4, therefore the expression of endoglucanase is inhibited during the tissue culture and plant development.

It was reported that the chemical azicidine could inhibit DNA methylation, and therefore reverse transgene silencing caused by DNA methylation[119]. In the future, it will be interesting

to test if we can reverse the gene silencing of the three non-expressed transgenic switchgrass lines in the late stage of plant development, which may minimize the toxic effect of endoglucanase.

It was reported that transgenic plants expressing endoglucanase could enhance the feedstock pretreatment [95]. In the future, we will test if the transgenic biomass of MD2 and MD4 is more efficient in releasing sugar during biofuel production.

In conclusion, we generated transgenic switchgrass plants expressing the endoglucanase that could be used in the biorefineries for ethanol production, though accumulation of the enzyme in chloroplast may have a negative effect on the plant development. Our results also suggest that it is possible to use switchgrass plants as production house for protein molecules of biological importance. Therefore more projects can be designed with the aim to produce bio-molecules as co-products of switchgrass biomass.

Acknowledgement: I thank Drs. Changhe Zhou, Bingyu Zhao, and Zhengxing Shen for constructing the plasmid vector pEnG used for switchgrass transformation and Dr. Qiang Cheng for help with RT-PCR.

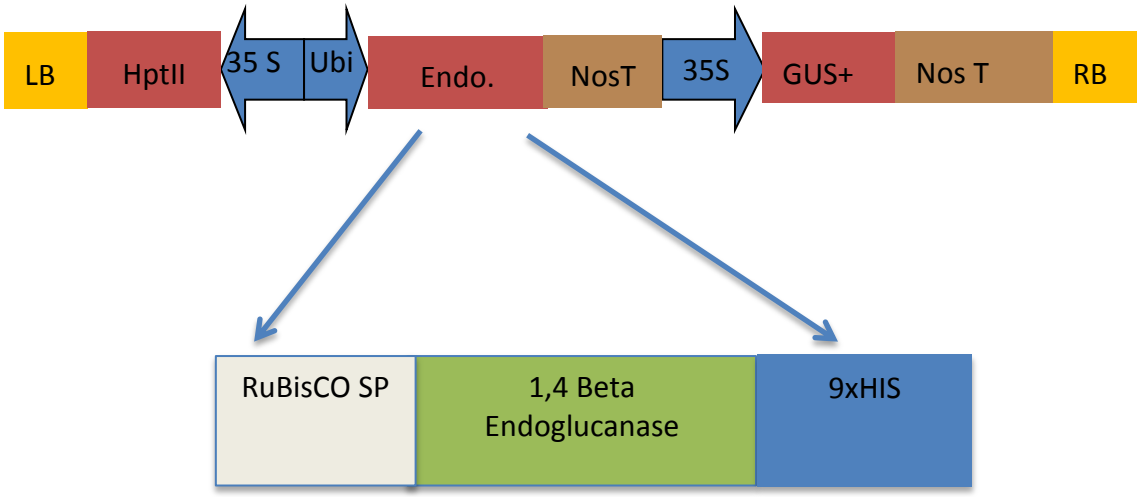

Figure 3 The diagram of the T-DNA region of transformation construct pEnG. T-DNA has *hpt* II as the antibiotic selection marker gene, GUS+ gene for reporter assay, and maize ubiquitin promoter for driving the expression of endoglucanase gene. Endoglucanase gene cassette has the RuBisCO signal peptide from rice at the N terminal, The codon optimized open reading frame of 1,4 Beta Endoglucanase gene from *Hypocrea jecorina* is fused with a C-terminal 9xHis tag for protein expression analysis. RB and LB: Right and left border of T-DNA respectively, 35S: CaM35S promoter, NosT: Nopaline synthase terminator.

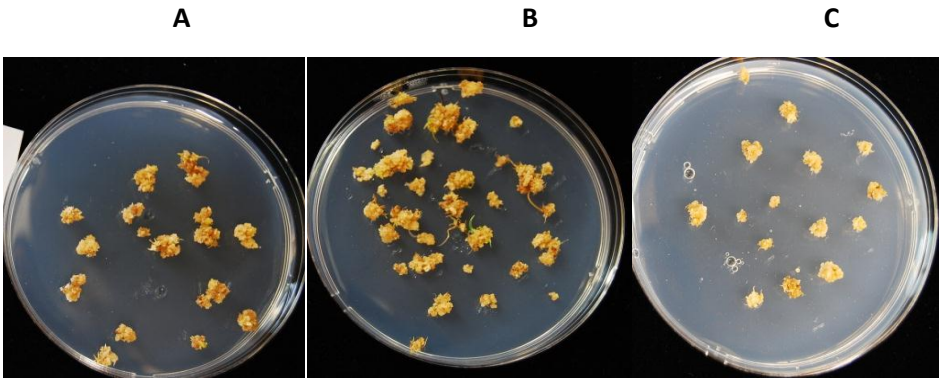

Figure 4 Optimize the hygromycin concentrations for effective selection against of the non-transgenic switchgrass callus on callus regeneration medium. (A): Callus growth was inhibited by 70 mg/L of hygromycin, (B): Callus growth was inhibited by 60 mg/L of hygromycin, (C): Callus growth was inhibited by 50 mg/L of hygromycin.

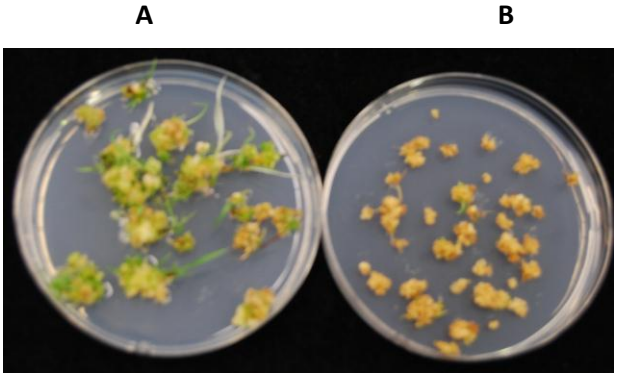

Figure 5 Optimize the Hygromycin concentrations for effective selection against of the non-transgenic switchgrass callus on callus regeneration medium. Non-transgenic calli were tested for hygromycin concentration optimization. **(**A): The non-transformed calli regenerated on the medium without antibiotic, (B): The non-transformed calli could not regenerate on media supplemented with 50mg/L of hygromycin.

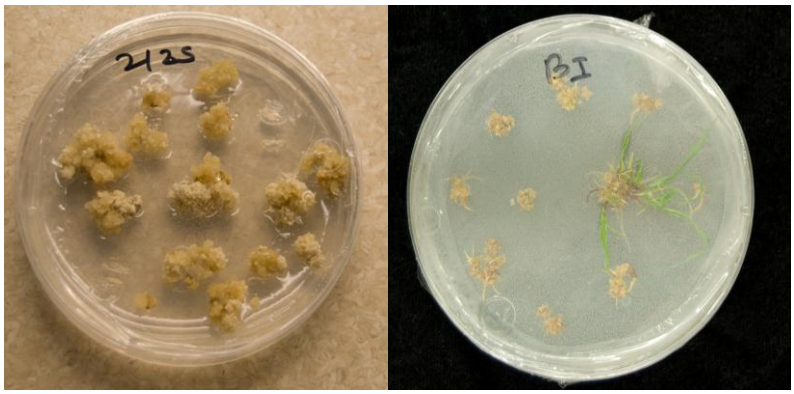

Figure 6 Selection of switchgrass callus for genetic transformation. (A): Embryogenic calli were selected on callus induction medium for two months that are ready for *Agrobacterium*-mediated transformation. (B): Transformed callus was regenerated on the regeneration medium supplemented with hygromycin (50 mg/L).

A **B**

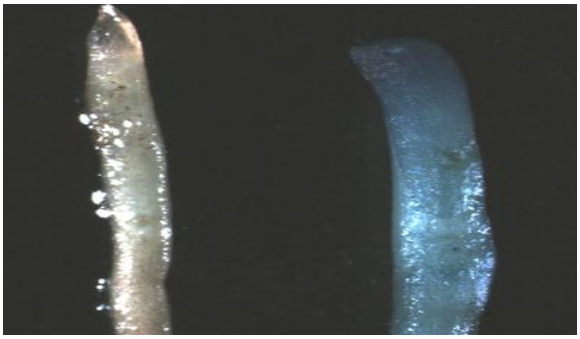

Figure 7 GUS Assay: The root sections of putative transgenic plant but not the control plant were stained in blue in GUS assay. Roots of control and putative transgenic plants growing greenhouse were incubated with X-gluc and phosphate buffer overnight. (A): Control Plant, (B): Transgenic plant.

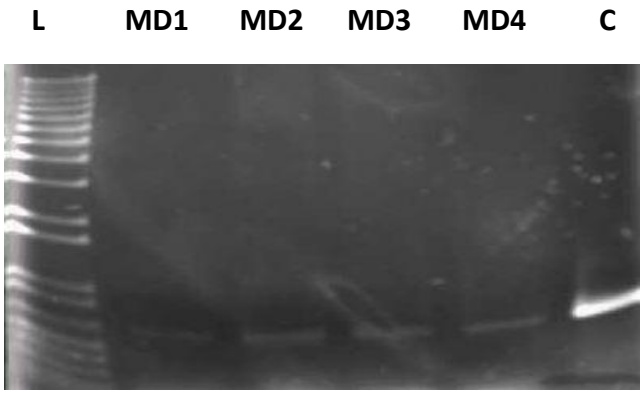

Figure 8 PCR amplification of the hygromycin resistance gene: Transgenic plants maintained in greenhouse were tested for the presence of hygromycin gene, which is the selection marker gene. L: 1 Kb+ DNA Ladder, MD1, MD2, MD3, MD4 are four independent transgenic lines, C: pEnG plasmid DNA as positive control. Part of the Hygromycin gene (650 bp) was amplified in transgenic as well as control plants

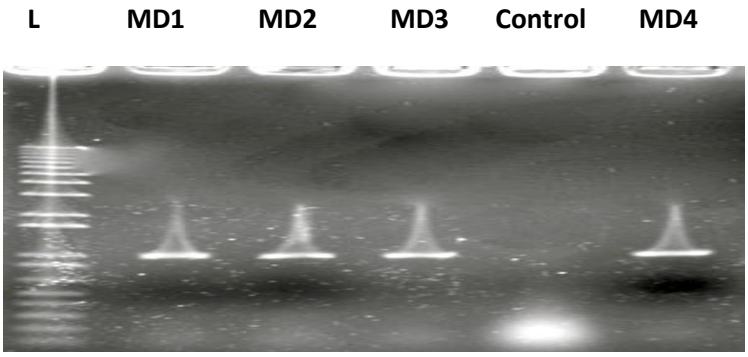

Figure 9 PCR analysis for the presence of endoglucanase gene in putative transgenic switchgrass plants: L: 1 Kb plus ladder, MD1, MD2, MD3, MD4 are four putative transgenic plants, Control: Wild type plant as negative control. DNA extracted from plants maintained in greenhouse was amplified by using endoglucanase gene specific primers and the expected PCR amplicon size is 1 kb.

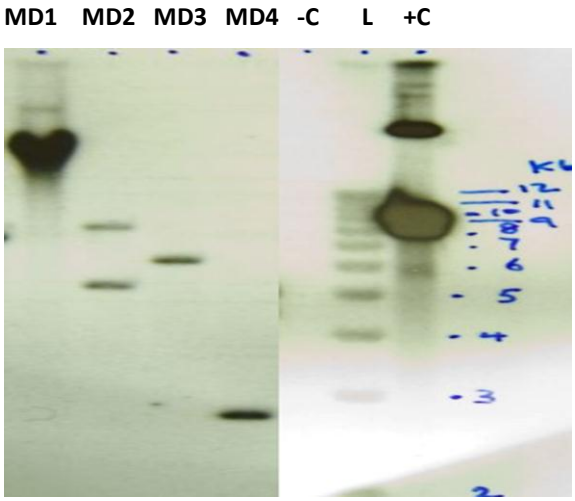

Figure 10 Southern blot analysis with hygromycin gene probe: Genomic DNAs extracted from green house plants was digested with *Hin*d III restriction enzyme. Restriction sites for *Hin*d III are absent in the Hygromycin gene. MD1, MD2, MD3, and MD4 are four putative transgenic plants transformed with the endoglucanase gene construct, -C: Water as negative control, L: 1 Kb DNA Ladder, +C: pCambia1305.2 plasmid DNA that carries a Hygromycin gene was used as positive control.

