Functional Genomic Analyses of Switchgrass Developmental Processes

Nathan Palmer

University of Nebraska-Lincoln, nathan.palmer@ars.usda.gov

Follow this and additional works at: http://digitalcommons.unl.edu/biochemdiss

Part of the Biochemistry Commons

http://digitalcommons.unl.edu/biochemdiss/20

This Article is brought to you for free and open access by the Biochemistry, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Theses and Dissertations in Biochemistry by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Functional Genomic Analyses of Switchgrass Developmental Processes

by

Nathan Andrew Palmer

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

Major: Biochemistry

Under the Supervision of Professor Gautam Sarath

Lincoln, Nebraska

April, 2015
Functional Genomic Analyses of Switchgrass Developmental Processes

Nathan Andrew Palmer, Ph.D.

University of Nebraska, 2015

Adviser: Gautam Sarath

Switchgrass (*Panicum virgatum* L.), a C₄ perennial grass species, is being developed as a bioenergy crop. Although much is known from a breeding perspective, there is limited information on the functional genomics of this crop, specifically regarding molecular mechanisms controlling aerial senescence, winter dormancy, and traits that confer winter hardiness. Using functional genomics to generate a transcriptional roadmap underpinning senescence and winter dormancy will provide researchers with a molecular understanding that can be applied to improve switchgrass germplasm.

In an initial study, a *de novo* assembly of the crown and rhizome transcriptome from an upland cultivar Summer was performed. This study added about 30,000 new expressed-sequence tags to the public databases and provided the first details on the molecular aspects of switchgrass rhizome metabolism.

The acquisition and remobilization of minerals between plant tissues is an essential feature of plant development. For switchgrass, seasonal remobilization of minerals can impact rhizome health and overall sustainable production of biomass. A total of 520 putative mineral transporters were identified in the draft version of the switchgrass genome (Pvi0), and their expression patterns were queried in publically available transcriptome datasets from various tissues and developmental stages. These
analyses suggested that some minerals more readily mobilized to the rhizomes at the end of the growing season. Transcripts for several mineral transporters were specifically enriched in rhizomes harvested at dormancy.

RNA-Seq analysis of field grown cv. Summer flag leaves resulted in the creation of a four-stage roadmap of flag leaf development and identified many genes specifically associated with expansion through senescence. Among these were two transcription factors homologous to a gene that regulates mineral remobilization from source to sink in wheat (*Triticum aestivum*) plants.

A related study established the molecular differences in rhizomes in contrasting cultivars of switchgrass. Results indicated that the latitudinally adapted cv. Summer was approaching a dormant state, whereas the rhizomes from the latitudinally non-adapted cv. Kanlow were actively growing. An improper timing of dormancy onset could negatively impact plant fitness.

These novel first results for any warm-season perennial grass species provide a robust molecular framework to understand winter dormancy in perenniating tissues.
Acknowledgements

I would like to thank my advisor, Dr. Gautam Sarath, for all of the mentorship and guidance he has given me throughout my Ph.D. program. I would also like to thank my advisory committee, Dr. Scott Sattler, Dr. Madhavan Soundararajan, Dr. Kenneth Vogel, and Dr. Michael Zeece, for all of their guidance. I would like to thank all of my co-workers in the USDA-ARS Grain, Forage, and Bioenergy Research Unit for all of the assistance they have given.
Table of Contents

Abstract .................................................................................................................................................... ii

Acknowledgements ................................................................................................................................ iv

Table of Contents .................................................................................................................................. v

CHAPTER 1: INTRODUCTION ........................................................................................................... 1

1.1 Switchgrass ........................................................................................................................................ 2

1.2 Perennial Senescence and Dormancy ............................................................................................... 7

  1.2.1 Senescence .................................................................................................................................. 7

  1.2.2 Dormancy ................................................................................................................................... 8

1.3 Sequencing Technologies .................................................................................................................. 11

  1.3.1 Sanger Sequencing .................................................................................................................... 11

  1.3.2 Roche 454 Sequencing .............................................................................................................. 12

  1.3.3 Illumina Sequencing .................................................................................................................. 12

  1.3.4 Plant Genomics Era .................................................................................................................. 14

1.4 Molecular and Bioinformatic Research in Switchgrass ....................................................................... 14

  1.4.1 Switchgrass Sequence-based Resources ..................................................................................... 15

  1.4.2 Genomic Markers ...................................................................................................................... 17

  1.4.3 Switchgrass Transformation Methodologies ............................................................................. 21

  1.4.4 Phenylpropanoid Biosynthesis Pathway ................................................................................... 23

  1.4.5 Transcriptomic and other General Molecular Research ......................................................... 29

1.5 Works Cited ....................................................................................................................................... 34

CHAPTER 2: NEXT-GENERATION SEQUENCING OF CROWN AND
CHAPTER 3: MINERAL TRANSPORTER GENES AND METABOLISM IN SWITCHGRASS

3.3.6 Mineral analyses ........................................................................................................ 107
3.3.7 Statistical analyses ...................................................................................................... 107
3.4 Results and Discussion .................................................................................................... 108
  3.4.1 Discovery of potential switchgrass transporters ..................................................... 109
  3.4.2 Differential regulation of transporter genes in switchgrass tissues over development ................................................................................................................. 109
  3.4.3 Expression profiles of selected mineral transporter gene families 114
  3.4.4 Mineral dynamics in crown and rhizome ................................................................. 126
3.5 Conclusions and Future Directions .................................................................................. 126
3.6 Conflict of Interest Statement ........................................................................................ 128
3.7 Acknowledgements ....................................................................................................... 128
3.8 Works Cited .................................................................................................................... 131

CHAPTER 4: CONTRASTING METABOLISM IN PERENNIATING STRUCTURES OF UPLAND AND LOWLAND SWITCHGRASS PLANTS LATE IN THE GROWING SEASON .................................................................................. 140
4.1 Abstract .......................................................................................................................... 141
4.2 Introduction .................................................................................................................... 143
4.3 Materials and Methods .................................................................................................. 145
  4.3.1 Plant materials, growth conditions, and selection of Harvest Date 145
  4.3.2 RNA-Seq .................................................................................................................. 146
  4.3.3 Mapping and Differential gene expression analysis ................................................. 147
  4.3.4 Genome functional annotation ................................................................................. 148
  4.3.5 Gene set enrichment analysis .................................................................................... 148
4.3.6 Metabolite analysis ................................................................. 148

4.3.7 Enzyme, protein, and immunoassays ........................................ 150

4.3.8 Statistical analyses ................................................................. 150

4.4 Results .......................................................................................... 151

4.4.1 80% of all reads map to the switchgrass draft genome .............. 151

4.4.2 Transcriptomes of Summer and Kanlow crowns and rhizomes are significantly different ................................................................. 151

4.4.3 Metabolite analysis differentiates Kanlow and Summer crowns and rhizomes ................................................................. 154

4.4.4 Nexus between gene-set enrichment and metabolite analyses ..... 156

4.4.5 Enzyme activity, protein levels, and enrichment of key transcription factors confirm GSEA of phenylpropanoid pathway ....... 157

4.5 Discussion ..................................................................................... 162

4.6 Conclusions .................................................................................. 167

4.7 Acknowledgements ....................................................................... 167

4.8 Supporting Information ................................................................. 168

4.9 Works Cited .................................................................................. 169

CHAPTER 5: SWITCHGRASS (PANICUM VIRGATUM L.) FLAG LEAF TRANSCRIPTOMES REVEAL MOLECULAR SIGNATURES OF LEAF DEVELOPMENT, SENESCENCE, AND MINERAL DYNAMICS ............ 178

5.1 Abstract ....................................................................................... 179

5.2 Introduction .................................................................................. 180

5.3 Experimental Methods ................................................................. 182
5.3.1 Field Layout, replication, and sample collection ........................................ 182
5.3.2 Chlorophyll quantification ............................................................................ 182
5.3.3 RNA extraction ............................................................................................ 182
5.3.4 RNA sequencing .......................................................................................... 183
5.3.5 Mapping and expression counting ................................................................. 183
5.3.6 Differential expression analysis ................................................................... 183
5.3.7 Gene set creation .......................................................................................... 184
5.3.8 Pathway Studios analysis ............................................................................. 184
5.3.9 Statistical analysis ....................................................................................... 185

5.4 Results ............................................................................................................ 185
5.4.1 Changes in gene expression define developmental stages ....................... 185
5.4.2 C and N assimilation genes are substantially up-regulated in phase 2 ........ 189
5.4.3 Genes associated with ascorbate and glutathione metabolism are differentially regulated ......................................................................................... 191
5.4.4 Gene set enrichment analysis identifies possible regulatory factors controlling leaf development ........................................................................................................ 191
5.4.5 Clustering of differentially expressed genes support developmental timeline of flag leaves ........................................................................................................... 192
5.4.6 Genes involved in epigenetic regulation were differentially expression between sampling dates .................................................................................................................. 195
5.4.7 Transport processes are enhanced during senescence ............................ 197
5.4.8 Senescence and growth processes are differential through seasonal
development ............................................................................................................. 199

5.4.9 Organic acids: the carbon currency from breakdown of lipids during senescence ............................................................................................................. 201

5.4.10 Roles for NAC transcription factors in remobilization ................... 201

5.5 Discussion ........................................................................................................ 203

5.6 Acknowledgements ........................................................................................ 211

5.7 Works Cited ..................................................................................................... 213

CHAPTER 6: SUMMARY AND CONCLUSIONS .................................................. 225

6.1 Final Discussion ............................................................................................... 226

6.2 Ongoing Work ................................................................................................. 232

6.3 Works Cited ..................................................................................................... 239

7 Appendix A Body of Work .................................................................................. 242

7.1 Genetic background impacts soluble and cell wall-bound aromatics in brownmidrib mutants of sorghum ................................................................. 242

7.2 A Nonsense Mutation in a Cinnamyl Alcohol Dehydrogenase Gene Is Responsible for the Sorghum brown midrib6 Phenotype ..................... 244

7.3 A Continuous, Quantitative Fluorescent Assay for Plant Caffeic Acid O-Methyltransferases ...................................................................................... 246

7.4 Identification and Characterization of Four Missense Mutations in Brown midrib 12 (Bmr12), the Caffeic O-Methyltransferase (COMT) of Sorghum ............................................................................................................... 248

7.5 Next-Generation Sequencing of Crown and Rhizome Transcriptome
7.6 Towards uncovering the roles of switchgrass peroxidases in plant processes ................................................................. 252
7.7 Global changes in mineral transporters in tetraploid switchgrasses (Panicum virgatum L) .................................................. 254
7.8 Contrasting Metabolism in Perenniating Structures of Upland and Lowland Switchgrass Plants Late in the Growing Season .......... 256
7.9 Switchgrass (Panicum virgatum L) flag leaf transcriptomes reveal molecular signatures of leaf development, senescence, and mineral dynamics ............................................................. 258
7.10 Characterization of Greenbug Feeding Behavior and Aphid (Hemiptera: Aphididae) Host Preference in Relation to Resistant and Susceptible Tetraploid Switchgrass Populations .................................. 260
7.11 Transcriptional responses of tolerant and susceptible soybeans to soybean aphid (Aphis glycines Matsumura) herbivory ....................... 262
7.12 Transcriptional Profiling of Resistant and Susceptible Buffalograsses in Response to Blissus occiduus (Hemiptera: Blissidae) Feeding .................................................................................. 264
Chapter 1

Introduction
1.1 Switchgrass

Switchgrass (*Panicum virgatum* L.) is a perennial C₄-grass native to North America and has been used as a warm-season forage crop in the Great Plains and Midwestern United States since the 1940s [1]. In 1992, switchgrass was designated as a model herbaceous energy crop by the United States Depart of Energy’s Bioenergy Feedstock Development Program (BFDP) [2]. Key factors leading to this BFDP designation include a broad adaptation range from the Atlantic Ocean to the Rocky Mountains and from southern Canada to northern Mexico, good biomass yields, ability to grow on marginal lands, and relatively simple seed processing [3]. Other switchgrass characteristics supporting its designation as a model herbaceous energy crop include long term (>10 years) productivity, low nutrient and water requirements, high net energy yields, and positive environmental benefits (such as wild-life habitat, reduced soil erosion, improved soil quality, reduced greenhouse gas emissions, and carbon sequestration) [2, 4-11].

There are two major ecotypes of switchgrass, upland and lowland [12], which diverged from a common ancestor approximately 0.5 to 0.8 million years ago based on modeled genomic information [13]. Lowland varieties of switchgrass are localized in southern latitudes and wetter soils while upland varieties are localized in northern latitudes and somewhat more arid environments. In addition to ecotype differences, switchgrass plants also have either a tetraploid or an octoploid genome [12, 14, 15]. Lowland plants have tetraploid genomes, while upland plants can be either tetraploid (2n = 4x = 36) or octoploid (2n = 8x = 72) [12]. Lowland varieties have higher biomass yields than upland varieties in general [16-19]. Switchgrass is a cross pollinating species
and is largely self-incompatible [20]. Lowland and upland plants are able to intermate, but crosses between tetraploids and octoploids are not viable [20].