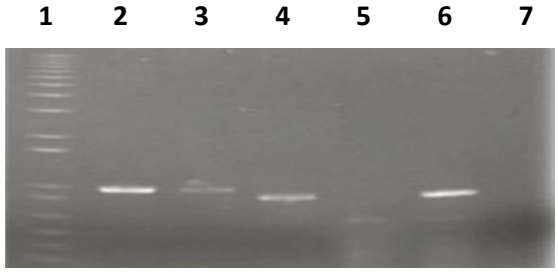

Figure 11 RT-PCR: The expression of endoglucanase gene in transgenic switchgrass plants was monitored by RT-PCR analysis by using gene specific primers flanking a native intron. Lane 1: 1 Kb+ DNA ladder, Lane 2 and 3: genomic DNAs of MD2 and MD4 plants were used as template for PCR amplification, Lane 4, 5, 6: cDNAs isolated from MD2, MD3 and MD4 plants respectively were used for PCR amplification. The endoglucanase gene amplified from genomic DNA is expected to be 1 Kb, while the RT-PCR product was expected to be 750 bp after removal of the native intron in the signal peptide.

Table 1 Protein concentrations of transgenic and control switchgrass plants: Total soluble protein was extracted in sodium acetate buffer and measured by Bradford assay with BSA standards. WT: wild type control plant, 4CL: a transgenic plant expression a RNAi silencing construct [108], MD2 and MD4 are two endoglucanase transgenic plants.

Plant Lines	**Total Soluble Protein** (mg protein/per gram of fresh leaf tissue)
WT	0.300
4CL	0.100
MD2	0.150
MD4	0.210

Figure 12 Chlorophyll content of leaves: WT: Wild Type. MD2, MD4: Transgenic plant. Transgenic leaves show pale yellow colored leaves as compared to wild type leaves because transgenic plant's chloroplast content is lowered due to targeting of signal peptide to the chloroplast.

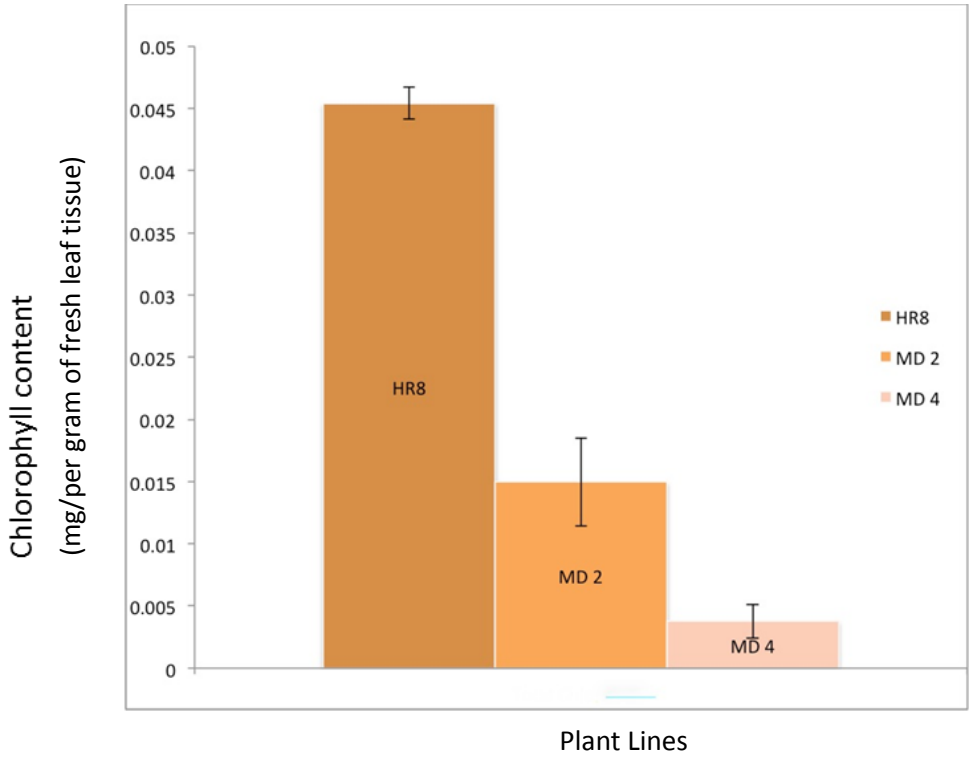

Figure 13 Total Chlorophyll content: Total chlorophyll from MD2, MD4 and HR8 (wild type control) was isolated in 80% acetone and were measured at 652 nm on spectrophotometer. The transgenic plants of MD2 and MD4 have reduced chlorophyll content than the HR8 control plants.

Table 2 Chlorophyll content: Chlorophyll content of two transgenic switchgrass lines (mg/ per gram of fresh leaf tissue) were compared to the wild type control plant (HR8). Three replicates for each sample were used.

Plant lines	**Chl a** (mg/g)	**Chl b** (mg/g)	**Total** (mg/g)
Wild Type	0.031±0.0005	0.012±0.0004	0.045±0.001
MD2	0.014±0.003**	0.004±0.001**	0.014±0.003**
MD4	0.003±0.001**	0.001±0.000**	0.003±0.001**

** Statistically significant differences ($p<0.01$) between wild type and transgenic plants.

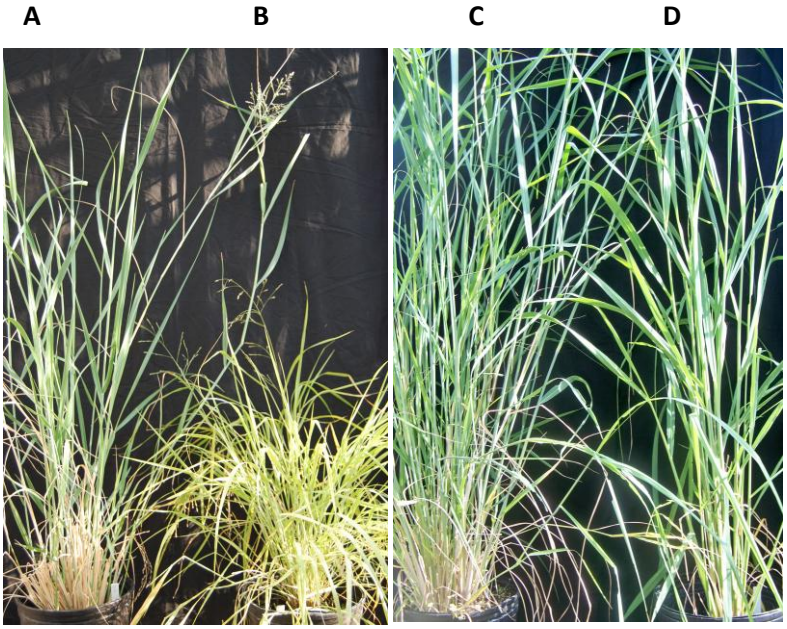

Figure 14 Phenotype of the endoglucanase transgenic plants: Transgenic plants showed negative effects on plant growth due to the presence of endoglucanase transgene. (A): Wild type control, (B): MD2, (C): Wild type control, (D): MD4. Transgenic plants are showing dwarf growth habit and pale yellow colored leaves.

Table 3 Phenotype characterization: Agronomic performances of five putative transgenic lines expressing endoglucanase gene was evaluated. All plants were grown and maintained in greenhouse under same growth conditions. Three plants of each line were used for the measurement. The averaged data was analyzed by one way ANOVA and compared by HSD Tukey test for significant difference.

Plant Name	**Plant Height** (cm)	**Stem Diameter** (cm)	**Tiller No.**
Wild Type	98.75± 10.65	0.35± 0.03	75±7
MD1	91.5 ± 8.675	0.41±0.02	85±5
MD2	45.57 ± 10.65**	0.25±0.03	65±7
MD3	59.375± 7.52**	0.39±0.02	97±5 **
MD4	52.5± 10.65**	0.37±0.03	85±7
MD5	113.25± 8.675	0.30±0.02	83±5

Phenotype characterization (continued from above)

Plant Name	**Leaf length** (cm)	**Leaf Breadth** (cm)	**Plant Biomass** (Kg)
Wild Type	18.5±3.75	0.69±0.09	0.19±0.04
MD1	16.83±3.03	0.77±0.09	0.09±0.03
MD2	11.50±3.75	0.31±0.09	0.03±0.04**
MD3	28.50±2.65	0.55±0.06	0.127±0.03
MD4	15.0±3.75	0.77±0.07	0.072±0.04
MD5	19.63±3.06	0.44±0.07	0.11±0.03

** Statistically significant differences ($p<0.01$) between wild type and transgenic plants.

CHAPTER 3. Activation tagging in switchgrass

3.1 Introduction

Switchgrass (*Panicum virgatum*) is one of the most promising biomass crops for the US, because it is a North American native, rich in natural diversity, and has adapted to a large portion of the continent [120]. Despite the increasing importance as a dedicated energy crop, we know little about the basic biology underlying the agronomic traits related to feedstock production. Compared to other crops or high-value forage plants, the genetic and genomic resources suitable for switchgrass molecular breeding, biotechnology, and molecular biology research are still lagging far behind.

As a self-incompatible and highly heterozygous perennial grass, switchgrass has two major ecotypes, lowland and upland, which are either tetraploid (*2n*=4*x*=36) or octoploid (*2n*=8*x*=72) with a few exceptions [36, 121, 122]. The genetic complexity of switchgrass largely contributes to the difficulty of direct application of genetic and genomic tools developed for model plants. Therefore, it is time to develop suitable reverse functional genetic tools that can help annotate the biological functions of switchgrass genes controlling biomass traits. Genetic mutants are critical for functional analysis of plant genes. Transposon tagging is a versatile tool for generating large collections of the gene knockouts in cereal plants like maize and rice, which are inefficient for T-DNA insertion mutagenesis. To harness the benefits of transposon insertion mutagenesis while circumventing the difficulties presented by the polyploid genome of switchgrass, researchers developed the transposon-based activation tagging system in different plant species [25, 28, 29, 123].

To identify the insertion sites of t-DNA or transposon in plant species with large genomes is challenging. TAIL (Thermally Asymmetric InterLaced) PCR was developed by Liu et al [124] that can be used for isolation of flanking sequence of T-DNA or transposon insertions. It involves use of thermally asymmetric PCR cycles for amplification of flanking sequence. The technique was further modified as hi (High efficiency) TAIL- PCR that improved its PCR

specificity and efficiency [125]. hi TAIL- PCR has been used to amplify both T-DNA and transposon activation tagging insertion sites.

In this study, I tried to test if we can establish the transposon activation tagging system in switchgrass. I also optimized the hi TAIL-PCR protocol and successfully identified the flanking DNA sequences of seven transposon activation mutants in switchgrass.

3.2 Material and Methods:

3.2.1 Plant material:

Eighteen transgenic plants, which were originally generated by Bin Xu by transformation of switchgrass genotype HR8 with a transposon activation tagging vector pSQ5 [109] were used for further transposon tagging analysis.

3.2.2 Vector for activation tagging:

The transposon activation tagging vector pSQ5 was previously developed for rice mutagenesis by Qu et al [123]. It has T-DNA carrying tetramer of CaM35S enhancer flanked by Ds inverted repeats. The CaM35S enhancer can cause overexpression of the adjacent genes when inserted in the plant genome. This vector also has the Ac gene in cis to a GFP marker gene, while RFP gene in cis with the Ds element. This design allows detecting the segregation of Ac and Ds elements in self or out crossed progenies.

3.2.3 GFP screening:

Leaves and roots of tranposants were observed for GFP or RFP signal under a fluorescent microscope (model SZXZ-RFL3, Olympus America, Melville, NY) with GFP long-pass, or RFP filters (Chroma Technology Corp., Rockingham, VT) and digital images were taken.

3.2.4 PCR analysis:

Plants were analyzed by PCR primers for presence of hygromycin. Primers for hygromycin were HYGRO_FOR 5 'CAAACTGTGATGGACGACACCG 3', HYGRO_REV 5'TATATGCTCAACACATGAGCG 3'

3.2.5 Southern blot:

Genomic DNA was extracted from pSQ5 tagged plants for southern blot analysis using a DNA fragment of the *HPTII* gene as a probe [109]. In brief, 10 µg of switchgrass genomic DNA was digested with restriction enzyme *Hin*dIII. Hind III enzyme restriction sites are absent in the *HPTII* gene. The digested DNA samples were electrophoresed in 0.8% agarose gel, and then transferred to a nylon membrane (Whatman Schleicher and Schuell, Keene, NH). After pre-hybridization in hybridization buffer [6xSSPE pH 7.0, 5x Denhardt Solution, and 0.5% sodium dodecyl sulfate (SDS)] containing 200 mg/ml of denatured Herring sperm DNA at 65^0 C for 3 h, hybridization with the probe was carried out at 68^0 C for 24 hrs. After hybridization, the filters were washed three times for 90 min each in a buffer (2 x SSC, 0.1% SDS), at 68^0C, and exposed to X-ray film at -80^0C for 3 days.

3.2.6 hi TAIL-PCR:

DNA concentration of 50 ng in working solution with hygromycin specific primers and LAD or Ac primers were used for PCR. TaKaRA kit with Taq polymerase, dNTP and PCR buffer was used with PCR cycles described by [125]. MyThermocycler by Biorad was used.

For R0 reactions Hyg0A and LAD 1-2 and LAD 1-4 were used. Water was used as negative control and pSQ5 plasmid was used as positive control. PCR products were diluted further with water in the ratio of 1:39 and 1 µl from this dilution was used for further PCR reactions. For R1 reaction Hyg1a and AC1 were used. PCR products were further diluted 1:10 with water and 1 µl from this reaction was used for further reactions. For the R2 reaction Hyg2a along with AC1 was used. Amplified products of R1 and R2 reactions along with 1Kb plus marker (Invitrogen) were run on the 0.8% agarose gel at 120 V for 30 minutes. It was stained with ethidium bromide. Amplified bands longer than 500 bp were supposed to have some flanking genomic sequence

and they were send for sequencing. Bands, which were less than 500 bp, were assayed to have only vector sequence and they were not analyzed further.

3.2.7 Gel purification and sequencing:

Bands were purified before sending for sequencing. Gel purification was done by using a kit from Epoch Life Science. Briefly, for each gram of gel band, binding buffer was added. Afterwards the gel was allowed to melt by incubating at 65^{0} C for 5-10 minutes. After the gel has completely dissolved it was loaded on a centrifugation column and spun at 13,000-rpm speed for 2 minutes, supernatant was discarded. Washing buffer is added to centrifugation column and it was spun at 13000 rpm for 2 minutes, supernatant was discarded. This step was repeated again. Afterwards, eluting buffer was added to column and it was allowed to elute for 2 minutes by keeping centrifuge column in microfuge tube. After 2 minutes, it was spun at 130000 rpm for 2 minutes to get complete recovery of elute in microfuge tube. After gel purification quantity of DNA in gel extract was measured by nanodrop at 280 nm. The amount of DNA was measured to be around 8-12 ng/ml. Then samples were sent for sequencing at the Core Lab of the Virginia Bioinformatics Institute.

3.2.8 Sequence annotation at NCBI:

DNASTAR Lasergene was used for alignment with vector sequence. Clustal W method was selected for alignment.

TDNA sequences of TAIL-PCR products were BLAST-searched against the GenBank sequences showing highest query coverage and lowest E value was selected as possible match.

3.3. Results:

3.3.1 GFP screening :

As shown in Figure 15, the T-DNA of pSQ5 carries both GFP and RFP genes that could be used as selection markers for identifying transformation events. In this study, 19 pSQ5 T0 transgenic plants were screened by observing GFP expression and PCR assay for detecting presence of transgenes. As shown in Figure 16, young roots and leaves of each putative transgenic plants were collected and observed under a fluorescent microscope. We screened 19 lines, where 13 lines showed strong GFP fluorescent signals. I used wild type control plants grown as negative control to correctly distinguish the presence of GFP signal. We also attempted to observe RFP signal by using the same method; however I failed to observe any RFP signal due to possible masking effect of autoflourescent signal caused by chlorophyll.

3.3.2 PCR analysis of T0 transgenic plants:

To further confirm transgenic plants previously selected on media with hygromycin, we used hygromycin gene specific primers to amplify the genomic DNA of all in-vivo pSQ5 lines. As shown in Figure 17, the genomic DNAs eleven pSQ5 lines could amplify the hygromycin gene fragment.

3.3.3 Southern blot:

Southern blot with a hygromycin probe was used to further confirm genome integration of transposon constructs and determine the copy number of T-DNAs. As shown in figure 18 seven out of eight transgenic lines showed presence of insert. pSQ5-25 does not show presence of any insert. pSQ5-32 and pSQ5-28 are showing presence of single insert. pSQ5-8, pSQ5-26, pSQ5-27, pSQ5-33 show two inserts. pSQ5-33 are show presence of three copies of insert. As we had previously screened plants with GFP marker gene, we know that many plants has Ac element. Presence of the Ac element indicates that it could activate jumping of the DS element.

Thus after transformation the Ds element can jump within plants many times. This probably gave rise to the multiple insertions in some plants.

3.3.4 TAIL-PCR:

As shown in figure 19, to identify the flanking DNA sequences of the T-DNA insertions, I used TAIL-PCR primers LAD1-2 and LAD 1-4 to amplify the transgenic genomic DNA. I tested all transgenic lines by using this primer combination. However I still could not get specific amplifications. So I used again another pair of primers namely LAD 1-1 and LAD 1-3. However, this primer combination could not amplify the products that were not previously amplified by LAD1-2 and LAD 1-4. In the future, products that were not amplified should be tested with new specific primers from either left border or right border. I found that amplified products, which were more than 500 bp in length, only could give some hits in NCBI search. Amplified products that were less than 500 bp gave vector sequence when blast search of NCBI Genbank.

3.3.5 Sequence annotation at NCBI :

As there is no comprehensive database for switchgrass, I used NCBI for query search. For some queries I got results from switchgrass C-DNA clones whereas other sequences showed alignment with other clones from the graminaceae family. R1 and R2 products gave the similar hits in NCBI search. It indicates efficiency of hi TAIL-PCR in giving clear products.

3.3.6 Phenotype characterization:

As shown in figure 20 there is phenotype difference in PSQ5 tagged plants and wild type plants. These differences are related to habit type, flowering timer and tiller number. Figure 20 is showing example of difference in tillers. PSQ5-1 is showing reduced number of tillers as compared to wild type.

3.4 Discussion:

To determine the functions of sequenced genes in an organism, gene is disrupted or mutation is generated in the genes and sequences are analyzed. However, site directed mutagenesis is not possible in all organisms. T-DNA insertion offers some advantages like stability of insertions. However, it is labor intensive for recalcitrant plants [126]. In traditional breeding chemical induced mutagenesis, radiation induced mutagenesis are tools that are used for mutations. However, these approaches are not applicable to switchgrass because switchgrass is recalcitrant for transformation and it is tetraploid so most of the genes are expected to act redundantly. However; T-DNA activation tagging is based on the overexpression of a gene and it can be applied to switchgrass.

Plants have some endogenous coloration due to plastids and hence the RFP signal was not observable. Due to lack of RFP, we could not prescreen for presence of Ds in the plants. We could not get amplification in all plants. So it is possible that due to activity of the tranposase Ds has already jumped. Activation tagging is useful as we can observe phenotype changes if plants are mutated. Also due to jumping of transposons we can get seeds with different mutations from single transformation. [127]. We can use these plants further for more activation tagging analysis. Plants with single insertion should be crossed with wild type to create seeds of T1 plants. Plants that gave hits in the switchgrass and graminaceae are useful for designing SNP marker for mapping in switchgrass crosses. If some mutation has occurred then we can develop allelic SNP markers and place it on the genetic map of switchgrass[123]. However, we need to get more flanking sequences from transposon tagging so that we can begin this mapping work. For this purpose we need to get more transposon-tagged mutants [128].

Table 4 Primary screening for transposons by observation of GFP under microscope and cloning of the flanking sequence:

Name of the Transgenic plants	GFP signal observed in roots	The flanking DNA sequence identified by TAIL-PCR
pSQ5-1	+	ND (no data)
pSQ5-2	+	ND
pSQ5-5	ND	*Panicum virgatum* clone PV_ABa006-K10
pSQ5-8	+	*Panicum virgatum* clone PV_ABa006-K1
pSQ5-25	-	*Fragaria vesca* clone VT_468_R T-DNA insertion flanking sequence
pSQ5-26	+	*Sorghum bicolor* hypothetical protein
pSQ5-27	+	*Panicum virgatum* clone PV_ABa107-B06
pSQ5-28	+	ND
pSQ5-29	-	*Panicum virgatum* clone PV_ABa051-K10
pSQ5-31	+	ND
pSQ5-32	-	*Panicum virgatum* clone PV_ABa107-B06
pSQ5-33	+	ND
pSQ5-34-1	+	ND
pSQ5-34-2	-	ND
pSQ5-36	+	ND
pSQ5-37	+	ND
pSQ5-38	+	ND
pSQ5-40	+	ND

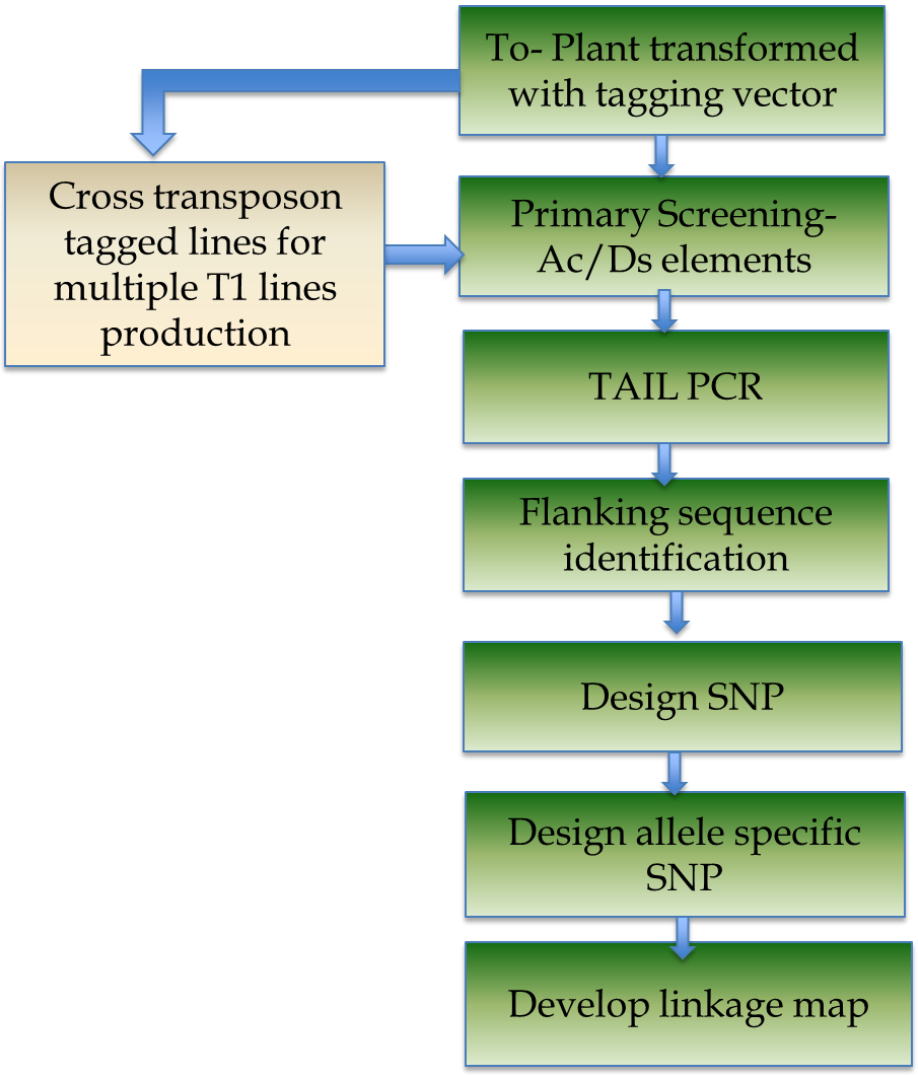

Figure 15 Overview of transposon tagging

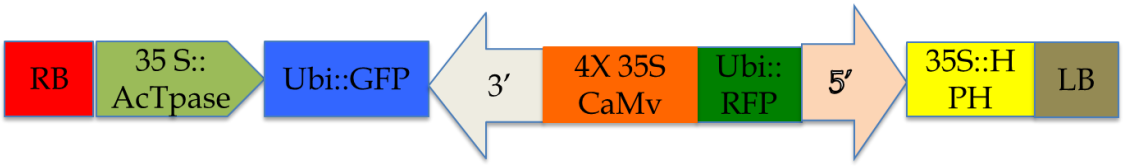

Figure 16 The T-DNA of the activation-tagging Ac-Ds vector pSQ5. RB and LB, Right and left borders of the T-DNA; Ubi: maize ubiquitin 1 promoter; GFP: GFP gene as a selection marker for Ac element; RFP: RFP (or DsRed) gene as selection marker for the Ds element; 4xEn: a tetramer of CaMV 35S enhancers; HPH : hygromycin phosphotransferase gene as plant transformation selection marker. The T-DNA is in the backbone of pCAMBIA-1305 [123].