Latitudinal adaptation is present in switchgrass populations [19, 21, 22]. The most obvious impact of latitudinal adaptation is observed in flowering times. Photoperiod is a key factor in the transition from vegetative to reproductive growth in switchgrass, as switchgrass is short-day dependent [23-25]. Switchgrass populations natively from northern latitudes flower much earlier when moved south, and conversely, switchgrass populations natively from southern latitudes flower much later, or not at all, when moved north [26, 27]. Most switchgrass populations are unable to be moved more than one hardiness zone north or south of their place of origin as a result of this latitudinal adaptation [19].

Traditional breeding methods for perennial grasses such as switchgrass may span the course of a decade or more. Vogel and Burson [28] outlined a typical breeding program as consisting of four phases (germplasm evaluation nursery, recurrent selection breeding program, regional small plot, trials of advanced lines), each of which lasting approximately five years. While the time required can be decreased if starting populations or cultivars are already well characterized, it is still a lengthy process. One way to significantly shorten the duration of this selection cycle is through the use of marker assisted and genomic selection, which has become a viable more route with the recent expansion of switchgrass genomic resources (see Section 1.4.2 Switchgrass Sequence-Based Resources) [29]. Once markers for a trait of interest are identified, the selection process is accelerated due to the fact that phenotypic scoring of all progeny plants in the field for several years is no longer required. A leaf clip from a seedling is all
that is necessary to score progeny for the presence of desired markers. The application of 
marker assisted and genomic selection have recently been demonstrated in switchgrass. 
Serba et al. [30] identified 11 genomic regions containing quantitative trait loci (QTLs) 
associated with biomass yield and plant height through analysis of a pseudo-F_1 
population resulting from cv. Alamo and cv. Summer crosses. Similarly, Lipka et al. 
analyzed a diverse association panel of 515 switchgrass plants (including lowland, upland 
tetraploid, and upland octoploid) with a set of nearly 17,000 SNPs (single nucleotide 
polymorphisms) to predict seven morphological traits and 13 biomass quality traits [29]. 
Consistent biomass yield per hectare is the most important trait when considering 
the economic viability of switchgrass as a biomass crop [9, 31]. However, yield is a very 
complex trait and heritability in progeny can be difficult to predict [32]. Secondary traits 
which are easier to phenotype such as plant height, leaf width, and heading date can be 
used to increase biomass yield [32]. Sustainable biomass yield, which is a function of 
minimal stand loss each year, is a central aspect for economically producing biomass 
from switchgrass. A high biomass producing population on a per plant basis which 
suffers 25% stand loss over winter each year (winter kill) would make a poor long-term 
biomass crop. Such a phenomenon was demonstrated in the selection of switchgrass 
populations for high in vitro dry matter digestibility (IVDMD) [33, 34]. In these studies, 
researchers were focused on creating populations having better forage properties as 
determined by IVDMD. While they succeeded increasing the IVDMD of select 
populations, they observed an unanticipated side effect of their selection process as an 
increase in the rate of winter kill [33, 34]. These studies highlight how occurrence of 
winter kill is an essential consideration when developing switchgrass populations as
biomass feedstock crops for regions with large and unpredictable temperature fluctuations. Whereas the exact causes of winter kill are not well known, incomplete preparation and transitioning of perenniating tissues (below ground crowns and rhizomes) to dormancy seems to be a likely factor.

The molecular aspects of dormancy signaling and onset are not well known in perennial grasses (see Section 1.2.2 Dormancy). Understanding the molecular processes essential to dormancy in switchgrass will provide additional phenotypic markers for selection of plants with improved winter survival. Enough genotypic and phenotypic variation exists within switchgrass populations and cultivars that it is possible to generate new experimental strains with significantly shifted timing for developmental processes. For example, experimental strains with significant divergence from a base population have been developed by the USDA-ARS in Lincoln, NE [Vogel; unpublished]. One of these experimental strains resulted from selection on the upland tetraploid cultivar Summer for high yielding and late maturing plants. Comparison of R1 heading dates between the base population and the experimental strain Summer Late Maturing, High Yield C2 indicated that the average heading date was shifted nearly a week later in the new population (Figure 1A). A similar selection process from a lowland tetraploid cultivar Kanlow for earlier maturing and high yielding plants resulted in an experimental strain Kanlow N1 Early Maturing, High Yield with heading dates shifted 11 days earlier (Figure 1B) as compared to the base population. Together these results highlight the ability to alter flowering time within a population through phenotypic selection, implying that similar strategies can be applied to other developmental events such as senescence and dormancy onset.
Figure 1. **Heading date shifts in experimental strain populations.** (A) The shift towards later heading in the experimental strain Summer Late Maturing, High Yield C2 (red) compared to the original cultivar Summer (black). (b) The shift towards earlier heading in the experimental straight Kanlow N1 Early Maturing, High Yield (red) compared to the original cultivar Kanlow (black). These data were collected from individual plants in a nursery located near Mead, NE. Summer plant data was collected in 2014 and Kanlow plant data was collected in 2011.

Extending the vegetative growth period for switchgrass by delaying the transition to reproductive growth or onset of above ground tissue senescence is one strategy which can be used to increase switchgrass biomass yields. Newell [26] observed that southern switchgrass populations flowered later, and as a result produced more biomass, when moved to more northern latitudes. However, significant winterkill often accompanied such transplants [19, 36]. Therefore different strategies would need to be applied for the selection of plants that can maintain desired high biomass yields from year to year in combination with high winter survival.

Winter dormancy in perennial grasses is poorly studied, making molecular phenotyping for end of year developmental processes difficult. Obtaining a better cellular and molecular picture of processes involved in switchgrass dormancy will enable more rapid selection for plants with improved winter survival.
1.2 Perennial Senescence and Dormancy

Flowering, senescence, and dormancy are interlinked developmental processes in perennial grasses such as switchgrass. In general, flowering and subsequent development of seeds initiate the processes leading to senescence of monocarpic tissues and coincident or senescence related signals initiate the onset of dormancy preparation in perenniating tissues. Specifically in switchgrass, this results in the senescence of all above ground tissues (tillers and leaves) coupled with dormancy onset in below ground tissues (crowns and rhizomes). Molecular signaling and related processes involved in senescence have been extensively investigated in a broad range of plant species. However, significantly more investigation into winter dormancy related processes is needed, especially for perennial monocots.

1.2.1 Senescence

In essence, senescence is a regulated process of cell death whereby accumulated nutrients in a tissue are recycled to other (sink) tissues. In monocarpic plants the sole sink tissue are seeds while in perennial plants sink tissues include seeds as well as storage and perenniating structures [37]. Plant senescence can be induced by endogenous and exogenous signals including age and stage of plant development, phytohormones, metabolites, light, pathogens, and abiotic stresses [37]. While molecular mechanisms initiating senescence can vary, once started, the process follows a more or less conserved route. NAC (NAM, ATAF1/2, and CUC2) and WRKY transcription factors (named for their characteristic protein domain sequence WRKYGQK) are key regulators of senescence processes. In the model plant Arabidopsis thaliana (At), two NAC transcription factors, AtANAC029 [38] and AtANAC092 [39], play pivotal roles in
promoting senescence. Similarly, WRKY genes including $AtWRKY6$ [40, 41], $AtWRKY53$ [42], and $AtWRKY22$ [43] have been shown to positively regulate senescence while, on the other hand, $AtWRKY54$ and $AtWRKY70$ [44] negatively regulate senescence. In addition to known transcription factors, gene expression studies in many plant species have resulted in the generation of senescence associated gene (SAGs) datasets. A majority of these SAGs have been combined in the Leaf Senescence Database (http://www.eplantsenescence.org/) which includes data from *Arabidopsis*, maize (*Zea mayes*), rice (*Oryza sativa*), wheat (*Triticum aestivum*), banana (*Musa acuminata*), and several other species [45].

1.2.2 Dormancy

Studies on the molecular aspects of dormancy in warm-season perennial grasses are spare or absent [46]. There are three states of plant dormancy as described by Lang et al. [47]: (1) paradormancy where signals from the above ground plant shoots prevent the growth of bud initials below ground, (2) ecodormancy where the environment prevents bud growth due to harsh conditions such as drought and low temperatures, and (3) endodormancy where signals from plant itself prevent further growth until sufficient exposure to low temperatures has occurred. The majority of current research on perennial dormancy has been done on woody dicots. However, in such woody systems, new vegetative growth initiates from buds which are located on above ground stems and branches. In switchgrass, and other herbaceous perennials, these above ground tissues seasonally senesce, and instead, all vegetative regrowth initiates from buds originating on below ground crown and rhizome tissues. However, extensive work has been performed on mechanisms impacting dormancy in leafy spurge (*Euphorbia esula* L.)
Although leafy spurge is an herbaceous perennial, data on the molecular aspects of dormancy obtained from this species can inform similar studies with switchgrass.

Initial investigation of perenniating tissues in leafy spurge focused on analysis of carbohydrate levels in crown buds. Anderson et al. [49] uncovered an inverse relationship in the amount of starch and sucrose across two growing seasons. Starch levels consistently increased throughout the growing season until their peak in mid-to-late October, when they significantly fell and remained low through the winter months [49]. Conversely, sucrose levels were relatively low through the summer months, and significantly increased in the fall, remaining high throughout the winter months [49]. The authors [49] proposed that the increased sucrose levels induced an endodormant state, preventing further growth. Chao et al. [50] tested this hypothesis in subsequent experiments by hydroponically treating decapitated plants with sucrose as well as phytohormones. These authors [50] demonstrated that the addition of 30mM sucrose was able to halt bud growth while micro-molar levels of gibberellic acid (GA) added with sucrose was sufficient to maintain bud growth. Testing other phytohormones also revealed that 4µM abscisic acid (ABA), 2.5 µM 1-naphthaleneacetic acid (NAA, a synthetic auxin), and 1 µM 6-benzylaminopurine (BA, a cytokinin) all inhibited bud growth [50]. These data indicated a causal relationship between sugar levels and phytohormones in determining the developmental state of buds.

Microarray analysis of an initial leafy spurge-based microarray chip, containing around 1,500 unigenes, enabled broader gene expression investigation into the dormancy process [51]. Horvath et al. [52] used this microarray to query gene expression changes occurring with the onset of endodormancy in leafy spurge. These authors [52] observed a
down-regulation of GA responsive genes along with an up-regulation of dehydration/ABA-responsive and cold-regulated genes, consistent with their earlier studies on sugars and phytohormones. Subsequent microarray experiments with improved chips, containing >23k unigenes, yielded a more detailed understanding of the dormancy process in leafy spurge [53]. Genes similar to the Arabidopsis genes *EARLY FLOWERING 4, GIGANTEA, CONSTANS, FALVIN BINDING KELCH REPEAT F-BOX1, CIRCADIAN CLOCK ASSOCIATED1, PHYA*, and several *PSEUDO-RESPONSE REGULATORS* were all up-regulated during dormancy progression [53]. In addition, a set of DAM (dormancy associated MADS-box) transcription factors, previously identified as mandatory for dormancy induction in peach [54, 55], were also up-regulated during dormancy onset in leafy spurge [53]. Further microarray experiments revealed that the transcription factors encoded by *ABI1, ABI4, DREB1a, ETR1, MAPK, MYB4, CCA1, COII, and MYC2* were regulatory hubs for genes differentially expressed in the para- to endo-dormancy transition [56].

More recent research on leafy spurge has highlighted the roles of temperature, photoperiod, and dehydration on endodormancy induction. Dogramaci et al. [57] demonstrated that low temperature alone was insufficient to induce endodormancy in crown buds in growth chamber settings; reduced photoperiod was required as well. Gene expression analysis by microarray revealed opposite expression patterns between reduced temperature versus reduced photoperiod treatments for genes involved in auxin stimulus and transport, flavanoid biosynthesis, ABA stimulus, photoperiod sensing, regulation of flower development, and regulation of transcription [57]. Further work by Dogramaci et al. [58] revealed that 14 days of dehydration in a greenhouse setting was sufficient to
induce endodormancy in leafy spurge crown buds. Metaanalysis of previously published expression data along with their most recent experiments resulted in hypothesis that genes with high homology to Arabidopsis \textit{HY5}, \textit{MAF3}, \textit{RVE1}, and \textit{RD22} may be molecular markers for endodormancy in leafy spurge [58].

Much of the significant amount of research done on various aspects of dormancy has been done on dicots. Since the divergence of monocots and dicots from a common ancestor occurred 140-150 million years ago [59], it is likely that there are many monocot-specific dormancy mechanisms. In depth analysis of the molecular aspects of dormancy in switchgrass may serve to highlight shared and unique features of this developmental process.