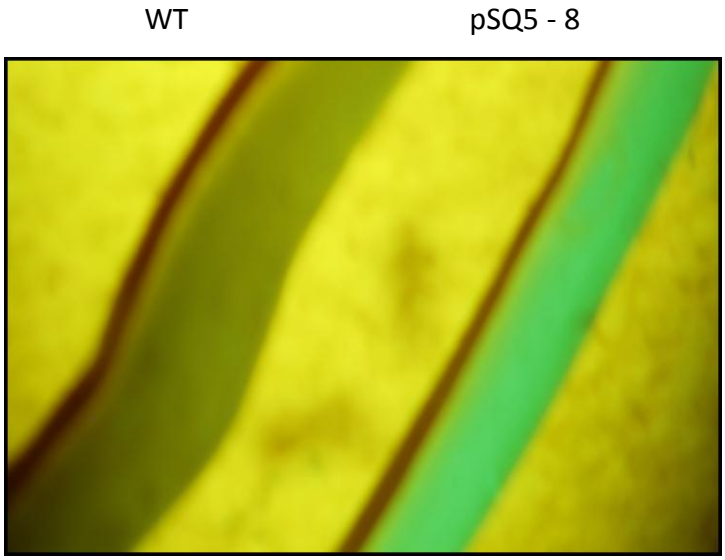

Figure 17 pSQ5 transgenic plants expressing GFP were observed under fluorescent microscope with yellow background in the dark. (WT): Young root section of wild type control plant showed no fluorescent signal, (pSQ5-8): Young root section of transgenic line showing strong GFP fluorescence in dark (20 x magnifications).

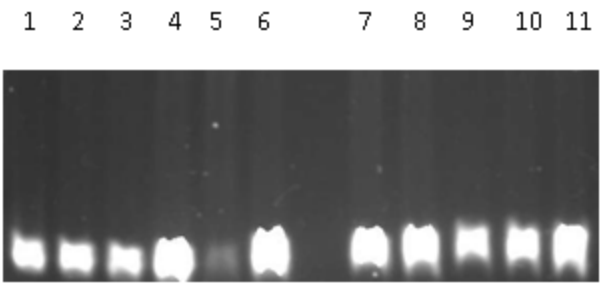

Figure 18 PCR analysis of genomic DNAs of pSQ5 plants for the presence of hygromycin gene which was selection marker: 1: pSQ5-8, 2:pSQ5-25, 3:pSQ5-27, 4:pSQ5-28, 5:pSQ5-29, 6:pSQ5-32, 7:pSQ5-33, 8:pSQ5-34, 9:pSQ5-36, 10:pSQ5-38, 11: pSQ5-40

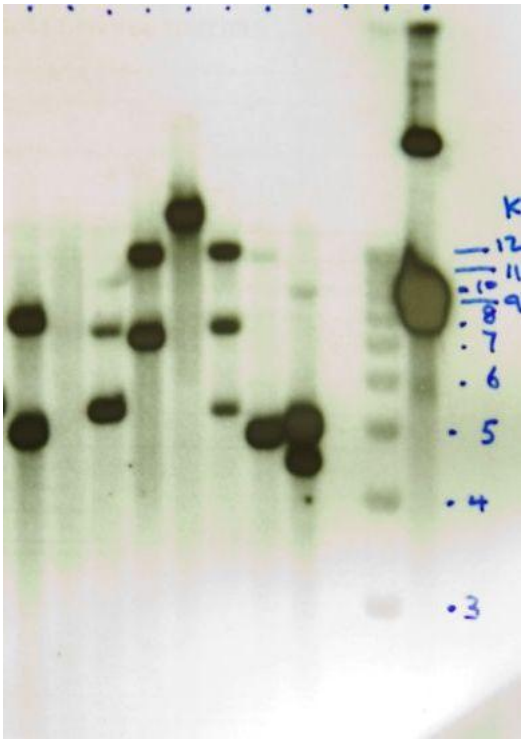

Figure 19 Southern blot analysis with hygromycin gene probe: Genomic DNA extracted from transgenic plants maintained in greenhouse was digested with *Hin*d III restriction enzyme. Restriction sites for hind III are absent in hygromycin gene. Lanes- 1:pSQ5-8, 2: pSQ5-25, 3: pSQ5-26, 4:pSQ5-27, 5:pSQ5-28, 6:pSQ5-29, 7:pSQ5-32, 8:pSQ5-33, L: Ladder, -C: water as negative control, C: pCambia1305 plasmid DNA that carry a Hygromycin gene was used as the positive control.

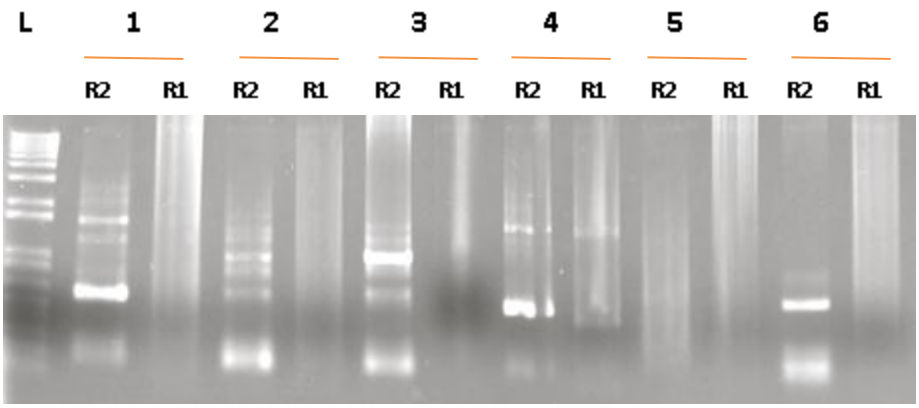

Figure 20 HI TAIL-PCR analysis: The flanking DNA sequences of selected pSQ5 lines were amplified by HI TAIL-PCR. (L): 1 Kb DNA marker. Six transgenic lines were used for TAIL-PCR: (1) pSQ5-8, (2) pSQ5-26, (3) pSQ5-28, (4) pSQ5-32, (5) pSQ5-33, (6) pSQ5-40. R1: primary hiTAIL-PCR amplification, R2: secondary Hi TAIL-PCR amplification.

Figure 21 Phenotype of one pSQ5 transgenic plant. (A): Wild type control plant grew normally in greenhouse, (B): Transgenic plant pSQ5-1 grown under the same conditions produced much less tillers in comparison to control plants.

REFERENCES

1. Moser LE VK, RF, Miller DA, Nelson J **Switchgrass, big bluestem and Indiangrass. In: Barnes** Ames: Iowa State Univ. Press; 1994.
2. McLaughlin S KL: **Development of switchgrass (Panicum virgatum) as a bioenergy feedstockin the United States**. *Biomass & Bioenergy* 2005, **28**(6):515-535.
3. Lewandowsky I SJ, Lindvall E, Christou M: **The development and current status of perrenial rhizomatous grasses as enegy crops in the US and Europe**. *Biomass & Bioenergy* 2003, **25**:335-361.
4. Hultquist S VK, Lee D, Arumuganathan K, Kaeppler S: **Chloroplast DNA and Nuclear DNA Content Variations among Cultivars of Switchgrass, Panicum virgatum L.** *Crop Sci* 1996, **36**:1049-1052
5. Costich D FB, Sheehan M, Casler M, Buckler E: **Genome-size Variation in Switchgrass (Panicum virgatum): Flow Cytometry and Cytology Reveal Rampant Aneuploidy**. *The Plant Genome* 2010, **3** (3):130-141.
6. Martinez-Reyna M VK: **Heterosis in Switchgrass: Spaced Plants**. *Crop Sci* 2008, **48**:1312-1322.
7. Liu L WY: **Identification of a Selfing Compatible Genotype and Mode of Inheritance in Switchgrass**. *Bioeng Res* 2012, **5**:305-313.
8. Young H HB, Anderton A, Lanzatella C, Tobias C: **Dihaploid Stocks of Switchgrass Isolated by a Screening Approach**. *Bioeng Res* 2010:305-313.
9. Vogel K MR: **Heterosis in switchgrass: Biomass yield in swards**. *Crop Sci* 2008, **48**:2159-2164.
10. Tobias CM, Sarath G, Twigg P, Lindquist E, Pangilinan J, Penning BW, Barry K, McCann MC, Carpita NC, Lazo GR: **Comparative Genomics in Switchgrass Using 61,585 High-Quality Expressed Sequence Tags**. *The Plant Genome* 2008, **1**(2):111-124.
11. Missaoui A PA, Bouton J: **Molecular markers for the classification of switchgrass (Panicum virgatum L.) germplasm and to assess genetic diversity in three synthetic switchgrass populations**. *Theoretical and applied genetics* 2006, **53**:1291-1302.
12. Wang Y ZX, Iyer N, Bryant D, Mockler T, Mahalingam R: **Exploring the switchgrass transcriptome using second- generation sequencing technology**. *PLoS* 2012, **7**:1-13.
13. Zhang J LY, Torres-Jerez I,Wang M, He J,Pennacchio C, Lindquist E, Yanbin Y, Wen-Chi Chou W-C,Shen H, Xu Y, Grimwood J,Schmutz J,Bartley L, Ronald P, Saha M, Dixon R, Tang Y, Udvardi M, Gilna P: **Development of resources for switchgrass functional genomics** *https://wwworaugov/gtl2011/abstracts/gilna_paul06pdf*.
14. Xie F FT, Baohong Zhang B,: **Identification and characterization of microRNAs and their targets in the bioenergy plant switchgrass (*Panicum virgatum*)**. *Planta* 2010, **232**:417–434.
15. Fu C SR, Zhou C, Shen H, Zhang J, Matts J,Wolf J, Mann D, Stewart N, Tang Y, Wang Z: **Overexpression of miR156 in switchgrass (Panicum virgatum L.) results in various morphological alterations and leads to improved biomass production**. *Plant Biotechnol J* 2012:1-10.
16. Matts J JG, Roe B, Sunkar R: **Identification of microRNAs and their targets in switchgrass, a model biofuel plant species**. *J Plant Physiology* 2010, **167**:896–904.
17. Chucka G TC, Sun L, Kraemerd F, Lic C, Dibblec D, Arorac R, Bragg J, Vogel J, Singh S, Simmons B, Pauly M, Hakea S: **Overexpression of the maize Corngrass1 microRNA prevents flowering, improves digestibility, and increases starch content of switchgrass**. *Proc Nat Acad Science USA* 2011, **108** (42):17550-17555

18. Li X, Song Y, Century K, Straight S, Ronald P, Dong X, Lassner M, Zhang Y: **A fast neutron deletion mutagenesis-based reverse genetics system for plants**. *The Plant journal : for cell and molecular biology* 2001, **27**(3):235-242.
19. Krysan PJ, Young JC, Sussman MR: **T-DNA as an insertional mutagen in Arabidopsis**. *Plant Cell* 1999, **11**(12):2283-2290.
20. Jeong DH, An S, Kang HG, Moon S, Han JJ, Park S, Lee HS, An K, An G: **T-DNA insertional mutagenesis for activation tagging in rice**. *Plant physiology* 2002, **130**(4):1636-1644.
21. Qu S, Desai A, Wing R, Sundaresan V: **A versatile transposon-based activation tag vector system for functional genomics in cereals and other monocot plants**. *Plant physiology* 2008, **146**(1):189-199.
22. Sikora P CA, Larsson M, Olsson J, Olsson O **Mutagenesis as a tool in plant genetics, functional genomics, and breeding**. In: *Int J Plant Genomics* Hindawi Publisher 2011: 1-13.
23. Weigel D AJ, Blazquez MA, Borevitz JO, Christensen SK, Fankhauser C, Ferrandiz C, Kardailsky I, Malancharuvil EJ, Neff MM et al: **Activation tagging in *Arabidopsis***. *Plant Physiol* 2000, **122**:1003-1013.
24. Odell JT, Nagy F, Chua NH: **Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter**. *Nature* 1985, **313**(6005):810-812.
25. Fladung M, Polak O: **Ac/Ds-transposon activation tagging in poplar: a powerful tool for gene discovery**. *BMC Genomics* 2012, **13**(61):doi:10.1186/1471-2164-1113-1161.
26. Scheinder A KT, Gigolashvili T, Mock HP, Sonnewald U, Simon R, Flugge UI, Werr W: **A transposon-based activation-tagging population in *Arabidopsis thaliana* (TAMARA) and its application in the identification of dominant developmental and metabolic mutations** *FEBES letters* 2005, **579**:4622-4628.
27. Ayliffe MA PM, Langridge P, Pryor AJ,: **A barley activation tagging system**. *Plant Mol Biol* 2007, **64**:329-347.
28. Schneider A, Kirch T, Gigolashvili T, Mock HP, Sonnewald U, Simon R, Flugge UI, Werr W: **A transposon-based activation-tagging population in Arabidopsis thaliana (TAMARA) and its application in the identification of dominant developmental and metabolic mutations**. *FEBS letters* 2005, **579**(21):4622-4628.
29. Ayliffe MA, Pallotta M, Langridge P, Pryor AJ: **A barley activation tagging system**. *Plant molecular biology* 2007, **64**(3):329-347.
30. Freeling M: **Plant transpososable elements and insertion sequences** *Annu Rev Plant Phys* 1984, **35**:277-298.
31. McClintock B: **The origin and behavior of mutable loci in maize**. *PNAS* 1950, **36**:344-355.
32. Lawrence CJ, Walbot V: **Translational genomics for bioenergy production from fuelstock grasses: maize as the model species**. *Plant Cell* 2007, **19**(7):2091-2094.
33. Bush D LJ: **Translational genomics for bioenergy production: There's room for more than one model** *Plant Cell* 2007, **19**:2971-2973.
34. Initiative TIB: **Genome sequencing and analysis of the model grass Brachypodium distachyon**. 2010, **463**:763-768
35. Eck J KE, zhu X, Jackson D, Goldachmldt A, Ewartwood K, Wnng L, Brutnell T: ***Setaria viridis*: A model for C4 photosynthesis**. *The plant cell* 2010, **22**.
36. JH B: **Molecular Breeding of switchgrass for use as a biofuel crops**. *Current Opinion Biotechnology* 2007, **17**:553-558
37. Missaoui A PA, Bouton J: **Investigation of genomic organization in switchgrass (*Panicum virgatum* L.) using DNA markers** *Theoretical and applied genetics* 2005, **110**(8):1372-1381.

38. Gunter LEB, A. S.; Ratnayeke, S.; Tuskan, G. A.; Wullschleger, S. D.: **Assessment of genetic similarity among 'Alamo' switchgrass seed lots using RAPD markers**. *Seed science and Tehcnology* 2003 **31**(3):681-689
39. Todd J WY, Wang Z, Samuels T: **Genetic diversity in tetraploid switchgrass revealed by AFLP marker polymorphisms**. *Genetics and Molecular Research* 2011, **10**(4):2976 - 2986.
40. Zhang Y ZJ, Jakubowski A,Price D, Acharya A, Wei Y,Brummer E , Kaeppler S, Casler M: **Post-glacial evolution of Panicum virgatum: centers of diversity and gene pools revealed by SSR markers and cpDNA sequences**. *Genetica* 2011, **2011**:933–948.
41. Huang S SX, Haselkorn R, Gornicki P: **Evolution of switchgrass (*Panicum virgatum* L.) based on sequences of the nuclear gene encoding plastid acetyl-CoA carboxylase**. *Plant Science* 2006, **164**:43-49.
42. Huang L BS, Zhanga X, Bales-Arcelob C, Xu B: **Genetic diversity of switchgrass and its relative species in Panicum genus using molecular markers**. *Biochem Syste Ecol* 2011, **39**: 685–693.
43. Okada M LC, Saha M,Bouton J, Rongling W,Tobias C: **Complete Switchgrass Genetic Maps Reveal Subgenome Collinearity, Preferential Pairing and Multilocus Interactions**. *Genetics* 2010, **185**:745–760.
44. Liu L WY, Wang Y, Samuels T: **A High-Density Simple Sequence Repeat-Based Genetic Linkage Map of Switchgrass**. *G3 Genes Genomes Genetics* 2012, **2**:357-370.
45. Gasser C.S FRT: **Genetically engineering plants for crop improvement**. *Science* 1989, **244**:1293-1299.
46. Chan M.T; Chang H.-H; Ho SLT, W.-F; Yu. S.-M: **Agrobacterium mediated production of trangenic rice plants expressing a chimeric alpha amylase promoter/ beta-glucurodinase gene**. *Plant Mol Biol* 1993, **22**:491-506.
47. Hiei Y OS, Komari T, Kumashiro T: **Eficient Transformation of rice(*Oryza sativa* L.) mediated by *Agrobaterium* and sequence analysis of the boundaries of the T-DNA**. *Plant Journal* 1994, **6**:271-282.
48. Birch RG: **Plant transfomation: problems and strategies for practical application**. *Annual review of plant biology* 1997, **48**:297-326.
49. Chen X ER, Baxter H, Kyle B, Han J, Agarwal S, Zale J: **A high throughput transient gene expression system for switchgrass (*Panicum virgtaum* L.) seedlings** *Biotechnology for biofuels* 2010, **3**(9).
50. Mazarei M A-AH, Rudis MR, Stewart CN Jr: **Protoplast isolation and transient gene expression in switchgrass, Panicum virgatum L**. *Biotechnology Journal* 2008, **3**( ).
51. Li JF PE, VonE, Von Arnim AG, Nebenfuhr A: **The fast technique: A simplified *Agrobacterium* based transformation method for transient gene expression analysis in seedlings in *Arabidopsis* and other plant species**. *Plant Methods* 2009, **5**(6).
52. Li R QR: **High throuhout *Agrobacterium* medaited switchgrass transformatio**. *Biomass & Bioenergy* 2010.
53. Burris J MD, Joyce B: **An improved tossue culture system for ebryogenic callus production and plant regenration in Switchgrass (*Panicum virgatum* L)**. *Bioeng Res* 2009(2):267-274.
54. Somleva M, Tomaszewski, Z., Conger, B.: **Agrobacterium-mediated transformation of switchgrass**. *Crop Sci* 2002, **42**:2080-2087.
55. Somleva MN SK, Beaulieu JJ, Peoples OP, Garrison BR, Paterson NA: **Production of polyhydroxubutyrate in swtichgrass, a value added co-product in an imp0ortant lignocellulosic biomass crop**. *Plant Biotehnology Journal* 2008, **6**:663-678.
56. Wang Z-Y GY: **Recent advances in genetic transformation of forage and turf grasses**. *In Vitro Cell Dev Biol, Plant* 2006(42):1-18

57. Repllin A BM, Jauhar PP, Chibbar RN,: **Genetic enrichment of cereal crops via alien gene transfer: New Challenges**. *Plant Tissue Culture Organ Cult* 2001, **64**:159-183.
58. Ishida Y SS, Ohta S, Hiei Y, Komari T, Kumashiro T: **High efficiency transformation of maize (Zea mays L.) mediated by abrobacterium tmefaciens**. *Nat Biotechnology* 1996, **14**:745-750.
59. Zhao ZY GW, Cai TS, Tagliani L, Hondrd D, Bond D et al: **High throughput genetic transformation mediated by *Agrobacterium tumefaciens* in maize**. *Mol Breed* 2001, **8**:323-333.
60. Cheng M HT, Layton J, Liu CN, FRy JE: **Dessication of plant tissue pot agrobacterium enhances T-DNA delivery and incrase stable transformation efficiency in wheat**. *In Vitro Cell Dev Biol, Plant* 2003, **39**:595-604.
61. Zhao ZY CT, Tagliani L, Miller M, Wang N, Pang H D et al: **Agrobacterium mediated sorghum transformation**. *Plant Mol Biol* 2000, **44**:789-798.
62. Yu TT SD, Liang GH, Trick HN, Huang B: **Agrobacterium mediated transformation of creeping bent grass using GFP as a reporter gene**. *Hereditas* 2000, **133**:229- 233.
63. Dong S QR: **High frequency transformation of tall fescue with *Agrobacterium tumefaciens***. *Plant Science* 2005, **168**:1453-1458.
64. Wang ZY GY: ***Agrobacterium* mediated high efficiency transformation of tall fescue (Festuca arundinaceae)**. *J Plant Physiology* 2005, **162**:103-113.
65. Pacurar DI HT, Kristian K, Lenk NI, : **A high though-put for *Agrobacterium* mediated transform sustem for the grass model species *Bracypodium distachyon* L**. *Transgenic Res* 2008, **17**:965-975.
66. Alexandrova KS, Denchev PD, Conger BV: **In vitro development of inflorescences from switchgrass nodal segments**. *Crop Sci* 1996, **36**(1):175-178.
67. Saski C LZ, Feltus F, Luo H: **New genomic resources for switchgrass: a BAC library and comparative analysis of homoeologous genomic regions harboring bioenergy traits**. *BMC Genomics* 2011, **12**:369.
68. Xi Y FC, Ge Y, Nandakumar R , Hiroshi Hisano H, Bouton J, Wang Z: **Agrobacterium-Mediated Transformation of Switchgrass and Inheritance of the Transgenes**. *Bioeng Res* 2009, **2**:275–283.
69. Song G WA, Hancock J: **Factors influencing Agrobacterium-mediated transformation of switchgrass cultivars**. *Plant Cell Tiss Organ Cult* 2012, **108**: 108.
70. Mann D KZ, Liu W, Joyce B, Percifield R, Hawkins J,LaFayette P, Artelt B, Burris J, Mazarei M, Bennetzen J, Parrott W, Stewart C: **Switchgrass (Panicum virgatum L.) polyubiquitin gene (PvUbi1 and PvUbi2) promoters for use in plant transformation**. *BMC Biotechnology* 2011, **11**.
71. Mann D LP, Abercrombie L, King Z, Mazarei M, Halter M, Poovaiah C,Baxter H, Shen H, Dixon R, Parrot W, Stewart C,: **Gateway-compatible vectors for high-throughput gene functional analysis in switchgrass (*Panicum virgatum* L.) and other monocot species**. *Plant Biotechnol J* 2012, **10**(2):226-236.
72. Xu B, Escamilla-Trevino LL, Sathitsuksanoh N, Shen Z, Shen H, Zhang YH, Dixon RA, Zhao B: **Silencing of 4-coumarate:coenzyme A ligase in switchgrass leads to reduced lignin content and improved fermentable sugar yields for biofuel production**. *New Phytol* 2011, **192**(3):611-625.
73. Fu C XX, Xi Y, Ge Y,Chen F , Bouton J , Dixon R, Wang Z: **Downregulation of Cinnamyl Alcohol Dehydrogenase (CAD) Leads to Improved Saccharification Efficiency in Switchgrass**. *Bioeng Res* 2011, **4**:153–164.
74. Fu C MJ, Xiaoa X, Gea Y, Hamiltonb C, Rodriguez M,Chenc F, Fostonc M, Ragauskasc A, Boutona J, Dixonc R, Wanga Z: **Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass**. *Proc Nat Acad Science USA* 2011, **108**:3803-3808.
75. Saathoff A TC, Sattler S, Haas E, Twigg P, Sarath G: **Switchgrass Contains Two Cinnamyl Alcohol Dehydrogenases Involved in Lignin Formation**. *Bioeng Res* 2011, **4**:120–133.

76. Saathoff A SG, Chow E, Dien B, Tobias C: **Downregulation of Cinnamyl-Alcohol Dehydrogenase in Switchgrass by RNA Silencing Results in Enhanced Glucose Release after Cellulase Treatment**. *PLoS One* 2011, **6**(1).
77. Shen H HX, Poovaiah C, Wuddineh W, Ma J, Mann D, Wang H, Jackson L, Tang Y, Stewart N, Chen F, Dixon R: **Functional characterization of the switchgrass (*Panicum virgatum*) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks**. *New Phytologist* 2012, **193**:121–136.
78. Greene N: **Growing Energy : How Biofuels Can Help End America's Oil Dependence** *Natural Resources Defense Council* 2004.
79. Wang M SC, Santini D: **Effects of Ethanol Use on Fuel-cycle Energy and Greenhouse Gas Emisssions**. *Argonne National Laboratory* 1999.
80. Sanchez C: **Lignocellulosic resudues: biodegradation and bioconversion by fungi**. *Biotechnology Advt* 2009, **27**:185-194.
81. Lin Y TS: **Ethanol fermentation from biomass resouces : current state and future prospects**. *Appl Microbiol Biotechnology* 2006, **69**:327-342.
82. Sun Ye CJ: **Hydrolysis of lignocellulosic material for ethanol production**. *Bioresouce technology* 2002, **83**:1-11.
83. Ragauskas A WC, Davison B,Britovsek G,, JCairney J EC, Frederick W, Hallett J, Leak D,, Liotta C MJ, Murphy R, Templer R,, Tschaplinski T: **The Path Forward for Biofuels and Biomaterials**. *Science* 2006, **311**:484-489.
84. Lynd LR CJ, Nicholas RJ, Wyman CE,: **Fuel ethanol from cellulosic biomass**. *Science* 1991, **251**:1318-1323.
85. Brett C WK: **Physiology and biochemistry of plant cell walls** London Chapman and Hall 1996.
86. Keshwani D.R. CJJ: **Switchgrass for bioethanol and other value added application: a review**. *Bioresouce technology* 2009, **100**:1515-1523.
87. Mosier N WC, Dale B, Elander R, Lee Y.Y, Holtzapple M, Ladisch M: **Features of promising technollogies or pretreatment of lignocellulosic biomass**. *Bioresouce technology* 2005, **96**:673-686.
88. Teeri T: **Cryataline cellulose degradation: new insight into the function of cellobiohydrolases**. *Trends Biotechnol* 1997, **15**:160-173.
89. Tomme P WR, Giles N, : **Cellulose hydrolysis by bacteria and fungi**. *Adv Microbe Physiology* 1995, **37**:1-81.
90. Sticklen M: **Plant Genetic Engineering to improve biomass characteristics for biofuels** *Current Opinion Biotechnology* 2006, **17**:315-319.
91. Gressel J: **Transgenic are imperative for biofuel crops**. *Plant Science* 2008, **174**:246-263.
92. Henrissat B DB, Viet C, Schulein C,: **Synergism of cellulases from *Trichosrma ressi* for degradation of cellulose**. *Nature Biotechnology* 1985, **3**:722-726.
93. Kusnadi AR NZ, Howard JA: **Production of recombinent proteins in trabsgenic plants: practical considerations.** . *Biotechnol Bioeng* 1997, **56**:473-484.
94. Brunckey R SM, Vinzant T, Himmel M, Lee D, Blaylock M, Decker S: **In planta expression of A. cellulolyticus Cel5A endocellulase reduces cell wall recalcitrance in tobacco and maize** *Biotechnology for biofuels* 2011, **4**(1).
95. Oraby H VB, Dale B, Ahmad R, Ransom C, Oehmke J, Sticklen M: **Enhanced conversion of plant biomass into glucose using transgenic rice-produced endoglucanase for cellulosic ethanol**. *Transgenic Res* 2007, **16**:739-749.
96. Dai Z HB, Anderson DB, Thomas SR: **Improved plant-based production of E1 endoglucanase using potato: expression optimization and tissue targeting.** 2000(6)::277–285.