\section*{1.3 Sequencing technologies}

\subsection*{1.3.1 Sanger Sequencing}

The first effective DNA sequencing methodologies were initially developed by Dr. Frederick Sanger starting in the early 1970s [60-62]. Sanger’s method of sequencing using chaining terminating nucleotides [63] was the core sequencing approach for over 30 years and is still used today. While there have been massive improvements to the Sanger approach since its introduction (fluorescent labeling instead of isotope labeled nucleotides [64, 65], thermostable polymerases and PCR [66, 67], transition from slab gel to capillary gel electrophoresis [68]), the heart of the procedure remained the same. However, in the mid-2000s, the DNA sequencing field underwent a massive change with the advent of “Next Generation” sequencing (NGS) technologies. Two separate methodologies were developed at approximately the same time which started a new genomics era.
1.3.2 Roche 454 Sequencing

The first system, Roche/454 FLX Pyrosequencer, was introduced commercially in 2005 [69]. Pyrosequencing works by measuring the pyrophosphate released upon incorporation of a nucleotide by DNA polymerase. The presence of pyrophosphate ultimately produces light through the action of luciferase coupled reactions [69, 70]. The amount of light released is directly proportional to the amount of pyrophosphate present, and therefore the number of nucleotides added by DNA polymerase. This reaction carried out on single templates in separate wells in a microplate containing approximately 1.6 million wells forms the basis of 454 pyrosequencing [69, 71]. In a typical 454 sequencing run, a single nucleotide is added to the plate and DNA polymerase is allowed to incorporate it in a template dependent manner. An imaging step then occurs allowing the system to detect the number of times a specific nucleotide was incorporated in each well in the plate [69]. The plate is then washed, a second single nucleotide is added, and the imaging cycle is repeated. The cycle of adding all four nucleotides individually repeats multiple times yielding significant sequencing data [69]. The first 454 sequencing instrument yielded approximately 200,000 100-150 bp long reads per run [69, 72]. In 2008 the 454 GS FLX Titanium system was released and sequencing yields increased to 1 million 500-700 bp long reads [72, 73].

1.3.3 Illumina Sequencing

The second key NGS system was the Illumina (Solexa) Genome Analyzer which was released in 2006 [74, 75]. Illumina sequencing is founded on a sequence-by-synthesis approach [76]. In this method, individual nucleotides are labeled with a specific fluorescent dye as well as a blocking group. A template of DNA or
complementary DNA (cDNA) is prepared from target tissues, sheared to specific size and purified. The purified DNA is ligated with known adapter sequences (unique to the 3' and 5' ends) and covalently attached randomly within a flow cell of the instrument. Subsequently, each individual oligonucleotide sequence is amplified through bridge amplification to create a large number of clonal copies, generating clusters at fixed locations [70]. The reverse strands are cleaved and washed from the flow cells. At this point the flow cell contains hundreds of clonal copies of the input library of DNA sequences originating from the target tissues. Next a mixture of all four nucleotides (singly-labeled with a unique fluorescent dye) is added along DNA polymerase. Because each nucleotide is blocked, the DNA polymerase can only add one nucleotide at a time. After a wash step, an image is taken and the color of light detect indicates which base was added at each location (specific to an individual target sequences). The fluorescent dye and blockings group are then removed and the cycle is repeated. Ultimately, the light signal originating from each specific location on the flow cell is analyzed and converted back to a DNA sequence. These DNA sequences are analyzed by downstream bioinformatics pipelines, often by mapping to existing genomes or de novo assembled scaffolds (genomic or transcriptomic). The first version of the Illumina Genome Analyzer generated nearly 25 million 35 bp long reads, or nearly 1 gigabases (Gb) per run [72, 73]. This yield has increased multiple times, by orders of magnitude, since the release of the initial Genome Analyzer. The release of the GAIIx system in 2008 could yield nearly 85 Gb/run in 100 bp long reads, while the HiSeq2000 system released in 2010 is able to generate 600 Gb/run in 150 bp long reads [72]. The newest Illumina System (HiSeq4000) is able to generate 1500 Gb/run in 150bp long reads [73].
1.3.4 Plant Genomics Era

The introduction of these new sequencing systems (among others including SOLiD by Applied Biosystems [77], Personal Genome Machine (PGM) by Ion Torrent [78], and PacBio RS by Pacific Biosciences [79]) has opened a flood gate in the generation of sequencing data, as well as massively reduced the actual cost of acquiring sequencing data. Starting in 2001, it cost approximately $5,000 to generate 1 megabase (Mb) of sequence data using Sanger methods and with the improvements in high throughput sequencing technologies this cost has been reduced to just $.05 / Mb in 2014, a five orders of magnitude decrease (Line Graph, Figure 2) [80]. This decrease in sequencing costs has led to a massive increase in the number of published plant genomes (Bar Graph, Figure 2) [81-113]. The first sequenced plant genome was from Arabidopsis thaliana which was completed in 2000 [114], followed by Rice (Oryza sativa) in 2002 [115] using Sanger sequencing technologies. After these two genomes, there were minimal additions until the introduction of NGS systems which initiated a true plant genomics era. Starting in 2005, there has been a continual increase in the number of published plant genomes, and as of 2014, the total number of published plant genomes has nearly reached 100 (Figure 2). The genomics era has benefited switchgrass research as well.

1.4 Molecular and Bioinformatic Research in Switchgrass

Numerous molecular and bioinformatic research articles have been published on switchgrass ever since its designation as a model herbaceous energy crop by the United States Depart of Energy’s Bioenergy Feedstock Development Program (BFDP) [2]. The focus of these publications can generally be classified into five categories: (a) sequence-
based data (b) genomic marker data, (c) plant transformation related methodologies, (d) phenylpropanoid biosynthesis pathway related, and (e) transcriptomic and other general molecular research.

Figure 2. Cost of sequencing with the number of published plant genomes on a per year basis. The line graph shows the decrease in the cost of sequencing per Mb along the left y-axis (log scale). The bar graph shows the number of published plant genomes per year along the right y-axis.

1.4.1 Switchgrass Sequence-based Resources

Whereas the development of switchgrass as a feedstock for biofuel production has been ongoing since the 1990s, it is only in the last decade that significant genomic resources have become available to assist in this endeavor. Much of the genomic resources have primarily generated through U.S. Department of Energy (DOE) projects. In 2005, the first set of expressed sequence tags (ESTs) were generated by USDA-ARS
scientists and their collaborators from different tissues in cv. Kanlow [116]. This set of ~12,000 Sanger-sequenced ESTs generated approximately 8,000 unigene sequences from switchgrass tissues (callus, crown, leaf, and stem). In 2008 this EST library was expanded with an addition ~65k ESTs in a collaborative project with DOE-JGI (Joint Genomes Institute) [117] and later a further addition of >400,000 ESTs originating from stem apex, early floral bud, etiolated seedling, late flowering bud, and root tissues [unpublished, publically available data]. The first advancement beyond Sanger-sequencing based resources in switchgrass occurred with the creation of de novo transcriptome assemblies using 454 generated sequences. The first such assembly was created with 930,000 sequences from cv. Summer crown and rhizome tissue [118] and subsequently on 980,000 total sequences from cv. Cimarron seed, seedling, tiller, and flower tissues [119]. Researchers at the Samuel L. Noble foundation have since created their own extensive transcript-based resources [120]. Using primarily 454 sequencing on cDNAs from multiple tissues, they generated approximately 1.5 million ESTs from two switchgrass populations (cv. Alamo and cv. Summer). These ESTs were combined with other publically available switchgrass ESTs [116, 117] to create a set of approximately 128,000 hypothetical transcripts, titled PviUTs (Switchgrass UniTranscripts) [120]. Expanding beyond a mere transcript database, these researchers then used their transcript database to develop an Affymetrix cDNA chip and used this microarray system to develop a broad switchgrass gene expression atlas (PvGEA) [120].

While these initial projects provided extremely useful resources, they were still primarily transcript based, and therefore only provided a small percentage of sequence information found in the entire estimated 1500 MB switchgrass genome [121-123]. The
next big advance in switchgrass genomics occurred in 2011 with the release of an early version of the switchgrass genome (Pvi0, www.phytozome.org) [85]. This first draft genome was based on 454 generated sequences from a lowland ecotype plant (cv. Alamo) which were assembled into approximately 410,000 thousand contigs, with a total size of 1,358 Mb, and included 65,878 putative gene models. Moving from transcript based sequencing to genome based sequencing resulted in the ability to process the large volumes of sequencing data provided by NGS platforms, and permitted the transition to more powerful genome-based bioinformatics methods to query switchgrass. Such approaches include transcriptome analysis (RNA-Seq), Genotype-by-Sequencing (GBS), Genome Wide Association Studies (GWAS), and more complete marker assisted breeding [124, 125]. The second version of the switchgrass draft genome (Pvi1.1, www.phytozome.org) [85] was released in the fall of 2013 and included improvements in gene annotation and in the combining of over half of the genomic sequence into chromosomes through the use of a genetic marker map generated from a cv. Alamo by cv. Summer cross. Work is still ongoing for improving the assembly and annotation of the switchgrass genome, and version 2 is expected to be released in late 2015. This third iteration of the switchgrass genome will leverage more thorough linkage maps, genomic sequence of a closely related species (Panicum halii), and extensive RNA-Seq.

1.4.2 Genomic Markers

Genomic markers distinguishing various switchgrass populations have been used extensively in switchgrass research. The first of these markers identified was a polymorphism in chloroplast DNA which was able to differentiate between upland or lowland ecotopes, using Southern blots and a unique BamHI restriction site near the
RUBISCO large subunit coding site in lowland populations [12]. Martinez-Reyna et al. [126] used this chloroplastic DNA (cpDNA) marker to demonstrate the maternal inheritance of cpDNA in switchgrass through reciprocal upland and lowland crosses. Missaoui et al. [127] reported the first genetic map of switchgrass by analyzing the segregation of 224 restriction fragment length polymorphism (RFLP) markers in full-sib progeny from a Summer X Alamo (upland X lowland) cross. In subsequent work, Missaoui et al. [128] applied 85 of their RFLP markers in analysis of plants from one upland cultivar (Summer) and two lowland cultivars (Alamo and Kanlow). These markers were able to clearly distinguish between the upland and lowland cultivars, however not between the two lowland cultivars Alamo and Kanlow [128]. Similar to the work by Hultquist et al. [12], Missaoui et al. [128] also investigated possible cpDNA markers for ecotype identification using a PCR and sequencing approach instead of Southern blots. In analyzing the chloroplast trnL (UAA) intron [129], Missaoui et al. [128] discovered that lowland ecotypes had a 49 bp deletion in this intron, allowing for more rapid upland vs lowland characterization by PCR compared to southern blot.

Casler et al. [130] performed the first broad scale marker analysis among multiple switchgrass populations. In this experiment, 125 random amplified polymorphic DNA (RAPD) markers were used to query 818 plants from 11 cultivars and 46 natural prairie populations. Results from this study showed that the switchgrass cultivars could not be distinguished from the natural prairie populations using these markers [130]. Using expressed sequence tag simple sequence repeats (EST-SSRs) instead of RAPD markers, Narasimhamoorthy et al. [131] queried 31 switchgrass populations from the Germplasm Resources Information Network (GRIN), including uplands, lowlands, tetraploids, and
octoploids. Using 63 EST-SSR makers, they were able to distinguish between upland and lowland populations, as well as sub-cluster genotypes into adaptive zones based on the geographic origin of each population [131]. Tobias et al. [117] generated an EST dataset consisting of nearly 62,000 ESTs generated from three different cDNA libraries (crown, seedling, and callus). Through analysis of these ESTs, the authors developed a set of nearly 2,400 potential SSR markers which could be used for switchgrass population characterization [117].

Okada et al. [122] developed the first complete genetic map for lowland, tetraploid switchgrass. Using a total of 637 genomic markers (SSR and Sequence-Tagged Sites (STS)) with 238 F1 progeny plants resulting from an Alamo X Kanlow cross, the authors were able to generate 18 linkage groups, corresponding to the 18 haploid chromosomes found in tetraploid switchgrass [122]. Okada et al. [132] subsequently used 21 SSR markers to separate upland ecotypes from lowland ecotypes, as well as documenting population structure amongst the upland tetraploid and octoploid plants. At approximately the same time as Okada et al. [122] developed their 637 genomic markers, Wang et al. [133] identified 1,030 genomic SSR markers as well, characterized from a mapping population resulting from a cross of two lowlands, “SL93 7x15” and “NL94 16x13.”

The application of genomic markers to differentiate between switchgrass populations and lines has accelerated more recently. Zalapa et al. [134] used genomic markers identified in previous work to analyze plants from 18 different cultivars, including lowlands and upland tetraploids and octoploids. Principal component analysis of their data clearly showed a separation of uplands and lowlands along the first
component axis and a separation between upland tetraploids and upland octoploids along
the second component axis [134]. Liu and Wu [135] developed a duplex SSR screening
system testing 48 SSR markers in 24 reactions, simplifying screening protocols.

Sharma et al. [136] provided the first significant genomic sequence based
potential marker set through the generation and analysis of over 200k bacterial artificial
chromosomes (BACs) from switchgrass. These libraries contained nearly 50 Gb of
switchgrass genomic sequence, and through Sanger sequencing the ends of each BAC,
the authors were able to identify nearly 280,000 known repetitive elements, 50,000 SSRs,
and 2,528 novel repetitive elements [136]. Utilizing EST libraries generated from 13
switchgrass cultivars and NGS technologies, Ersoz et al. [137] were able to identify
nearly 150,000 SNPs. These new resources will be useful to the continued breeding
efforts for developing elite switchgrass lines.

Additional switchgrass genetic linkage maps have been developed from different
plant populations by researchers at Oklahoma State University and the Samuel Roberts
Noble Foundation. In two reports, Liu et al. [138, 139] at Oklahoma State University
identified SSR markers in a unique selfed progeny population derived from a northern
lowland plant, NL94. Meanwhile, Serba et al. [140] at the Noble Foundation identified
upland and lowland specific markers in a Summer (VS16) x Alamo (AP13) full-sib
population. Lu et al. [125] described the first true genomics-era experiment combining
NGS with large and diverse population numbers. Here the authors used reduced
representation sequencing and carried out GBS on an association panel representing
plants collected from 66 locations in the United States, 130 plants resulting from a bi-
parental cross of upland tetraploids, and 168 half-sib plants, also from an upland
tetraploid [125]. Aside from the development of linkage maps containing nearly 90,000 SNPs, one of the most significant conclusions reached by these authors was that tetraploid switchgrass behaves like a diploid when it comes to inheritance of polymorphisms [125].

Finally, two publications document switchgrass marker analysis results which made use of the draft version of the switchgrass genome (Pvi0). Childs et al. [124] isolated RNA from seedlings from seven switchgrass populations (4 upland, 3 lowland), sequenced them on the Illumina HighSeq2000 platform, aligned them to the Pvi0 reference genome, and identified SNPs within these alignments. In total this group identified 1,305,976 SNPs, 438,464 of which were unique to lowland cultivars and 723,678 were unique to upland cultivars [124]. Evans et al. [141] developed an exome capture system in collaboration with Roche-NimbleGen and used this to sequence gene specific regions of the switchgrass genome. To demonstrate the power of this method, exomes isolated from 10 different cultivars of switchgrass were sequenced and aligned to the Pvi0 reference genome. Variant analyses of these alignments resulted in the identification of 1,395,501 SNPs, 8,173 copy number variants, and 3,336 presence/absence variants [141]. A key challenge is to being using these extensive molecular resources to speed the selection process within switchgrass breeding programs.

1.4.3 Switchgrass Transformation Methodologies

Plant transformation is a standard procedure used in molecular biology research to discover gene function and/or develop transgenic plants with new characteristics. Having the ability to stably transform select genes into switchgrass will be very important in its development as a feedstock crop [142]. Somleva et al. [143] reported the first genetic
transformation of switchgrass using *Agrobacterium*. Callus tissue from cv. Alamo plants were transformed with a plasmid containing the *bar* gene for bialaphos selection, and the *uidA* (*gus*) reporter gene [143]. No further reports on switchgrass transformation occurred until 7 years later when Xi et al. [144] documented *Agrobacterium*-mediated transformations of cv. Alamo with pCAMBIA plasmids containing a hygromycin resistance gene and a GUS reporter gene. However, transformation is not without problems because trans-gene silencing was found to occur at high rates in plants containing multiple transformations, although a reversal of gene silencing in segregating progeny containing only one copy of the transgene was also observed [144]. Transformation efficiencies could be improved through the use of some additives including acetylsyringone, L-cysteine, and dithiothreitol [145].

A comparison of *Agrobacterium*-mediated transformation efficiency for lowland cultivars Alamo, Colony, and Performer indicated that callus derived from Performer had significantly higher efficiency of transformation than for tissues derived from either Alamo and Colony plants [146]. Ramamoorthy and Kumar detailed a simplified transformation protocol that while it has a lower efficiency than Li and Qu’s [146] method, it led to a faster generation of transformed plants and required fewer steps. *Agrobacterium*-mediated multigene transformations have been successful in switchgrass [147]. These researchers suggest that up to 8 genes could be potentially transformed simultaneously with their protocol [147].

One caveat to switchgrass transformations is that they have only been demonstrated to work on lowland populations, primary cv. Alamo, but also cultivars Performer and Colony. Transformations of upland switchgrass ecotypes have failed,
primarily due to the fact that regenerating plants from callus tissue was difficult [148, 149]. However, Liu et al. [150] have recently documented a modified callus regeneration method which yielded regeneration rates of 50% and 76% for two upland cultivars Blackwell (octoploid) and Dacotah (tetraploid), respectively. A comparison of the transformation efficiency of Alamo and Dacotah calluses generated with this new approach demonstrated that Alamo had a transformation efficiency of 73% and Dacotah had an efficiency of 8% [150]. While the efficiency of transformation of the upland Dacotah is nearly an order of magnitude less than the lowland Alamo, it still demonstrates that upland switchgrass populations could be transformed using Agrobacterium. The larger picture utility of this and other protocols remain to be fully validated.

1.4.4 Phenylpropanoid Biosynthesis Pathway

Lignin content in switchgrass has been a focus for research even prior to its designation as a model herbaceous energy crop [151]. High lignin levels negatively impact biochemical conversion of biomass into ethanol and related liquid fuels, and as a result, breeding for reduced lignin levels has been a common approach in improving switchgrass germplasm [152, 153]. A more direct way to reduce lignin is through manipulation of the phenylpropanoid biosynthesis pathway [154, 155]. The suite of enzymes compromising the phenylpropanoid pathway generate a number of aromatic precursors that are substrates for a number of other metabolic routes. In the pathway leading to lignin biosynthesis, the end products are alcohol monomers (monolignols) which are transported from the cytosol to the cell walls were they polymerize to form lignin. Lignin is composed of primarily of three types of units, named H-lignin, G-lignin,
and S-lignin. Changes or loss of specific lignin pathway genes/enzymes will impact the levels and relative ratios of the three types of lignin monomers. Both natural and engineered mutations in the pathway have been utilized to understand the specific roles of individual genes encoding for enzymes and/or transcription factors in lignification. As limited examples, in alfalfa (*Medicago sativa* L.) transgenic methods were used to down-regulate individual enzymes in this pathway [156]. In sorghum (*Sorghum bicolor*), nonsense and missense mutations in three enzymes in the phenylpropanoid biosynthesis pathway have resulted in reduced lignin levels (bmr6 [157], bmr12 [158], bmr2 [159]). Invariably, reduced lignin levels lead to increased saccharification of cell wall carbohydrates and improved ethanol conversion efficiencies [156, 159, 160].

Studies with switchgrass have largely corroborated work done with other plants. In switchgrass, plants selected for lowered lignin or mutated for a loss in a specific gene displayed significantly enhanced biochemical conversion of biomass into ethanol. Escamilla-Trevino et al. [161] analyzed the cinnamoyl CoA reductase-like (CCR) gene family in switchgrass. These authors identified four CCR like genes, two of which (*PvCCR1* and *PvCCR2*) demonstrated actual CCR activity [161]. After investigation of the expression patterns of *PvCCR1* and *PvCCR2*, the authors concluded that *PvCCR1* is most likely associated with lignin biosynthesis (and therefore a potential knockdown target) while *PvCCR2* appeared to have defense related functions [161]. Similarly to CCR, Saathoff et al. [162] reported the presence of two related cinnamyl alcohol dehydrogenase (CAD) genes in switchgrass (*PviCAD1* and *PviCAD2*). While kinetic characterization of both CADs revealed similar activity on appropriate phenylpropanoid biosynthesis pathway substrates, expression analysis showed *PviCAD1* had significantly
higher expression levels (by over an order of magnitude) in stem tissue, and therefore PviCAD1 is most likely key in monolignol biosynthesis [162]. In subsequent research, Saathoff et al. [163] carried out a more thorough activity survey of PviCAD1 by using site-directed mutagenesis to identify specific residues key for enzyme activity.

_Agrobacterium_-mediated transformations of RNAi constructs for several phenylpropanoid biosynthesis genes have been used to modify switchgrass lignin content. Fu et al. [164] targeted a caffeic acid O-methyltransferase (_COMT_ ) gene for down-regulation in switchgrass due to the positive outcomes of similar approaches in alfalfa [165]. Analysis of their down-regulated plants revealed a large reduction in S-lignin levels as well as a 38% increase in ethanol yield compared to wild-type plants [164]. Xu et al. [166] focused on the enzyme 4-coumarate:CoA ligase (4CL) for similar reasons. These authors identified two homologous 4CL genes in switchgrass (Pv4CL1 and Pv4CL2) and expression and activity analysis suggested Pv4CL1 was involved in monolignol biosynthesis [166]. RNAi mediated down-regulation of Pv4CL1 resulted in plants with improved saccharification efficiency and approximately a 50% increase in glucan yield with dilute acid pretreatment [166]. Cinnamyl alcohol dehydrogenase (CAD) was selected for down-regulation by two different sets of researchers. Saathoff et al. [167] analyzed CAD down-regulated lines and demonstrated these lines had reduced CAD protein levels, reduced CAD activity, reduced lignin levels, and increased glucose release. Fu et al. [168] also generated CAD down-regulated plants and observed similar results as Saathoff et al. [167] in reduced CAD activity and increased sugar release during saccharification. One unique observation made by Fu et al. [168] was a significant increase in chlorogenic acid (a phytoalexin) levels, suggesting a reallocation
of phenolic precursors to produce phytoalexins due to reduced monolignol biosynthesis.

Many plant cells contain primary and secondary cell walls. As tissues mature, secondary cell walls are formed. In many types of cells, such as sclerenchyma and xylem, the secondary cell walls can become extensively lignified. Although the enzymes of the phenylpropanoid pathway produce the monolignols needed for lignin deposition, a number of other proteins ultimately influence this process. Transcription factors are the master regulators of the cell’s transcriptomic machinery and thereby impact all aspects of cell development. Several transcription factors have been identified in model plants that directly impact lignification. Among these transcription factors, several NAC transcription factors genes in *Arabidopsis thaliana* which control cell wall and lignin levels have been identified. The knockout or overexpression of these genes in *Arabidopsis* lead to significant changes to cell wall structure and composition (for example *AtXND1* [169]; *AtANAC101* and *AtVND7* [170]; *AtNST1*, *AtNST2*, and *AtNST3* [171, 172]; *AtSND2* and *AtSND3* [173]). The importance of NACs to cell well structure and composition led Shen et al. [174] phylogenetically classified NAC genes in switchgrass along with 10 other plant species. The overall goal was to identify potential switchgrass homologs to Arabidopsis and other NACs with known impacts on lignin content and cell wall structure. These switchgrass NACS would then be good candidate genes for manipulating cell walls in switchgrass. Work using transgenics to understand the function(s) of switchgrass NAC genes have yet to appear in literature.

Shen et al. [175, 176] created an over-expression (OE) line for the transcription factor *PvMYB4*. Over-expression of this transcription factor in switchgrass resulted in an overall suppression of lignin biosynthesis through down-regulation of enzymes in the
phenylpropanoid biosynthesis pathway including PAL, C4H, C3H, COMT, F5H, 4CL, CCoAOMT, CCR, CAD, and HCT [175]. This reduction of lignin lead to a slight increase of cell wall polysaccharide content and approximately a 3-fold increase in sugar release efficiency [175, 176].

In a manner similar to Shen et al.’s [175, 176] approach to modifying switchgrass through the overexpression of PvMYB4, Xu et al. [177] selected an Arabidopsis NAC transcription factor (LONG VEGETATIVE PHASE ONE, AtLOV1) to overexpress in switchgrass plants. AtLOV1 has been found to improve cold tolerance through induction of COLD-REGULATED 15A (COR15A) and COLD INDUCED 1 (KIN1) and delay flowering time through repression of CONSTANS (CO) and FLOWERING LOCUS T (FT) in A. thaliana [178]. Xu et al. [177] reported that ectopic overexpression of AtLOV1 in switchgrass plants resulted in increased lignin levels, erect leaves, and delayed flowering time but no improvement in cold tolerance was observed.

As mentioned earlier, in the down-regulation of CAD expression in switchgrass, Fu et al. [168] observed a significant increase in chlorogenic acid levels. The biosynthetic pathway to generated chlorogenic acid in monocots is poorly understood, and therefore Escamillia-Trevino et al. [179] attempted to identify enzymatic routes for production of chlorogenic acid in switchgrass. The key enzyme in the production of chlorogenic acid in dicots is hydroxycinnamoyl-CoA : quinate hydroxycinnamoyltransferase (HQT) [180], however a homologous enzyme was not found in monocot plants [179]. Instead, the authors identified a gene somewhat similar to the phenylpropanoid biosynthesis enzyme hydroxycinnamoyl-CoA shikimate hydroxycinnamoyltransferase (HCT) which they named PvHCT-Like1 [179]. Upon
activity characterization of recombinant PvHCT-Like1 protein, Escamillia-Trevino et al. [179] discovered that PvHCT-Like1 exhibits strong HQT activity and concluded that PvHCT-Like1 probably functions in chlorogenic acid biosynthesis in switchgrass.

Another route for modifying plant development or even specific genes is through microRNAs (miRNAs). MiRNAs are small, noncoding RNAs that play a very important role in a variety of developmental processes by inhibiting gene expression through complementary base pairing to target mRNAs, leading to mRNA degradation. Xie et al. [181] published the first survey of miRNAs in switchgrass and identified 121 potential miRNAs belonging to 44 families through alignment of known miRNAs in other plant species and to publically available switchgrass ESTs. In a similar time frame, Matts et al. [182] identified 34 miRNAs in a direct sequencing approach by isolating small RNAs from cultivar Alamo plants and sequencing using 454 technology. Xie et al. [183] undertook a more thorough analysis of switchgrass miRNAs from seedlings under drought and salt stress. The authors used Illumina sequencing on small RNA libraries to generate 50 million reads which were then mined for the presence of potential miRNAs [183]. A total of 273 miRNAs were identified in this approach, 126 of which were conserved among other plant species and 147 of which were novel [183].