97. Ziegler MT TS, Danna KJ: **Accumulation of a thermostable endo-1,4-b-D-glucanase in the apoplast of Arabidopsis thaliana leaves.** *Mol Breed* (6):37–46.
98. Ziegelhoffer TJ RA, Austin-Phillips S: **Biomass ethanol: Technical progress,opportunities, and commercial challenges**. *Mol Breed* 2001(8):147–158.
99. Ziegelhoffer TJ WJA-PS: **Expression of bacterial cellulase genes in transgenic alfa alfa, potato and tobacco** *Mol Breed* 1999, **5**:309-318
100. Harrison M GJ, Coleman H, Shand K, Kinkema M, Palupe A, Hassall R, Sainz M, Lloyd R, Miles S, Dale J,: **Accumulation of recombinant cellobiohydrolase and endoglucanase in the leaves of mature transgenic sugar cane**. *Plant Biotech J* 2011, **8**:1-13.
101. Ziegelhoffer TJ RA, Austin-Phillips S: **Expression of Acidothermus cellulyticus E1 endo- beta 1.4 glucanase catalytic domain in transplastomic tobaco**. *Plant Biotehnology Journal* 2009(7):527-536.
102. Yu L GB, Rutzke C, Walker L, Wilson D, Hanson M: **Expression of thermostable microbial cellulases in the chloroplasts of nicotine-free tobacco**. *J Biotechnol* 2007, **131**:362-369.
103. Kim S LD, Choi I, Ahn S, Kim Y, Bae H: ***Arabidopsis thaliana* Rubisco small subunit transit peptide increases the accumulation of *Thermotoga maritima* endoglucanase Cel5A in chloroplasts of transgenic tobacco plants** *Transgen Res* 2010, **19**:489-497.
104. Dai Z HB, Quesenberry RD, Thomas SR: **Optimization of Acidothermus celluloyticus endogluendoglucanase (E1) production in transgenic tobacco plants by transcriptional, post-transcription and post-translational modification.** *Transgenic Res* 2005(14):627–643.
105. Jung S KS, Bae H, Hyoun S, Bae H-J: **Expression of thermostable bacterial beta glucosidase (BgIB) in transgenic tobacco plants**. *Bioresouce technology* 2010, **101**:7144-7150.
106. Harrison M GJ, Coleman H, Shand K, Kinkema M, Palupe A, Hassall R, Sainz M, lloyd R, Miles S, Dale J: **Accumulation of recombinant cellobiohydorlase and endoglucanase in the leaves of matue transgenic sugar cane**. *Plant Biotehnology Journal* 2011:1-13.
107. Kim S LD-S, Choi s, , Ahn S-j KY-H, Bae H-J: ***Arabidopsis thaliana* Rubisco small subunit transit peptide increases the accumulation of *Thermotoga maritima* endoglucanase Cel5A in chloroplasts of transgenic tobacco plants** *Transgenic Res* 2010, **19**:489–497.
108. Xu B E-TL, Sathitsukanosoh N, Shen Z, Shen H, Zhang P,Dixon R, Zhao B,: **Silencing of 4-coumarate:coenzyme A ligase in switchgrass leads to reduced lignin content and improved fermentable sugar yields for biofuel production**. *New Phytologist* 2011, **192**:611-625.
109. Xu B HL, Shen Z, Welbaum G, Xang X, Zhao B,: **Selection and characterisation of a new switchgrass (*Panicum virgatum*) line with high embryogenic capacity for genetic transformation** *Scientia Horticulturae* 2011 **129**:8854-8861
110. Jefferson R KT, Bevan M: **GUS fusions: ,B-glucuronidase as a sensitive and versatile gene fusion marker in higher plants**. *The EMBO Journal* 1987, **6**(13):3901 -3907.
111. Aronoff S WJ: **Quantitative measurement of leaf chlorophylls by spectrophotometry of their pheophytins in aqueous alcoholic extracts**. *Plant physiology* 1962, **37**(5):584-589.
112. Dere S GT, Sivaci R,: **Spectrophotometric determination of chlorophyll- A, B and total carotenoid contents of some algae species using different solvents** *Tr J Botany* 1998, **22**:13-17.
113. Tawyman R SE, Kohil A, Capell T, Christou P: **Testing of genetic manipulation in plants**, vol. XVIII: Springer 2002.
114. Sun Ye CJ, Himmel M, Skory C, Adney W, Thomas S, Tisserat B, Nishimura Y, Yamamoto Y: **Expression and characterisation of *Acidothermuc cellulolyticus* E1 endoglucanase in transgenic duckweed *Lemna minor* 8627**. *Bioresource Technol* 2007, **98**:2886-2872.

115. Dai Z HB, Anderson D, Thomas S: **Expression of *Acidothermaticus cellulolyticus* endoglucanase E1 in transgenic tobacco: biochemical characteristics and physiological effects**. *Transgen Res* 2000, **9**:43-45.
116. Ye S CJ, Himmel M, Skory C, Adney W, Thomas S, Tisserat B, Nishimura Y, Yamamoto Y: **Expression and characterisation of *Acidothermus cellulolyticus* E1 endoglucanase in transgenic duckweed *Lemna minor* 8627**. *Bioresource Technology* 2007, **98**:2866-2872.
117. Kim S V, Gelvin S: **Genome-wide analysis of Agrobacterium T-DNA integration sites in the Arabidopsis genome generated under non-selective conditions**. *The Plant Journal* 2007, **51**:271-279.
118. Forsbach A SD, Lechtenberg B, Gils M, Renate, R S: **A comprehensive characterization of single-copy T-DNA insertions in the *Arabidopsis thaliana* genome**. *Plant Mol Biol* 2003, **52**:161-173.
119. Emani C SG, Rathore K: **Trasgene silencing and reactivation in *Sorghum***. *Plant Sci* 2002, **162**:181-192.
120. McLaughlin S, Walsh M: **Evaluating environmental consequences of producing herbaceous crops for bioenergy**. *Biomass and Bioenergy* 1998, **14**(4):317-324.
121. Talbert LE TD, Burns JC, Rawlings JO, Moll RH: **Estimates of genetic-parameters in switchgrass** *Crop Sci* 1983, **23**:725-728.
122. Parrish DJ FJ: **The biology and agronomy of switchgrass for biofuels** *Crc Rv Plant Sci* 2005, **24**:423-459.
123. Qu S DA, Wing R, Sundersan V: **A versatile transposon based activation tag vector system for functional genomics in cereals and other monocot plants** *Plant Physiology* 2008, **146**:189-199.
124. Liu Y MN, Oosumi T, Whittier R: **Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal assymetric interlaced PCR**. *Plant J* 1995:457-463.
125. Liu Y CY: **High-efficiency thermal assymetric interlaced PCR for ampliication of unknown flanking sequences** *BioTechniques* 2007, **43**:649-656.
126. Ramchandran S SV: **Transposons as tool for functional genomics** *Plant Physiol Bioch* 2001, **39** 243-252.
127. Fitzmaurice W NL, Wernsman E,Thompson W, Conkling M,: **Transposon tagging of the sulfur gene of tobacco using engineered maize Ac/Ds elements**. *Genetics* 1999, **153**:1919-1928.
128. Fladung M PO: **Ac/Ds-transposon activation tagging in poplar: a powerful tool for gene discovery**. *BMC Genomics* 2012, **13**:1-15.

APPENDIX I

Tissue culture media for switchgrass

Medium Name	Purpose	Compositions
CI Medium	SE callus induction	MS salts and vitamins, 3% maltose, 0.8% agar, 22.5 µM 2, 4 D and 5 µM 6 BA
RE Medium	SE callus regeneration	MS salts and vitamins, 3% maltose, 0.8% agar, 1.44 µM GA3
Infection medium	Agrobacterium suspension and infection	½ MS salts and vitamins, 3% maltose, 200µM acetosyringone
Co-cultivation medium	Agrobacterium cocultivation	½ MS salts and vitamins, 3% maltose, 0.8% agar, 22.5 µM 2, 4 D and 5 µM 6 BA
CS medium	Transformed callus selection	CI medium with 375 augmentin and 50 mg/l kanamycin
RS medium	Transformed callus regeneration	RE medium with 375 augmentin and 50 mg/l kanamycin

APPENDIX II

Optimized endoglucanase sequence

The DNA and amino acid sequences of original and optimized endoglucanase

The restriction sites used for cloning are highlighted in yellow. The DNA sequence encodes rice Rubisco signal peptide is underlined. The native intron was highlighted in blue. The open reading frame (ORF) of endoglucanase gene is in italic. The DNA sequence encode 9xHis tag is highlighted in green. The start and stop codons are highlighted in red and bold.

>The original beta-glucanase gene

ACTAGTGCAGAG**ATG**GCCCCCTCCGTGATGGCGTCGTcggccaccaccgtcgctcccttccaggggctcaagtccaccg
ccggcatgcccgtcgcccgccgctccggcaactccagcttcggcaacgtcagcaatggcggcaggatcaggtgcatgcaggtAATAAC
CTACTGACCCAACACACATTATTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCAACATTAACCAATA
ATTCAATTATCGTTTATTTAGGTGTGGtctaga***ATG****AAGTTCCTTCAAGTCCTCCCTGCCCTCATACCGGCCG*
CCCTGGCCCAAACCAGCTGTGACCAGTGGGCAACCTTCACTGGCAACGGCTACACAGTCAGCAACAACCT
TTGGGGAGCATCAGCCGGCTCTGGATTTGGCTGCGTGACGGCGGTATCGCTCAGCGGCGGGGCCTCCTG
GCACGCAGACTGGCAGTGGTCCGGCGGCCAGAACAACGTCAAGTCGTACCAGAACTCTCAGATTGCCAT
TCCCCAGAAGAGGACCGTCAACAGCATCAGCAGCATGCCCACCACTGCCAGCTGGAGCTACAGCGGGAG
CAACATCCGCGCTAATGTTGCGTATGACTTGTTCACCGCAGCCAACCCGAATCATGTCACGTACTCGGGA
GACTACGAACTCATGATCTGGCTTGGCAAATACGGCGATATTGGGCCGATTGGGTCCTCACAGGGAACA
GTCAACGTCGGTGGCCAGAGCTGGACGCTCTACTATGGCTACAACGGAGCCATGCAAGTCTATTCCTTTG
TGGCCCAGACCAACACTACCAACTACAGCGGAGATGTCAAGAACTTCTTCAATTATCTCCGAGACAATAA
AGGATACAACGCTGCAGGCCAATATGTTCTTAGCTACCAATTTGGTACCGAGCCCTTCACGGGCAGTGGA
*ACTCTGAACGTCGCATCCTGGACCGCATCTATCAAC*catcaccatcaccatcaccatcaccac**TAA**AAGCTT

> The original beta-glucanase

MKFLQVLPALIPAALAQTSCDQWATFTGNGYTVSNNLWGASAGSGFGCVTAVSLSGGASWHADWQWSG
GQNNVKSYQNSQIAIPQKRTVNSISSMPTTASWSYSGSNIRANVAYDLFTAANPNHVTYSGDYELMIWLGK
YGDIGPIGSSQGTVNVGGQSWTLYYGYNGAMQVYSFVAQTNTTNYSGDVKNFFNYLRDNKGYNAAGQYV
LSYQFGTEPFTGSGTLNVASWTASINHHHHHHHHHH*

>The optimized beta-glucanase gene

ACTAGTGCAGAG**ATG**GCCCCCTCCGTGATGGCGTCGTCGGCCACCACCGTCGCTCCCTTCCAGGGGCTC
AAGTCCACCGCCGGCATGCCCGTCGCCCGCCGCTCCGGCAACTCCAGCTTCGGCAACGTCAGCAATGGC
GGCAGGATCAGGTGCATGCAGGTAATAACCTACTGACCCAACACACATTATTCTTCTTCTTCTTCTTCTTC
TTCTTCTTCTTCTTCTTCAACATTAACCAATAATTCAATTATCGTTTATTTAGGTGTGGTCTAGA***ATG****AAGT*