Chuck et al. [184] transformed a known miRNA from maize (Corngrass1) into switchgrass in order to change physical characteristics of the plant. Corngrass1 (Cg1) overexpression has been shown to prevent plant development from progressing past a juvenile phase as well as having decreased lignin and increased glucose levels [185, 186]. The authors reported that expression of Cg1 in switchgrass resulted in plants with 250% increase in starch levels, reduced biomass yield on high expression lines, and complete
inhibition of flowering [184]. Fu et al. [187] targeted a similar miRNA, miR156 from rice, for overexpression in switchgrass. These authors observed similar phenotypes as Chuck et al. [184] in high miR156 expression lines, but in lower expression lines, overexpression plants were able to actually flower and had nearly a 50% increase in above ground biomass due to an increased number of tillers per plant [187]. In general, global down-regulation of pathways in plants leads to unpredictable phenotypes, invariably with much poorer agronomic traits as compared to wild type plants.

1.4.5 Transcriptomic and other General Molecular Research

Development of a switchgrass specific Affymetrix gene chip [120] enabled researchers to query more specific aspects of switchgrass development. Li et al. [188] used this system to measure the effect of prolonged heat stress on cv. Alamo plants in a growth chamber setting. The authors observed that prolonged heat stress resulted in smaller plants, an induction of 2,002 genes, and a repression of 2,809 genes [188]. Induced genes included heat shock genes, other chaperone genes, and oxido-reductase genes while repressed genes included all glycolysis genes, nitrogen and phosphorus metabolism related genes, and a wide range of transcription factors [188]. Wang et al. [189] used the switchgrass Affymetrix chip to query gene expression differences between buds and nodes from high tillering and low tillering field grown plants. 750 genes were identified as being consistently upregulated in both tissues in high tillering plants compared to low, and 390 were consistently down regulated in the same manner [189]. Several WRKY, AP2 (Apetala 2), bZIP (basic leucine zipper), and B3 (three basic regions) transcription factors were among the genes more highly expressed in high tillering lines, but no NAC transcription factors were found in this gene set [189]. Genes
involved with transport (vesicular and lipid) as well as amino acid biosynthesis were upregulated in low tillering plants [189]. The relative contributions of these up/down-regulated genes to observed phenotypes are not known.

The advent of NGS technologies has provided useful tools for switchgrass transcriptomic analyses as well. A first use of this technology was in the de novo assembly of transcriptomes using the Roche 454-FLX system. Palmer et al. [118] published an assembly of transcripts isolated from crowns and rhizomes of an upland, tetraploid cultivar Summer. Around the same time, Wang et al. [119] released their own assembly of transcripts isolated from multiple tissues (seedlings, tillers, flowers, and dormant seeds) from the lowland cultivar Cimarron.

More recently, there have been two publications detailing the first uses of RNA-Seq with Illumina sequencing in analyzing switchgrass transcriptomes, leveraging the switchgrass draft genome. First, Palmer et al. [190] compared the transcriptomes of crowns and rhizomes from field grown upland cv. Summer and lowland cv. Kanlow plants harvested prior to a killing frost. The authors reported that while Summer perenniating tissues appeared to have initiated dormancy related processes while Kanlow tissues were still in a state of growth and concluded that delayed dormancy onset in the lowland cv. Kanlow may be one contributing factor to winter kill observed in the population [190]. In other studies, Palmer et al. [191] investigated gene expression profiles of flag leaves in field grown upland Summer plants as a function of a developmental time series from plant heading through the onset of aerial senescence. RNA-Seq analysis documented the overall changes in the transcriptome during the progression of flag leaf development and transcriptional changes that trigger leaf
senescence. By comparing changes in leaf chlorophyll levels and the expression of genes for chlorophyll biosynthesis and degradation, a four-phase molecular roadmap for switchgrass flag leaf ontogeny was developed. Genes associated with early leaf development were up-regulated in phase 1. In Phase 2, leaves had increased expression of genes for chlorophyll biosynthesis, and those needed for full leaf function. Phase 3 coincided with the most active phase for leaf C and N assimilation. Phase 4 was associated with the onset of senescence, as observed by declining leaf chlorophyll content, a significant up-regulation in transcripts coding for enzymes involved with chlorophyll degradation, and a large number of senescence-associated genes. Of considerable interest were switchgrass NAC transcription factors with significantly higher expression in senescing flag leaves. Two of these transcription factors were closely related to a wheat NAC gene that impacted mineral remobilization. The third switchgrass NAC factor was orthologous to an Arabidopsis gene with a known role in leaf senescence. Other genes coding for nitrogen and mineral utilization, including ureide, ammonium, nitrate, and molybdenum transporters, shared similar expression profiles with the three NAC transcription factors and could be direct targets of these NACs [191].

Finally, the availability of a sequenced and annotated genome for switchgrass has enabled various researchers to conduct systems-level analyses of various gene families. Saathoff et al. [192] mined the Pvi0 draft genome in order to identify class III plant peroxidases which are involved in many plant processes including lignification, response to stress, and herbivory. 263 class III peroxidases were identified in switchgrass, with 33 belonging to the monocot-specific Group V.I clade [192]. Zhao et al. [193] undertook
the analysis of the switchgrass MYB (myeloblastosis) transcription factor family, one of the largest transcription factor families in plants. In other plant systems, MYBs have been shown to regulate secondary cell wall biosynthesis, and therefore are of interest in biomass feedstock research [194-196]. The authors identified a total of 230 switchgrass MYBs through mining the Pvi0 draft genome along with the Switchgrass Functional Genomics Server [120, 193]. In depth phylogenetic analysis of MYBs from switchgrass, Arabidopsis, poplar, rice, and maize resulted in the creation of 48 subgroups, three of which are dicot specific and size being grass specific [193]. Yuan et al. [197] mined the Pvi1.1 version of the switchgrass genome in order to identify members of the CCCH-type zinc finger gene family. This is a very diverse gene family which binds to zinc, DNA, RNA, proteins and lipids and are involved in processes including cell fate specification, stress responses, and hormone regulated signaling [198]. In switchgrass, the authors identified a total of 103 CCCH genes which separated in 21 clades by phylogenetic analysis and had diverse expression in various tissues [197]. In an analysis of mineral transporters in switchgrass, Palmer et al. [199] identified 520 mineral transporters in 40 classes by mining the Pvi0 version of the genome. Expression analysis of transporter dataset based on 454 data from various tissues and developmental states highlighted temporal expression patterns for select transporters which could be key in remobilization of nutrients [199].

The development of the genomic resources discussed above and those currently in development (for example version 2.0 of the switchgrass genome) indicate that molecular studies on switchgrass are entering a period of rapid growth. These future studies, both
applied and basic, are likely to vastly increase the currently knowledge about this important crop and improve their utility in breeding programs.
1.5 Works Cited


25. Benedict HM: **Effect of day length and temperature on the flowering and growth of four species of grasses.** *J Agric Res* 1940, **61:**0661-0671.


61. Sanger F, Donelson JE, Coulson AR, Kossel H, Fischer D: **Determination of a Nucleotide-Sequence in Bacteriophage-F1 DNA by Primed Synthesis with DNA-Polymerase.** *J Mol Biol* 1974, **90**(2):315-&.


63. Sanger F, Nicklen S, Coulson AR: **DNA Sequencing with Chain-Terminating Inhibitors.** *P Natl Acad Sci USA* 1977, **74**(12):5463-5467.


76. Seo TS, Bai XP, Kim DH, Meng QL, Shi SD, Ruparelt H, Li ZM, Turro NJ, Ju JY: **Four-color DNA sequencing by synthesis on a chip using photocleavable fluorescent nucleotides.** *P Natl Acad Sci USA* 2005, **102**(17):5926-5931.


80. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)

81. Michael TP, Jackson S: The First 50 Plant Genomes. *Plant Genome-Us* 2013, 6(2).


98. Leushkin EV, Sutormin RA, Nabieva ER, Penin AA, Kondrashov AS, Logacheva MD: The miniature genome of a carnivorous plant Genlisea aurea contains a
low number of genes and short non-coding sequences. *Bmc Genomics* 2013, 14.


137. Ersoz ES, Wright MH, Pangilinan JL, Sheehan MJ, Tobias C, Casler MD, 
Buckler ES, Costich DE: **SNP Discovery with EST and NextGen Sequencing**
in **Switchgrass (Panicum virgatum L.).** *Plos One* 2012, 7(9).

Repeat-Based Genetic Linkage Map of Switchgrass. *G3-Genes Genom Genet*

139. Liu LL, Huang YL, Punnuri S, Samuels T, Wu YQ, Mahalingam R:
**Development and integration of EST-SSR markers into an established**
linkage map in **switchgrass.** *Mol Breeding* 2013, 32(4):923-931.

Brummer EC, Saha MC, Devos KM: **Linkage Maps of Lowland and Upland**

141. Evans J, Kim J, Childs KL, Vaillancourt B, Crisovan E, Nandety A, Gerhardt DJ,
Richmond TA, Jeddeloh JA, Kaeppler SM et al: **Nucleotide polymorphism and**
copy number variant detection using exome capture and next-generation
sequencing in the polyploid grass Panicum virgatum.** *Plant J* 2014, 79(6):993-
1008.

142. Gressel J: **Transgenics are imperative for biofuel crops.** *Plant Sci* 2008,
174(3):246-263.

143. Somleva MN, Tomaszewski Z, Conger BV: **Agrobacterium-mediated genetic**


152. Sarath G, Akin DE, Mitchell RB, Vogel KP: Cell-wall composition and accessibility to hydrolytic enzymes is differentially altered in divergently


160. Dien BS, Sarath G, Pedersen JF, Sattler SE, Chen H, Funnell-Harris DL, Nichols NN, Cotta MA: \textit{Improved Sugar Conversion and Ethanol Yield for Forage Sorghum (Sorghum bicolor L. Moench) Lines with Reduced Lignin} 


172. Mitsuda N, Iwase A, Yamamoto H, Yoshida M, Seki M, Shinozaki K, Ohme-Takagi M: NAC transcription factors, NST1 and NST3, are key regulators of


molecular signatures of leaf development, senescence, and mineral dynamics.


Chapter 2

Next-generation sequencing of crown and rhizome transcriptome from an upland, tetraploid switchgrass

Note: The results described in this chapter have been published and all text is modified from the original version. Supplemental data can be found online at the journal’s website.

2.1 ABSTRACT

The crown and rhizome transcriptome of an upland tetraploid switchgrass cultivar cv Summer well adapted to the upper-Midwest was investigated using the Roche 454-FLX pyrosequencing platform. In all approximately 1 million reads consisting of 216 million bases were assembled into 27,687 contigs and 43,094 singletons. Analyses of these sequences revealed minor contamination with non-plant sequences (< 0.5 %), indicating that a majority were for transcripts coded by the switchgrass genome. Blast2Go comparisons resulted in the annotation of ~65 % of the contig sequences and ~40 % of the singleton sequences. Contig sequences were mostly homologous to other plant sequences, dominated by matches to the *Sorghum bicolor* genome. Singleton sequences while displaying significant matches to *Sorghum bicolor*, also contained sequences matching non-plant species. Comparisons of the 454 dataset to existing EST collections resulted in the identification of 30,177 new sequences. These new sequences coded for a number of different proteins and a selective analysis of two categories, namely peroxidases and transcription factors resulted in the identification of specific peroxidases and a number of low-abundance transcription factors expected to be involved in chromatin remodeling. KEGG maps for glycolysis and sugar metabolism showed high-levels of transcripts coding for enzymes involved in primary metabolism. The assembly provided significant insights into the status of these tissues, and broadly indicated that there was active metabolism taking place in the crown and rhizomes at the post-anthesis, seed maturation stage of plant development.
2.2 Introduction

Increasingly, there is interest in use of switchgrass as a feedstock for biofuels because it can be effectively grown on marginal croplands [1, 2]. In order to fulfill anticipated biomass demand, improvements in agronomic properties, particularly biomass yields, yield stability, and quality of lignocellulosic materials need to be accomplished by the year 2030 to meet a national goal of replacing 30% of petroleum gasoline with liquid fuels derived from renewables [3]. Accomplishing these goals in the upper-Midwest will also be met with the challenge of sustaining high-productivity from potential cold weather-related losses in stand (plants per m²) over time that could both reduce yields and increase production costs to replant fields.

Sustainable production of switchgrass for biofuels in the upper-Midwest will require cultivars that withstand great fluctuation in temperatures and rainfall. At least two different factors are believed to contribute to switchgrass production under these conditions. The first is the overall health of the below ground components of the plant. Depending on the genotype, switchgrass produces either short or long rhizomes. Each spring, new tillers arise from rhizomes, crowns and axillary buds present on stem bases. There is significant genetic variation for new tiller production, and for the proportion of tiller initials derived from different sources. Thus, breeding efforts will have to capitalize on this diversity to produce cultivars with optimal biological efficiency for tiller meristem initiation and growth. In addition to tillering, a second factor that is related with winter-hardness is lignin. Selection of forages for increased dry matter digestibility (e.g. for animal feed) is accompanied by lowering lignin in plant tissues, but plants bred for lowered lignin also have displayed a loss in agricultural fitness in some genetic
backgrounds [4]. Unfortunately little is known about the underlying reason for this observation. However, several studies have also seen that selection for increased digestibility also negatively impacts winter-hardiness in some switchgrass populations [5, 6].

To aid breeding and selection, molecular markers that are associated with below-ground tissue health are necessary. While genomic biology provides a systematic means for identifying such markers, the transformational step of establishing a whole genome sequence is difficult to realize in plants such as switchgrass that have polyploid genomes, and are likely to contain large families of dispersed repetitive DNA elements [7]. To circumvent this problem, transcriptomes of these plants are generally evaluated by *de novo* sequencing of cDNA to provide a fundamental overview of the coding-capacity of their genomes [for example [8-14]. ESTs from sequencing of switchgrass tissues, including young crowns and roots have been produced and made publically available [11, 15, 16]. However, these ESTs suffer from the limitations of being produced from traditional clone-based libraries and are not from crowns and rhizomes of field-grown plants, especially from a cultivar well adapted to the Upper Midwest of the US. To more systematically characterize the transcriptome of plants relevant to the Upper Midwest, we have capitalized on the capacity of next-generation sequencing technologies that can provide a more comprehensive overview of the transcriptome. In addition to capacity, the availability of longer reads (250-500 bases) from the Roche-454 FLX Titanium platform allows relatively accurate assembly of data into contigs, permitting better overall annotation and data mining.

Here we have analyzed the transcriptome of crowns and rhizomes obtained from
field grown switchgrass cv Summer plants. This cultivar is an upland tetraploid with
good winter-hardiness [17], and has been used to create hybrids which show heterosis for
yields [18, 19].

2.3 Materials and Methods

2.3.1 Plant material

Stands of switchgrass cv Summer had been established in the field near Mead, NE
for several years [20]. Above ground portions of the plants were cut and below ground
portions of the plants were then harvested in late August 2009 at the post-anthesis, seed
maturation stage of development using a lever action hole cutter for golf greens. Four
soil plugs containing crown, roots and rhizomes were placed in plastic bags and kept on
ice until cleaned. Soil plugs were cleaned by hand within 1 hr of harvest. Adherent soil
was removed using tooth brushes. Crowns and rhizomes were trimmed to remove much
of the roots and tiller buds and immediately flash frozen in liquid nitrogen. Flash-frozen
tissues were placed on dry ice for transport to the laboratory and stored at -80 C until
used. Crowns and rhizomes were fine milled either by hand or using a cryogenic grinder
(6870 Freezer Mill (Spex Sample Prep, Metuchen, NJ). Pulverized plant material was
used to extract RNA.

2.3.2 RNA extraction and cDNA library generation

Total RNA was extracted from all switchgrass tissues using the modified Trizol
(Invitrogen, Carlsbad, CA) protocol of Tobias, Twigg, et al. (2005). In short, total RNA
was extracted from 16-20 100 mg aliquots of switchgrass tissue. During extraction, the
RNA from two 100 mg aliquots was combined for resuspension in 50 µl of RNase-free
water with RNaseOUT ribonuclease inhibitor added. The pellets from both aliquots were
resuspended sequentially in the same 50 µl of water with heating at 60°C for 5 minutes each. Any undissolved pellet material was discarded. From these samples, mRNA was isolated using the FastTrack MAG Maxi isolation kit and 100 µl of magnetic beads as directed (Invitrogen, Carlsbad, CA). The mRNA was quantitated using a Nanodrop spectrophotometer (Thermo Fisher, Waltham, MA). Synthesis of cDNA was performed using the high yield protocol of the QuantiTect Whole Transcriptome kit (Qiagen, Valencia, CA) with 100 ng of mRNA as the starting material. The cDNA was purified from the reaction mixture using the QIAamp DNA Blood Mini kit (Qiagen, Valencia, CA). Clean-up was achieved using the supplementary protocol for the purification of REPLI-g amplified DNA. The cDNA was again quantitated using the Nanodrop and adjusted to a concentration of 400-500 ng/µl (total cDNA of 40-50 µg provided for 454 sequencing).

2.3.4 454 Pyrosequencing

Switchgrass cv Summer crown and rhizome cDNA was fractionated and sequenced using a 454 GS-FLX sequencer with Titanium chemistry according to the manufacturer’s instructions (Roche, IN, USA) at the Core for Applied Genomics and Ecology (CAGE), The University of Nebraska-Lincoln, Lincoln, NE.

Briefly, 10ng of cDNA was nebulized. Fragment end polishing, adaptor ligation, and library immobilization reactions were subsequently carried out using GS FLX Titanium General Library Prep Kits (Roche, IN, USA). The single stranded (sst) template DNA was eluted with 25 µl of the EB buffer (QIAGEN, Valencia CA) and DNA profile and quantification were measured by running 1 µl of the samples on the Agilent Bioanalyzer 2100 (Santa Clara, CA) using a RNA Pico 6000 chip. The final sst DNA
library was quantified using Qubit (Invitrogen, Carlsbad, California) and was diluted to a normalized concentration of $1 \times 10^8$ molecules/µl for the emulsion PCR reactions. Emulsion PCR and sequencing were performed according to the FLX titanium protocols. The read number, average read length, and average quality of the reads for each run is shown in Table 1. Sequence files are available in the NCBI Sequence Read Archives under study: SRP009076; and Runs: SRR358964; SRR358965; and SRR358966.

2.3.5 Bioinformatics and de novo transcriptome assembly

454 GS FLX Titanium sequence data was assembled using Roche’s GS De Novo Assembler (gsAssembler) software, version 2.3. The –cDNA option was used since the sequence data source was mRNA. The default assembly parameters were used to assemble all three half-plates in a single assembly, and the software automatically excluded reads <50 bp. The assembly output consisted of a series of 27,687 contigs, all of which were greater than 100 bp in length. These contigs were used for downstream analysis. In addition to the contig sequences, individual sequencing reads that had no significant overlap with any other read were classified as “singletons” by Roche’s software and not included in the assembly output. These singleton reads were separated from the initial data set and all of these reads greater than 250 bp in length were also used in downstream analysis. Although we have used contig coverage as an approximation of transcript abundance, the actual relationship between these two parameters has not been quantified.

2.4 Results

Switchgrass crowns and rhizomes obtained from field grown cv Summer plants were used to generate cDNA libraries for pyrosequencing on a Roche Inc 454 GS-FLX
instrument. Three aliquots from two different library preparations were sequenced. The pooled raw read exhibited a bimodal distribution with a broad peak centered around 125 bp and a sharper peak centered around 515 bp (Figure 1A). Quality trimming of these reads prior to assembly by Newbler version 2.3 (Roche Inc) resulted in a bulk of reads under 250 bp, although the biomodal distribution was still evident (Figure 1B).

The trimmed component yielded a total of 929,820 reads containing over 216 million bases. These reads were assembled into 27,687 contigs of 100 base pairs or larger with a total assembly length of 12.9 million bases. 641,443 (69%) reads of the original 929,982 were included in this assembly with an inferred read error of 2.23%. This error term was generated by the Newbler assembler, and was defined as: Number of read alignment differences/Number of mapped bases. About 18% (170,312) of the reads failed to assemble into contigs using the Newbler 2.3 program and were categorized as singletons. The other aligned reads could be placed into 12,548 isogroups (gene models), and 27,687 contigs (Table 1). The average contig length was 722 bp and the median was 568 bp.

To assess potential contamination in this assembly, switchgrass contig and singleton sequences were first compared to proteins present in diverse taxonomic groups of organisms contained in the Refseq databases (NCBI) using the blastX algorithm at a e-value threshold of 1 x 10^{-7}. These analyses showed that the contig and singleton sequences displayed a match of 0.02% and 0.05% respectively to microbial sequences; 0.08% and 0.70% to fungal proteins; and 0.01% and 2.91% to invertebrate sequences within the Refseq collections. These data indicated that most of the assembled sequences were from switchgrass tissues.
Figure 1. Size and distribution of 454 reads for dataset (A) raw reads and (B) trimmed reads after removal of adapter sequences and sequences of poor quality. Reads under 50bp were excluded from further analyses.
Assembled contigs 100 bp and longer and singletons longer than 250 bp were annotated with Blast2GO (Conesa and Gotz 2008; Gotz, Garcia-Gomez et al. 2008) (www.Blast2GO.org/) in a two step process. First, the blastx algorithm [21] was used searching against the NCBI non-redundant protein database using an e-value of $1 \times 10^{-3}$ cutoff and saving up to 20 blast-hits for each sequence. Second, every significant blast hit for each sequence was searched against a Gene Ontology database to collect all of the GO terms associated with related proteins. Out of the 27,687 contigs, approximately 70 % (19,505 sequences displayed at least one blast hit at the e-value of $1 \times 10^{-3}$), and the remaining 30 % (8,174 sequences) did not have a blastx hit (Figure 2A). The top 50 most abundant contigs are shown in Supplementary Table 1. This list contained ESTs coding for metabolic enzymes, transcription factors and proteins involved in signaling.

The contig sequences that did not display a blastx similarity, were next analyzed by the blastn algorithm [21] against the NCBI ALL_EST database (Figure 2B). Analysis of these 8,174 contigs indicated that approximately 87 % of the remaining contig sequences matched other sequences with an e-value of $1 \times 10^{-3}$, and almost 73 % of the remaining contig sequences had an EST match of less than $1 \times 10^{-30}$.

Similar analyses for the 43,094 singleton sequences were performed (Figure 3A and 3C). Over 54 % of the sequences did not have a match against the NCBI protein databases of less than a value of $1 \times 10^{-3}$. Of the sequences showing protein matches, 87 % had at least one GO assigned term and 13 % (2,505 sequences) had a blastx hit, but no GO terms assigned (Figure 3A). The singleton sequences (23,344) without a blastx hit were compared against the NCBI EST database using the blastn algorithm (Figure 3B).
Figure 2. Pie chart showing results of Blast2GO alignments for contig sequences. (A) Distribution of sequences with GO terms assigned (green), Blast hit but no GO terms assigned at an e-value of $1 \times 10^{-3}$ (blue), and no Blast hits at an e-value of $1 \times 10^{-3}$ (red). (B) Reanalysis of contig sequences with no Blast2GO assignments by Blastn, high confidence blast hits with an e-value of $1 \times 10^{-30}$ or lower (green), this with an e-value between $1 \times 10^{-3}$ and $1 \times 10^{-30}$ (blue), and no blast hits (red). Numbers in each section are the total number of contigs assigned to each category.
A majority (65%) of the queried sequences displayed a match to existing EST sequences at an e-value of < 1 x 10^{-3} with 36% of the sequences having an EST match at the e-value of 1 x 10^{-30} cutoff. 8,165 sequences did not have a significant match. Of the reanalyzed singleton sequences with a match to an existing EST in the NCBI database, 56% displayed a match to EST sequences at an e-value threshold of 1 x 10^{-3}. These initial analyses suggested that even a small DNA sequence error rate (1 bp) in base assignment in the non-overlapping regions of a contig or within a singleton could lead to the virtual translation of short or incorrect protein sequences resulting in no matches in a Blast2GO search. However, reanalysis of these Blast2GO unmatched sequences using the blastn algorithm indicated that many of these 454-derived sequences indeed matched existing ESTs.

We next performed a database search using blastn with an e-value threshold of < 1 x 10^{-25} with the contig (27,687) and singleton (43,094) sequences to the available switchgrass UniGenes from NCBI (Build #2 from August 25, 2010), sorghum [Phytozome.org version 7.0], and the Brachypodium [Phytozome.org version 7.0], transcriptomes [22, 23] (Table 2). As expected greater than 81% of the contig sequences had a match in the available switchgrass ESTs, the remaining 19% (5,199) sequences appear to be new to this dataset. Matches to the sorghum and Brachypodium transcriptomes were considerably less, approximately ~ 63% and ~53% respectively (Table 2). For the singleton sequences, approximately 39% shared a significant identity to the available switchgrass ESTs, and the matches to the sorghum transcriptome were somewhat lower (~27%). The least identity was observed with the Brachypodium transcriptome. A bulk of the predicted singleton sequences did not have a match to the
Figure 3. Pie chart showing results of Blast2GO alignments for singleton sequences. (A) Distribution of sequences with GO terms assigned (green), Blast hit but no GO terms assigned at an e-value of $1 \times 10^{-3}$ (blue), and no Blast hits at an e-value of $1 \times 10^{-3}$ (red). (B) Reanalysis of contig sequences with no Blast2GO assignments by Blastn, high confidence blast hits with an e-value of $1 \times 10^{-30}$ or lower (green), this with an e-value between $1 \times 10^{-3}$ and $1 \times 10^{-30}$ (blue), and no blast hits (red). Numbers in each section are the total number of contigs assigned to each category.
three plant transcript databases that were queried (Table 2).

To better understand the robustness of the contig assembly, we took all the ~18,000 contig sequences with an assigned GO term (See Figure 2), and performed an annotation using increasingly stringent cutoff values. We were also interested in finding a stringency parameter (blastx e-value) that afforded good annotation, and that could be routinely used to analyze the total dataset. Of the total contig sequences, approximately 92 % (18,135 sequences) showed a match to existing annotated sequences at e-value of $1 \times 10^{-3}$. As might be expected, increasing the stringency from e-value of $1 \times 10^{-3}$ to e-value of $1 \times 10^{-50}$ resulted in a loss of 39 % in the number of annotated sequences, with similar decreases in the total number of enzyme codes and annotated sequences with enzyme codes [Table 3]. We selected an e-value of $1 \times 10^{-15}$ for further annotation of sequences. This value appeared to be a reasonable compromise between discovering true protein/enzyme matches in our dataset, arising from imperfect assembly of sequences introducing frameshifts, short translated reads and potential lack of orthologs in the existing databases [also see Table 2].