TCCTCCAGGTGCTCCCAGCTCTGATCCCAGCAGCTCTCGCTCAGACCTCCTGCGACCAGTGGGCTACTTTC
ACCGGCAACGGCTACACTGTCAGCAACAACCTGTGGGGTGCTTCTGCCGGCTCAGGATTCGGATGCGTG
ACCGCTGTCTCCCTCAGCGGAGGCGCGTCTTGGCACGCTGATTGGCAGTGGTCAGGAGGACAGAACAAC
GTGAAGTCCTACCAGAACAGCCAGATTGCTATCCCTCAGAAGAGGACTGTCAACTCTATTTCATCCATGC
CCACCACTGCGAGCTGGTCTTACTCAGGCTCCAACATCCGCGCCAACGTGGCTTACGACCTGTTCACCGC
GGCCAACCCGAACCATGTCACTTACAGCGGCGATTACGAGCTCATGATTTGGCTGGGCAAGTACGGCGA
CATCGGCCCTATTGGCTCTTCACAGGGCACCGTGAACGTCGGCGGCCAGTCCTGGACTCTCTACTACGGC
TACAACGGCGCTATGCAGGTGTACAGCTTCGTCGCGCAGACCAACACTACCAACTACTCTGGCGATGTGA
AGAACTTCTTCAACTACCTGAGGGACAACAAGGGCTACAACGCCGCTGGCCAATACGTGCTCTCATACCA
GTTCGGCACTGAGCCGTTCACCGGCTCCGGCACTCTGAACGTGGCCTCCTGGACCGCCTCTATCAACCAG
*CTGCA*CCATCATCACCACCATCACCATCAC**TGA**AAGCTT

> The optimized beta-glucanase

MKFLQVLPALIPAALAQTSCDQWATFTGNGYTVSNNLWGASAGSGFGCVTAVSLSGGASWHADWQWSGGQNNVKSYQNSQIA
IPQKRTVNSISSMPTTASWSYSGSNIRANVAYDLFTAANPNHVTYSGDYELMIWLGKYGDIGPIGSSQGTVNVGGQSWTLYYGYNG
AMQVYSFVAQTNTTNYSGDVKNFFNYLRDNKGYNAAGQYVLSYQFGTEPFTGSGTLNVASWTASINQLHHHHHHHHHH*

>>Alignment the DNA sequences of original and optimized endoglucanase genes

Original 1 ACTAGTGCAGAG**ATG**GCCCCCTCCGTGATGGCGTCGTCGGCCACCACCGTCGCTCCCTTC
Optimized 1 ACTAGTGCAGAG**ATG**GCCCCCTCCGTGATGGCGTCGTCGGCCACCACCGTCGCTCCCTTC

Original 61 CAGGGGCTCAAGTCCACCGCCGGCATGCCCGTCGCCCGCCGCTCCGGCAACTCCAGCTTC
Optimized 61 CAGGGGCTCAAGTCCACCGCCGGCATGCCCGTCGCCCGCCGCTCCGGCAACTCCAGCTTC

Original 121 GGCAACGTCAGCAATGGCGGCAGGATCAGGTGCATGCAGGTAATAACCTACTGACCCAAC
Optimized 121 GGCAACGTCAGCAATGGCGGCAGGATCAGGTGCATGCAGGTAATAACCTACTGACCCAAC

Original 181 ACACATTATTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCAACATTAACCAAT
Optimized 181 ACACATTATTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCAACATTAACCAAT

Original 241 AATTCAATTATCGTTTATTTAGGTGTGGTCTAGA**ATG**AAGTTCCTTCAAGTCCTCCCTGC
Optimized 241 AATTCAATTATCGTTTATTTAGGTGTGGTCTAGA**ATG**AAGTTCCTCCAGGTGCTCCCAGC

Original 301 CCTCATACCGGCCGCCCTGGCCCAAACCAGCTGTGACCAGTGGGCAACCTTCACTGGCAA
Optimized 301 TCTGATCCCAGCAGCTCTCGCTCAGACCTCCTGCGACCAGTGGGCTACTTTCACCGGCAA

Original 361 CGGCTACACAGTCAGCAACAACCTTTGGGGAGCATCAGCCGGCTCTGGATTTGGCTGCGT
Optimized 361 CGGCTACACTGTCAGCAACAACCTGTGGGGTGCTTCTGCCGGCTCAGGATTCGGATGCGT

Original 421 GACGGCGGTATCGCTCAGCGGCGGGGCCTCCTGGCACGCAGACTGGCAGTGGTCCGGCGG
Optimized 421 GACCGCTGTCTCCCTCAGCGGAGGCGCGTCTTGGCACGCTGATTGGCAGTGGTCAGGAGG

Original 481 CCAGAACAACGTCAAGTCGTACCAGAACTCTCAGATTGCCATTCCCCAGAAGAGGACCGT
Optimized 481 ACAGAACAACGTGAAGTCCTACCAGAACAGCCAGATTGCTATCCCTCAGAAGAGGACTGT

Original 541 CAACAGCAT--CAGCAGCATGCCCACCACTGCCAGCTGGAGCTACAGCGGGAGCAACATC
Optimized 541 CAACTCTATTTCATC--CATGCCCACCACTGCGAGCTGGTCTTACTCAGGCTCCAACATC

Original 599 CGCGCTAATGTTGCGTATGACTTGTTCACCGCAGCCAACCCGAATCATGTCACGTACTCG
Optimized 599 CGCGCCAACGTGGCTTACGACCTGTTCACCGCGGCCAACCCGAACCATGTCACTTACAGC

Original 659 GGAGACTACGAACTCATGATCTGGCTTGGCAAATACGGCGATATTGGGCCGATTGGGTCC
Optimized 659 GGCGATTACGAGCTCATGATTTGGCTGGGCAAGTACGGCGACATCGGCCCTATTGGCTCT

Original 719 TCACAGGGAACAGTCAACGTCGGTGGCCAGAGCTGGACGCTCTACTATGGCTACAACGGA
Optimized 719 TCACAGGGCACCGTGAACGTCGGCGGCCAGTCCTGGACTCTCTACTACGGCTACAACGGC

Original 779 GCCATGCAAGTCTATTCCTTTGTGGCCCAGACCAACACTACCAACTACAGCGGAGATGTC
Optimized 779 GCTATGCAGGTGTACAGCTTCGTCGCGCAGACCAACACTACCAACTACTCTGGCGATGTG

Original 839 AAGAACTTCTTCAATTATCTCCGAGACAATAAAGGATACAACGCTGCAGGCCAATATGTT
Optimized 839 AAGAACTTCTTCAACTACCTGAGGGACAACAAGGGCTACAACGCCGCTGGCCAATACGTG

Original 899 CTTAGCTACCAATTTGGTACCGAGCCCTTCACGGGCAGTGGAACTCTGAACGTCGCATCC
Optimized 899 CTCTCATACCAGTTCGGCACTGAGCCGTTCACCGGCTCCGGCACTCTGAACGTGGCCTCC

Original 959 TGGACCGCATCTATCAACCA------TCACCATCACCATCACCATCACCACTAAAAGCTT
Optimized 959 TGGACCGCCTCTATCAACCAGCTGCACCATCATCACCACCATCACCATCACTGAAAGCTT

>>Alignment the amino acid sequences of original and optimized endoglucanase proteins

Original 1 MKFLQVLPALIPAALAQTSCDQWATFTGNGYTVSNNLWGASAGSGFGCVTAVSLSGGASW
optimized 1 MKFLQVLPALIPAALAQTSCDQWATFTGNGYTVSNNLWGASAGSGFGCVTAVSLSGGASW

Original 61 HADWQWSGGQNNVKSYQNSQIAIPQKRTVNSISSMPTTASWSYSGSNIRANVAYDLFTAA
optimized 61 HADWQWSGGQNNVKSYQNSQIAIPQKRTVNSISSMPTTASWSYSGSNIRANVAYDLFTAA

Original 121 NPNHVTYSGDYELMIWLGKYGDIGPIGSSQGTVNVGGQSWTLYYGYNGAMQVYSFVAQTN
optimized 121 NPNHVTYSGDYELMIWLGKYGDIGPIGSSQGTVNVGGQSWTLYYGYNGAMQVYSFVAQTN

Original 181 TTNYSGDVKNFFNYLRDNKGYNAAGQYVLSYQFGTEPFTGSGTLNVASWTASIN--HHHH
optimized 181 TTNYSGDVKNFFNYLRDNKGYNAAGQYVLSYQFGTEPFTGSGTLNVASWTASINQLHHHH

Original 239 HHHHH
optimized 241 HHHHH

APPENDIX III

TAIL Primers and insertion flanking sequences

Name of primers	Sequence of primers
Hyg 0a	CGGGCGTATATGCTCCGCATTGGTCTTG
Hyg 1a	ACGATGGACTCCAGTCCGGCCCTCTATCAGAGCTTGGTTGACGGC
Hyg2a	GAGCCGGGACTGTCGGGCGTACACAAAT
LAD 1-1	ACGATGGACTCCAGAGCGGCCGCVNVNNNGGAA
LAD 1-2	ACGATGGACTCCAGAGCGGCCGCBNBNNNGGTT
LAD 1-3	ACGATGGACTCCAGAGCGGCCGCVVNVNNNCCAA
LAD 1-4	ACGATGGACTCCAGAGCGGCCGCBDNBNNNCGGT
AC1	ACGATGGACTCCAGAG

List of DNA sequences of seven TAIL-PCR products: Red color indicates vector sequence whereas black color indicates insertion-flanking sequence.

1.

pSQ5-5

TGGCATAGCGCGATCGTCTGGACGATGGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCC
AGCACTCGTCCGAGGGCAAAGAAATAGAGTAGATGCCGACCGGGATCTGTCGATCGACAAGCTCGAGT
TTCTCCATAATAATGTGTGAGTAGTTCCCAGATAAGGGAATTAGGGTTCCTATAGGGTTTCGCTCATGTG
TTGAGCATATAAGAAACCCTTAGTATGTATTTGTATTTGTAAAATACTTCTATCAATAAAATTTCTAATTCC
TAAAACCAAAATCCAGTACTAAAATCCAGATCCCCCGAATTAATTCGGCGTTAATTCAGTACATTAAAAA
CGTCCGCAATGTGTTATTAAGTTGTCTGACCGAAACCTGCCGCTTATTCTTGCATCATCTCATTGCCGACA
AGGAACCGAGTACGCCTCGTTCAGCGTCGAGCTTCTCGGGGAAATGAACGGTAGTAAGGCCCTGTAAG
TTCTAAAAAAAAATTACAGTACCTGTCACATTGAATGTTCAAATATATGCATGGAGCATTAAATGTAGTT
AAAAAATAACTAATTACACAGCCTAACTGATAAGTACGAGATGAATCTTTTAAATTTAATTAATTCATAAT
TGAATATTATTTATTAAATAACAACAAAATGTGCTACAGTACCAAAACCCAAACCTTTTCGCGAACTAAAC
AGAGCCTAAGTTGGAAAATCGCTGTCAAGGTGTTTGCCCTGTTTGGCTTTGGCATGGCGCCCATGTGAG
AGGAGGGATTGTACAATTTGTAAGTACCACGCAAATGCACGGCAACCTTCAAATGGGAAATGCATGCTA
GGCAATAGGCATCTACCGTGCCCCGCGCGCGCCTCTCGAGAGCTCTCATCGATA

Panicum virgatum clone PV_ABa006-K10, complete sequence
Length=136391

2.
PSQ5 – 8

CGGCATAGCGCGGCGTCTGGACGATGGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCA
GCACTCGTCCGAGGGCAAAGAAATAGAGTAGATGCCGACCGGGATCTGTCGATCGACAAGCTCGAGTT
TCTCCATAATAATGTGTGAGTAGTTCCCAGATAAGGGAATTAGGGTTCCTATAGGGTTTCGCTCATGTGT
TGAGCATATAAGAAACCCTTAGTATGTATTTGTATTTGTAAAATACTTCTATCAATAAAATTTCTAATTCCT
AAAACCAAAATCCAGTACTAAAATCCAGATCCCCCGAATTAATTCGGCGTTAATTCAGTACATTAAAAAC
GTCCGCAATGTGTTATTAAGTTGTCTGACCGAAACCTGCCGCTTATTCTTGCATCATCTCATTGCCGACAA
GGAACCGAGTACGCCTCGTTCAGCGTCGAGCTTCTCGGGGAAATGAACGGTAGTAAGGCCCTGTAAGTT
CTAAAAAAAAATTACAGTACCTGTCACATTGAATGTTCAAATATATGCATGGAGCATTAAATGTAGTTAA
AAAATAACTAATTACACAGCCTAACTGATAAGTACGAGATGAATCTTTTAAATTTAATTAATTCATAATTG
AATATTATTTATTAAATAACAACAAAATGTGCTACAGTACCAAAACCCAAACCTTTTCGCGAACTAAACA
GAGCCTAAGTTGGAAAATCGCTGTCAAGGTGTTTGCCCTGTTTGGCTTTGGCATGGCGCCCATGTGAGA
GGAGGGATTGTACAATTTGTAAGTACCACGCAAAATGCACGGCAACCTTCAAATGGGAAATGCATGCTA
GGCAATAGGCATCTACCGCCCCCCGCGGCCGCTCTGGAGTTTCCATCGATAAA