A blastx comparison of the proteins coded by the switchgrass contig and singleton sequences with other plant and non-plant species also provided further insights to this dataset. The best matches of the translated switchgrass contig sequences were to sorghum (*Sorghum bicolor* L. Moench.) proteins, followed by maize (*Zea mays* L.) and rice (*Oryza sativa* L.) protein sequences [Figure 4A]. There were fewer hits to other plant species, including a range of dicots [Figure 4A]. Very few switchgrass contigs displayed a significant match to non-plant sequences. In contrast, for the switchgrass singleton sequences, although the best matches were to sorghum, the next highest scores were to
### Figure 5. Top ten matches of switchgrass crown and rhizome sequences to other species.

(A) Contigs. (B) Singletons. An e-value cutoff of $1 \times 10^{-3}$ or lower was used to designate matches to the species identified on the left axis.

#### A. Contigs
- **Saccharum hybrid** 33
- **Arabidopsis lyrata** 38
- **Populus trichocarpa** 40
- **Triticum aestivum** 40
- **Ricinus communis** 51
- **Hordeum vulgare** 63
- **Vitis vinifera** 67
- **Oryza sativa** 2,990
- **Zea mays** 5,133
- **Sorghum bicolor** 10,468

#### B. Singletons
- **Acrithosiphon pisum** 48
- **Ricinus communis** 52
- **Bombyx mori** 76
- **Populus trichocarpa** 81
- **Acinetobacter junii** 106
- **Vitis vinifera** 140
- **Tribolium castaneum** 1,474
- **Zea mays** 2,895
- **Oryza sativa** 4,157
- **Sorghum bicolor** 8,872
rice followed by maize. Significant numbers of singleton sequences matched to proteins present in insects *Tribolium castaneum* (Red flour beetle), *Bombyx mori* (silkworm), and the pea aphid (*Acyrthosiphon pisum*). Matches to a gram negative bacterium (*Acinetobacter junii*) were also present in the singleton sequences [Figure 4B].

A comparison of the assembled crown and rhizome contigs and singletons against available switchgrass ESTs derived from different tissues [16] was performed to obtain new sequences not yet present in these databases, and an approximation of the distribution of ESTs (expression snap-shot) in the different tissues analyzed [Figure 5]. Of the ~70,000 contigs and singletons, 30,177 sequences did not have a match to the available EST sequences in cDNA libraries generated from seedlings, callus, young crowns and roots, vegetative and floral apices and developing seeds. A total of 11,043 sequences were common to all of the tissue/stage specific ESTs queried [Figure 5], ESTs common between all of the other library comparisons were below ~3,500. The relatively small number of matches (1,715) to the existing crown and root ESTs derived from Sanger sequencing of young tissues suggested that crowns and rhizomes in field grown plants had a considerably more complex transcriptome. Overall these comparisons indicated that the 454 sequencing had yielded a significant coverage of the crown and rhizome transcriptome. The 30,177 sequences (“new”) to the current 454 assembly yielded ~2,000 gene models that were analyzed by Blast2GO. We compared the resultant output to the whole crown and rhizome transcriptome assembly to determine overall distribution of GO terms in these ~2,000 gene models, and to detect any over/under representation within broad and narrow GO terms.

At the “Biological Processes, Level 2 Terms” there were some variations between
Figure 5. Comparison of 454 contig and singleton sequences to switchgrass ESTs. ESTs for seedlings, callus, young crown and root, stem, stem apices, and floral organs were obtained from publically available databases. The assembled 454 sequences were compared against each library by Blastn with an e-value of $1 \times 10^{-25}$ or lower to identify transcripts found in common between the compared datasets. Crown and rhizome sequences without a significant match (30.117) were considered to be “new” sequences and are shown within the largest ellipse. The numbers of common transcripts among and between the compared databases are shown in the appropriate areas.
the new sequences and the whole assembly in classifications into terms. The ‘new’ contained slightly greater proportion of sequences that were assigned to the cellular process, cellular component organization and death categories, and lower representation in the biological regulation, response to stimulus, and signaling categories as compared to the whole assembly [Figure 6A]. Representation (percentage) in other categories was essentially similar. A comparison between these sequences at the “Molecular Function, Level 3 Terms” showed a greater abundance of new sequences that matched to nucleotide/nucleoside/nucleic acid and chromatin binding as compared to the whole assembly [Figure 6B]. These new sequences also appeared to be enriched in transcripts coding for hydrolases and transferases as compared to the whole assembly. There was a slight decrease in transcripts coding for proteins assigned into the cofactor binding, structural constituent of ribosomes, and transcription factor activity categories [Figure 6B]. We did not statistically evaluate these differences since the descriptive value of this analysis would not have been affected.

Two different subsets of proteins, containing relatively few sequences (peroxidases) and a much larger number of sequences (transcription factors) within the “new” were selected for more detailed analysis. There were a total of 56 sequences coding for 21 different proteins putatively identified as peroxidases by Blast2GO. Of these 56 DNA sequences, a large number (26) were classified as retrotransposons of an unknown category or as putative copia-like retrotransposons and were not analyzed further. The remaining 30 sequences coded for 16 proteins [Table 4] that contained a “peroxidase descriptor”. Two sequences coding for a putative acid phosphatase were included in this annotation due to the association of the “haloacid peroxidase like” term
Figure 6. Comparison of the percent of sequences associated with a specific Blast2GO term for the whole assembly (blue) and the “new” sequences (red) (see Figure 5.). (A) Biological processes terms and (B) molecular processes terms.
within GO. The contigs coding for catalase displayed a moderate match to the two catalases encoded within the sorghum genome (e-value of $1 \times 10^{-20}$ to SORBIDRAFT_04g001130; and e-value of $1 \times 10^{-19}$ to SORBIDRAFT_10g030840), in contrast to strong matches ($< e$-value of $1 \times 10^{-90}$) to these catalases within the available switchgrass ESTs in the public databases. It is unclear if this rhizome sequence codes for a catalase or a catalase-like protein. Analysis of the 5 class three peroxidases present in the “new” sequences showed they belonged to 4 different clades described by Passardi et al. [24] for the class III peroxidases encoded by the rice genome [Table 4]. Two other proteins appeared to be transcripts coding for cytosolic ascorbate peroxidase and a 2cys-peroxiredoxin.

New sequences coding for transcription factors are shown in [Table 5]. There were a total of 175 DNA sequences that were identified by Blast2GO as transcription factors. Filtering out multiple sequences that coded for the same protein and annotation by hand of the remaining sequences resulted in the identification of over 30 sequences that coded for transcription factors with orthologs with a known function in other plants. Many of these switchgrass genes were present as singletons in the assembled transcriptome, and coded for proteins controlling a number of important cellular process, including control of cell cycle [E2F and its repressor E2L], organ development [Rolled leaf; bZIP42, BHL4, and HUA2], interactions with the environment [RAV1 and 2, CBF-7, LEC1, SPL7, NLP7 and NFXL1], hormone signaling [ARF-7, ABI-5, MYB101], DNA repair and remodeling [SWI3D, DUO1 and NAC8], histone H3 demethylation, specifically at H3K4 [ELF6] and at H3K27 [REF6] [Table 5]. A number of proteins belonging to the WRKY family were annotated, but it was more difficult to accurately
predict their exact orthologs due to incomplete sequence coverage. Although not fully explored in this study, many of these factors will show variable levels of interaction, and their identification will be the prelude to dissecting their role(s) in switchgrass crowns and rhizomes. Two KEGG (http://www.genome.jp/kegg/) pathways were populated with DNA sequences identified by Blast2GO as coding for metabolic enzymes involved in glycolysis/gluconeogenesis (MAP00010), and starch and sucrose metabolism (MAP00500). These pathways were chosen to ascertain that sequences expected to be abundant in metabolically active tissues were present. The entire pathway relevant to plants for glycolysis/gluconeogenesis was populated with varying levels of transcript abundances [colored shading on appropriate boxes, Figure 7]. Transcripts for glyceraldehyde-3-phosphate dehydrogenase, pyruvate kinase, phosphoglycerate kinase, and aldolase were most abundant and transcripts coding for enzymes such as phosphoenolpyruvate carboxykinase and aldose-1-epimerase were least abundant.

A similar protocol was used to study the presence of enzymes in the starch and sucrose metabolism pathways. Sugar metabolism can be expected to play a key role in the growth and adaptation of the below-ground tissues in switchgrass to changes in photosynthate supply over the course of the growing season. Transcripts for all of the expected enzymes were found with the notable exception of a 1,4-β-D-xylan synthase (E.C. 2.4.2.24) [Figure 8]. However, transcripts for UDP-D-xylose synthetase and xylan 1-4-β-xylosidase were not very abundant. In contrast, transcripts for enzymes involved in pectin synthesis were quite abundant. Similarly, these tissues appeared to have high levels of transcripts for sucrose, starch and cellulose synthesis. The presence of enzymes involved in the biosynthesis of polysaccharides, suggested that active growth was
Figure 7. KEGG map for glycolysis/gluconeogenesis populated with transcripts coding for specific enzymes in the pathway. The abundance of sequences identified by Blast2GO for a given enzyme is shown with different colored boxes. The abundance range for the assignment of transcripts is shown on the left margin. Blue (0-99), light green (100-249), yellow (250-499), orange (500-999), and red (>1,000).
**Figure 8.** KEGG map for starch and sugar metabolism populated with transcripts coding for specific enzymes in the pathway. Other details are as described for Figure 7.
probably still occurring in the crowns and rhizomes of these plants.

2.5 Discussion

Perenniality in switchgrass is likely to be controlled by several mechanisms that ultimately impact the physiological status of the below-ground components of the plant [25]. These below ground structures include the roots, rhizomes and crowns. It can be anticipated that over a growing cycle there will be significant, but cyclical, changes in the physiology of these organs underpinned by significant changes in the transcriptome. These changes will include nutrient remobilization and regeneration of new shoots (tillers) at the onset of green-up in early spring, increased tissue accretion over the growing season, transition to the slowing of developmental processes in the fall, and followed by potentially a quiescent stage in the winter. At present, we lack meaningful molecular insights into these processes.

As a first step in understanding how gene expression patterns change in the below ground tissues of switchgrass plants over a growing season, we have assembled a preliminary transcriptome using over 900,000 sequences obtained by next generation sequencing of tissues obtained from cv Summer. Tissues were obtained from plants at the S4 stage; seeds at physiological maturity; [26]. Contamination within the assembled sequences from other organisms was quite low, indicating that a majority of these sequences were derived from switchgrass tissues. However, some level of contamination from insect, fungal and bacterial sources can be expected in field-harvested tissues. As might be expected, the level of non-switchgrass transcripts was greater in the singleton pool as compared to the assembled contigs. A majority of the contigs (~67%) had a GO term assigned, consistent with studies in several other non-model species lacking
annotated genomes [12, 14, 27, 28] [Supplementary Table 2]. The numbers of singletons reported in other studies have been variable, and success in finding GO terms have been dependent on whether these sequences were analyzed separately or combined with contigs. The switchgrass singletons reported here were analyzed as a separate pool, to maximize the discovery of new (and potentially rare) transcripts and to understand relative distribution of similarities to other organisms. A combination of Blast2GO and Blastx searches resulted in effectively identifying ~87% of the singleton sequences as sharing significant similarity to other plants. Nonetheless, actual GO annotation in this singleton pool (~40%) was lower as compared to the contig pool indicating that deeper coverage might have improved discovery.

A sizeable fraction (~>40%) of the 454-derived sequences were not present in earlier EST collections [16], suggesting that these “new” transcripts not currently in the databases could contain some proportion of rhizome and crown specific sequences. However, the unequal distribution of sequences across all libraries (ESTs and 454; see Figure 5) might have skewed these comparative analyses, and “new” sequences could contain low abundance and rare transcripts that might have escaped detection during Sanger-sequencing of the different switchgrass tissues [16]. The relatively large number of overlapping sequences present in all the libraries can be expected to contain transcripts coding for many metabolic processes common to all switchgrass tissues.

New sequences generally occupied a similar distribution within the biological processes and molecular function GO terms as compared the overall 454 assembly, although certain categories broadly classified as “binding” and “hydrolytic/transferase activities” within GO were somewhat overrepresented. These data indicated that
singleton sequences could code for specific, but rare proteins such as transcription factors that are needed for normal functioning of crowns and rhizomes. Indeed, a manual annotation of two sets of sequences classified as coding for peroxidases and transcription factors showed the utility of such a detailed analysis. Five new peroxidases, a cytosolic ascorbate peroxidase, and thioredoxin peroxidase were identified. A bacterial-induced peroxidase homolog [29] present in the tissues analyzed indicated that rhizome and crowns could have been under biotic stress. Although care was exercised to process field-harvested tissues as quickly as possible (within 1 hour of harvest), some elevation of transcripts associated with stress and or wounding might be expected.