Panicum virgatum clone PV_ABa006-K10, complete sequence

3.

pSQ5 – 25 R2

GCGTCAAGCGCTTCGTCTGGACGATGGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCA
GCACTCGTCCGAGGGCAAAGAAATAGAGTAGATGCCGACCGGGATCTGTCGATCGACAAGCTCGAGTT
TCTCCATAATAATGTGTGAGTAGTTCCCAGATAAGGGAATTAGGGTTCCTATAGGGTTTCGCTCATGTGT
TGAGCATATAAGAAACCCTTAGTATGTATTTGTATTTGTAAAATACTTCTATCAATAAAATTTCTAATTCCT
AAAACCAAAATCCAGTACTAAAATCCAGATCCCCCGAATTAATTCGGCGTTAATTCAGTACATTAAAAAC
GTCCGCAATGTGTTATTAAGTTGTCTAAGCGTCAATTTGTTTCCTAATCCGATAGCTCTGGAGTCTAGTCT
GACATTCGGAGGCACTTGCACGGATCCAGCTGAAGGTGTGGATGCGCCCAGCTCTGACCGCCCCCCCGC
GGCCGCTCTGGAGTCCATCGTAAA

Fragaria vesca clone VT_468_R T-DNA, Insertion flanking sequence
length=263

4.
pSQ5- 26
CGGCGTCTGGACCGATGGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCACTCGTCC
GAGGGCAAAGAAATAGAGTAGATGCCGACCGGGATCTGTCGATCGACAAGCTCGAGTTTCTCCATAAT
AATGTGTGAGTAGTTCCCAGATAAGGGAATTAGGGTTCCTATAGGGTTTCGCTCATGTGTTGAGCATAT
AAGAAACCCTTAGTATGTATTTGTATTTGTAAAATACTTCTATCAATAAAATTTCTAATTCCTAAAACCAA
AATCCAGTACTAAAATCCAGATCCCCCGAATTAATTCGGCGTTAATTCAGTACATTAAAAACGTCCGCAA
TGTGTTATTAAGTTGTCTAAGCGTCAATTTGTTTACACCACTGCTGTTCCATGTGAACCATAAATATCATTT
TCGTAGCAGCAACCAGAAGCTAACCATGCTTTTTAACTAAGAAGAACTAGGTGCAGGATGGGCCTTTTG
GTTTTCTAATCACTGACTAATGTGCGAAAATATATTTCA

Sorghum bicolor hypothetical protein, mRNA Length=1671 GENE ID: 8057514 SORBIDRAFT_04g002650

5.

pSQ5- 27 R1-1

GGCGGCCTAAGCGGGGCCGTCTGGACGATGGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGC
AAGAACTCGTCCGAGGGCAAAGAAATAGAGTAGATGCCGACCGGGATCTGTCGATCGACAAGCTCGAG
TTTCTCCATAATAATGGGGGAGTAGTTCCCAGATAAGGGAATTAGGGTTCCTATAGGGGTTCGCTCATG
TGTTGAGCATATAAGAAACCCTTAGTATGGATGTGTATTTGGAAAATACTTCTATCAATAAAATTTCTAAT
TCCTAAAACCAAAATCCAGTACTAAAATCCAGATCCCCCGAATTAATTCGGAGTTGATTGAGTACATTAA
AAACGTCCGCAATGTGTTATTAATTCGGCGTTAATTCAGTACATTAAAGACGGCTCTGGCAGATTCTCCA
AAACTTTTGAATACCAACTTTGTTCAGGTTGTCGAGATCGACAACTTTTGTTTCAGGAACTTTTTCATTTG
AGCAATGGTTTGAAAGTTATTCATGTATTGCAAAAACTACCTATTAAAAGGCTTGTCATTGCATGTGACA
ACTCTCATTTTCTCTCTTAGGCCTCAACTGGACTTTTATTATTCTTGATATGGCGGATATTCCTATAAATTG
TTGGGGACATTATTTGATCCAATTGAAAGTTGTGTTAAATCCAGGACAAACATGATTTGAATTGGTTATC
AATTCATGTGACAGTGACTGAATTTTTTGTTTGATTCTCTTTGTATAGAGAATTGTTAAAGCATTCTGAAA
GGCGTGTTTTCCGGCCCTTCTCCACTGGAAAGCTTTCCCCTGCTCTATAGATGGAGACAATTCTCTTGGG
GCCAAGAGTTTGGTCCCATGCATGCTGTGGAATGCTAGAGTTTGACCTTCTCTTCTTATCCTCCGAGACT
GGAATGAGCGTCGGAGAAATTCAATGGAATCAAGGTTACAAGTCACCGGAACTACTATTGCATGGCAA
ACTTTTCTTCGTGAATGTGGGGCAGATGTACGCCTCGACTGAATTAATCTTGGATCGCCCATGCCGATGA
TTTGGCGATGTATTTTACCGCCTTTCAGCAATGTCCATCTCTTTCAAAAAAAATATGTTCACGATTTCCTCA
GTTACCGGCTACAGGAGGTAGGAACAGGACGCTATATGAGAGTCCTGGGTAAGAGTGATTTAACTACG
ACAAAACAAGAGCTTTTGCGTCATTCCTATTAAACTTCACCGCTCCATACAAGAACAGGGCTTGTTTTGG
TGAGATCT

Panicum virgatum clone PV_ABa107-B06, complete sequence length=158729

6.

pSQ5 29-2

AAATAGAGTAGATGCCGACCGGGATCTGTCGATCGACAAGCTCGAGTTTCTCCATAATAATGTGTGAGT
AGTTCCCAGATAAGGGAATTAGGGTTCCTATAGGGTTTCGCTCATGTGTTGAGCATATAAGAAACCCTTA
GTATGTATTTGTATTTGTAAAATACTTCTATCAATAAAATTTCTAATTCCTAAAACCAAAATCCAGTACTAA
AATCCAGATCCCCCGAATTAATTCGGCGTTAATTCAGTACATTAAAAACGTCCGCAATGTGTTATTAAGTT
GTCTAAGCGTCACCACAAACCCTCCAACCTAGGGTGAGAGAGTAGCTATGGACTAGCTATCTCTAAACA
AAAACCCTTAGGAAGAGAAACGGAAGCAATGCGGAAAAACTCCCCAAACAGCAGCTCTGCTGAGTCAG
ACCGGTTTGACCGCTAGCACAGACCGG

Panicum virgatum clone PV_ABa051-K10, complete sequence length=172117

7.

pSQ5 32-2-1

CGGCGTCTGGACCGATGGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCACTCGTCC
GAGGGCAAAGAAATAGAGTAGATGCCGACCGGGATCTGTCGATCGACAAGCTCGAGTTTCTCCATAAT
AATGTGTGAGTAGTTCCCAGATAAGGGAATTAGGGTTCCTATAGGGTTTCGCTCATGTGTTGAGCATAT
AAGAAACCCTTAGTATGTATTTGTATTTGTAAAATACTTCTATCAATAAAATTTCTAATTCCTAAAACCAA
AATCCAGTACTAAAATCCAGATCCCCCGAATTAATTCTGAGCCTTTTCATCACATTAAGCCCGCTTTTGGA
GTGTTATTAATTCCGCGTTAATTCGGTACATTAAAGACGGCTCTGGTAGATTCTCCAAAACTTTTGAATAC
CAACTTTGTTCAGTTTTTCGAGATCTACAACTTTTGTTTCAGGAACTTTTTCATTTGAGCAATGGTTTGAAA
GTTATTTAATTATTTCAAAAACTACCAATTAAAAGTCTTGTCATTTCATGTGACAACTTTCATTTTCTCTCTT
AGGCCTCAACTGGACTTTTATTATTCTTGTTATGGCGGATATTCCTATAAATTTTTAGGGACATTAGTTGA
TCCAATTGAAAGTTGTTTTAAATCCAGGACAAACATGATTTGAATTGGTTATCAATTCATGTGACAGTGTT
TGAATTTTTTGTTTGATTCAATTTGTATAGAGAATTGTTAAAGCATTCTGAAAGGTGTGTTTTCCGGCAGT
TTTCCAATGGAAAGTTTTTCCCTGCTCTATAGATGGTGACAATTCTCTTGTGGCCTAAGATTATGGTCACA
TGCATGCTGTGGATGCAAGAATTTGACCTTCTCTTCTTATCCTCCGGACTGGGANNNNCGTCGGANGGA
AATTCACTGGAATCAAGGTTACNGTCACCGATCCTACTANTTGCNANGGNAAAGNTTNCTNCGACNNN
GTNGNGGGCAGATC

1.*Panicum virgatum* clone PV_ABa107-B06, complete sequence length=158729

2.*Panicum virgatum* clone PV_ABa003-K10, complete sequence length=128623

APPENDIX IV

Switchgrass genomic DNA extraction with 5% CTAB

1. Freeze 5-10 g plant material at -80^{0}C
2. Grind the plant material with liquid nitrogen; don't let the sample (powder) thaw. (It will degrade your DNA). If you have too many samples, grind them by batch and freeze in -80°C until all samples have been processed.
3. Add 15ml 5X CATB (plus 0.1% Mercaptoethanol, must be fresh), mix well (make sure mix the powder with CTAB before it is thawed)
4. Incubate your sample at 65^0 C for 2 hour (mix occasionally), you should see your sample separated into dark green layer and light green layer.
5. Add 15ml chloroform/isogamy alcohol (24:1) and mix well, then shake your sample on the rotary shaker for 1 hour at room temperature
6. Spin at 5000 g for 15 min at room temperature (20°C)
7. Transfer the supernatant to a clean tube (be careful, you don't want to transfer any leaf debris or any dirty stuff), add 12 ml isopropyl alcohol to the sample, and mix slowly by invert the tube. You should see the solution turn cloudy and your DNA come out from the solution.
8. Mix slowly and let the DNA wired together, then hook out your DNA using disposable glass pipets (seal the end to make a hook), transfer the hooked DNA to a new 1.5ml tube that has 1ml 70% ethanol.
9. Shake the tube by hands and discard the 70% ethanol. Spin and pipe out as much as possible of the leftover ethanol
10. Add 500 ul of T5E, and dissolve your DNA (you can incubate at 65°C for 1hour to dissolve your DNA, but usually your DNA should be easily dissolved in T5E)
11. Add 3 ul RNAse (10mg/ml), incubate at 37°C for 1hour
12. Add 500ul phenol to your sample, mix well, shake for 15min, then spin at 7000 rpm for 5min
13. Transfer the supernatant into a new tube, add 500ul chloroform/ isoamyl alcohol (24:1), mix well, shake for 15min, then spin at 7000rpm for 5min

14. Repeat step 13.
15. Transfer the supernatant into a new tube and add 500ul isopropyl alcohol, plus 50ul 3M NaOAc (pH=5.2), mix well, and then spin at high speed for 5min (you can also hook out your DNA instead of spin down, which will result in better quality DNA)
16. Wash your sample with 70% ethanol for at least 4 times, air dry the sample (make sure it is dry, but not over dry)
17. Add 200ul 1XTE (depends on the size of your pellet) to dissolve your DNA. Store your DNA in -20^0 C

5x CTAB**:** *(high CTAB and high Salt to precipitate the* polysaccharides and phenolic compounds, which cannot be removed by phenol or chloroform, very sticky*)*

0.7 M NaCl

0 mM Tris pH 8.0

5% CTAB (Hexadecyltrimethylamonium bromide)

10 mM EDTA

0.5% sodium metabisulfite

Note: This mixture can be stored at RT

To make this buffer, it is fastest to add NaCl till all other chemicals have been completely dissolved.

0.1% 2-Mercaptoethanol (always add freshly, never put in your 5XCTAB stock)

For 500 ml 5xCATB

Dissolve 25 g of CTAB in 390ml ddH_2O (pre-warmed to >65°C) with a stir bar, stir while adding the CTAB slowly, till it is completely dissolved)

Then add the following components in order

2.5 g of sodium metabisulfite

25 ml of 1M Tris-Cl (pH8.0)

10 ml of 0.5 M EDTA (pH=0.8)

Once everything is dissolved, then add 70ml of 5M NaCl

This mixture can be stored at RT

Add 0.1 % 2-Mercaptoethanol just before to use.

T5E: (50mM Tris pH 8.0, 10mM EDTA pH 8.0)

1xTE: (10mM Tris pH 8.0, 10mM EDTA pH 8.0)