Singleton sequences also coded for a number of transcription factors, which could be expected to be of lower abundance in the transcriptome. Although they belonged to diverse families, several factors that control chromatin remodeling were identified. Among these, the switchgrass orthologs of the Jumonji-type ELF6 [30], REF6 [31] and SWI3D [32], could provide a future means to understand chromatin remodeling that might occur in response to the seasonal growth habits of the switchgrass plants. Additionally, many of these transcription factors control the expression of genes involved in basal cell metabolism and responses to biotic and abiotic stress. Transcript abundances for genes coding for metabolic pathways were variable, although abundances for those involved in primary metabolism were quite high. The assembly provided significant insights into the status of these tissues, and broadly indicated that there was active metabolism taking place in the crown and rhizomes at this stage of plant development. Future next-generation analyses of crown, rhizome and root transcriptomes across the growing season, among switchgrass populations with divergent winter-hardiness
responses should yield even greater insights into tissues that impact perenniality and are essential to the sustainable production of this important bioenergy feedstock.

2.6 Acknowledgements

We thank Steve Masterson and Patrick Callahan for excellent technical support. This work was supported by the Office of Science (BER), U. S. Department of Energy Grant Number DE-AI02-09ER64829, and by the USDA-ARS CRIS project 5440-21000-028-00D. The U.S. Department of Agriculture, Agricultural Research Service, is an equal opportunity/affirmative action employer and all agency services are available without discrimination. Mention of commercial products and organizations in this manuscript is solely to provide specific information. It does not constitute endorsement by USDA-ARS over other products and organizations not mentioned.
### Table 1  Assembly of 454 data

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total reads</td>
<td>928,820</td>
</tr>
<tr>
<td>Total bases</td>
<td>216,450,730</td>
</tr>
<tr>
<td>Aligned reads</td>
<td>641,443</td>
</tr>
<tr>
<td>Aligned bases</td>
<td>146,297,845</td>
</tr>
<tr>
<td>Inferred read error</td>
<td>2.23%</td>
</tr>
<tr>
<td>Singletons</td>
<td>170,312</td>
</tr>
<tr>
<td>Isogroups</td>
<td>12,548</td>
</tr>
<tr>
<td>Isogroups with one contig</td>
<td>7,288</td>
</tr>
<tr>
<td>Contigs</td>
<td>27,687</td>
</tr>
</tbody>
</table>

### Table 2  Blastn of contigs and singletons with selected plant transcriptomes

<table>
<thead>
<tr>
<th>Species</th>
<th>Blast hits(^a)</th>
<th>No blast hits</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Contigs</td>
<td>Singletons</td>
<td>Contigs</td>
</tr>
<tr>
<td>Sorghum bicolor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brachypodium distachyon</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panicum virgatum (EST)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) An e-value of 1 x 10\(^{-15}\) was used for these analyses

### Table 3  GO annotation of contig sequences at different stringencies

<table>
<thead>
<tr>
<th>e-value</th>
<th>Sequences(^a)</th>
<th>Annotations(^b)</th>
<th>Number of sequences with EC(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number (%)</td>
<td>EC</td>
<td></td>
</tr>
<tr>
<td>1 x 10-3</td>
<td>16,708 (92.1)</td>
<td>72,772</td>
<td>6,291</td>
</tr>
<tr>
<td>1 x 10-10</td>
<td>15,979 (88.1)</td>
<td>69,473</td>
<td>6,079</td>
</tr>
<tr>
<td>1 x 10-15</td>
<td>15,102 (83.3)</td>
<td>65,704</td>
<td>5,781</td>
</tr>
<tr>
<td>1 x 10-25</td>
<td>13,484 (74.4)</td>
<td>58,683</td>
<td>5,273</td>
</tr>
<tr>
<td>1 x 10-50</td>
<td>10,223 (56.4)</td>
<td>44,388</td>
<td>4,167</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EC: number of enzyme codes recovered at each e-value threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>(^a) Total number of annotated sequences and percentage of total input sequences with a match</td>
</tr>
<tr>
<td>(^b) Total number of annotations across all GO terms</td>
</tr>
<tr>
<td>(^c) Total number of sequences with an enzyme code as identified by GO analysis</td>
</tr>
</tbody>
</table>
Table 4  Peroxidases and related proteins identified in cv Summer crowns and rhizome transcripts (identified within the “new” group with significant matches to heme-containing oxidoreductases)

<table>
<thead>
<tr>
<th>C or S</th>
<th>Protein Family</th>
<th>Predicted ortholog and group⁹</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>Ascorbate peroxidase</td>
<td>Cytosolic</td>
</tr>
<tr>
<td>C</td>
<td>Bacterial-induced peroxidase</td>
<td>Os02g0237000, group IV.4</td>
</tr>
<tr>
<td>S</td>
<td>Catalase</td>
<td>Catalase</td>
</tr>
<tr>
<td>C</td>
<td>Class III peroxidase</td>
<td>PviPrx-19, monocot-specific group V</td>
</tr>
<tr>
<td>C</td>
<td>Class III peroxidase</td>
<td>Os09g0471100, group IV</td>
</tr>
<tr>
<td>C</td>
<td>Class III peroxidase</td>
<td>Rice peroxidase 124, group III</td>
</tr>
<tr>
<td>S</td>
<td>Class III peroxidase</td>
<td>OS03g0762400, group VI</td>
</tr>
<tr>
<td>S</td>
<td>Peroxidasin homolog</td>
<td>Fatty acid dioxygenase</td>
</tr>
<tr>
<td>S</td>
<td>Thioredoxin peroxidase</td>
<td>2-cys peroxiredoxin</td>
</tr>
</tbody>
</table>

C contig, S singleton, PviPrx-19 *Panicum virgatum* peroxidase 19 (identified in earlier work by Tobias et al. [16])

⁹ Predicted rice ortholog based on Passardi et al. [24]
Table 5  Switchgrass transcription factor orthologs identified within cv Summer crowns and rhizome transcripts (only “new” switchgrass transcripts with orthologs of known function are shown; orthologs defined as best-hit by BLASTP to translated switchgrass sequences)

<table>
<thead>
<tr>
<th>C or S</th>
<th>TF family</th>
<th>Predicted ortholog</th>
<th>Function of orthologs</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>AP-2 domain</td>
<td>RAV 1 and 2 (Atha)</td>
<td>Interactions with CONSTANS and FT Represses E2F activation of genes</td>
<td>[33]</td>
</tr>
<tr>
<td>S</td>
<td>Antagonist of E2F</td>
<td>E2L (Atha)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Auxin response</td>
<td>ARF-7 (Slyc)</td>
<td>Auxin and GA signaling</td>
<td>[35]</td>
</tr>
<tr>
<td>S</td>
<td>bHLH family</td>
<td>HEC-1 (Atha)</td>
<td>Rolled leaf Involved in rooting/root hairs</td>
<td>[36]</td>
</tr>
<tr>
<td>C</td>
<td>bHLH family</td>
<td>HEC-1 (Atha)</td>
<td>Female flower development</td>
<td>[37]</td>
</tr>
<tr>
<td>S</td>
<td>Basic leucine zipper</td>
<td>LIGULELESS-2 (Zmay)</td>
<td>Organ development Demarcates boundary of leaf and sheath</td>
<td>[39]</td>
</tr>
<tr>
<td>S</td>
<td>Basic leucine zipper</td>
<td>ABI-5 (Atha)</td>
<td>Abscisic acid signaling</td>
<td>[40]</td>
</tr>
<tr>
<td>S</td>
<td>bZIP</td>
<td>bZIP 42 (Atha)</td>
<td>Organ formation</td>
<td>[41]</td>
</tr>
<tr>
<td>C</td>
<td>CBF/DREB-like CCAT-binding factor</td>
<td>CBF-7 (Atha)</td>
<td>Cold hardiness Embryo/fatty acid biosynthesis</td>
<td>[42]</td>
</tr>
<tr>
<td>S</td>
<td>Chromatin remodeling</td>
<td>LEC1 (Zmay)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>E2F</td>
<td>E2F (Atha)</td>
<td>Control of cell cycle</td>
<td>[34]</td>
</tr>
<tr>
<td>C</td>
<td>Heat shock</td>
<td>SPL7 (Osat)</td>
<td>Control of leaf spots</td>
<td>[45]</td>
</tr>
<tr>
<td>S</td>
<td>Hox-family</td>
<td>BHL4 (Atha)</td>
<td>Leaf margin organization</td>
<td>[46]</td>
</tr>
<tr>
<td>S</td>
<td>Jumonji-domain</td>
<td>ELF6 (Atha)</td>
<td>Histone H3K4 demethylation Histone H3K27</td>
<td>[47]</td>
</tr>
<tr>
<td>S</td>
<td>Jumonji-domain</td>
<td>REF6 (Atha)</td>
<td>Demethylation</td>
<td>[48]</td>
</tr>
<tr>
<td>S</td>
<td>Myb family</td>
<td>DUO-1 (Atha)</td>
<td>Male germline formation</td>
<td>[49]</td>
</tr>
<tr>
<td>S</td>
<td>Myb r2r3 family</td>
<td>MYB101 (Atha)</td>
<td>Hormone signaling Suppressor of gamma response 1</td>
<td>[50]</td>
</tr>
<tr>
<td>S</td>
<td>NAM-superfamily</td>
<td>NAC8 (Atha)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>NIN-Like</td>
<td>NPL7 (Atha)</td>
<td>Nitrate sensing</td>
<td>[52]</td>
</tr>
<tr>
<td>S</td>
<td>NF-X1 type Tudor/PWWP/MBT domain-containing</td>
<td>NFXL1 (Atha)</td>
<td>Protection under stress</td>
<td>[53]</td>
</tr>
<tr>
<td>C</td>
<td>WRKYs</td>
<td>Several members (Atha)</td>
<td>Flowering/shoot morphology ABA, UV responses; biotic and abiotic stresses</td>
<td>[54]</td>
</tr>
</tbody>
</table>

C contig, S singleton, TF transcription factor, Atha Arabidopsis thaliana, Osat Oryza sativa, Slyc Solanum lycopersicum, Zmay Zea mayes
2.7 Works Cited


50. Reyes JL, Chua NH: **ABA induction of miR159 controls transcript levels of two MYB factors during Arabidopsis seed germination.** *Plant J* 2007, **49**(4):592-606.

51. Yoshiyama K, Conklin PA, Huefner ND, Britt AB: **Suppressor of gamma response 1 (SOG1) encodes a putative transcription factor governing multiple responses to DNA damage.** *PNAS USA* 2009, **106**(31):12843-12848.


Chapter 3

Global changes in mineral transporters in tetraploid switchgrasses (*Panicum virgatum* L.)

Note: The results described in this chapter have been published and all text is modified from the original version. Supplemental data can be found online at the journal’s website.

3.1 Abstract

Switchgrass (*Panicum virgatum* L) is perennial, C₄ grass with great potential as a biofuel crop. An in-depth understanding of the mechanisms that control mineral uptake, distribution and remobilization will benefit sustainable production. Nutrients are mobilized from aerial portions to below-ground crowns and rhizomes as a natural accompaniment to above-ground senescence post seed-set. Mineral uptake and remobilization is dependent on transporters, however, little if any information is available about the specific transporters that are needed and how their relative expression changes over a growing season. Using well-defined classes of mineral transporters, we identified 520 genes belonging to 40 different transporter classes in the tetraploid switchgrass genome. Expression patterns were determined for many of these genes using publically available transcriptomic datasets obtained from both greenhouse and field grown plants. Certain transporters showed strong temporal patterns of expression in distinct developmental stages of the plant. Gene-expression was verified for selected transporters using qRT-PCR. By and large these analyses confirmed the developmental stage-specific expression of these genes. Mineral analyses indicated that K, Fe, Mg, Co and As had a similar pattern of accumulation with apparent limited remobilization at the end of the growing season. These initial analyses will serve as a foundation for more detailed examination of the nutrient biology of switchgrass.
3.2 Introduction

Plant mineral composition depends on uptake and translocation of minerals from the rhizosphere, through the root-shoot junction (crown), and into the aboveground tissues. These processes are influenced by both environmental and genotypic factors. In the perennial plant growth cycle, certain minerals can be recycled or remobilized from senescing tissues in the autumn, stored in perennial tissues during winter dormancy, and then remobilized and translocated to growing tissues in the spring. In switchgrass (*Panicum virgatum* L.), the stems and leaves survive for one year, while perennial tissues including the crowns, rhizomes, and older roots survive for much longer periods of time. The crown and rhizome tissues connect the root system to the shoot system, thus, minerals that are remobilized from annual or perennial tissues must pass through the crown tissue. These tissues can also serve as repository for remobilized nutrients at the end of the growing season. Thus, mineral uptake and recycling are cornerstones for the sustainable production of biomass from switchgrass and other perennial herbaceous bioenergy crops.

In *Arabidopsis thaliana*, concentrations of several minerals were highest in young tissues [1], suggesting that stage of harvest may be important for mineral concentration in plant tissues. In switchgrass, harvests during late vegetative stage or during summer had substantially higher P, Cl, K, and S than at post-senesce stage [2-6], demonstrating that these compounds are remobilized from leaf and stem tissue. Genotypic differences in remobilization were shown to be present [4, 6, 7].

The most abundant minerals in above ground switchgrass tissues are Si, K, Cl, Ca, and P [7, 8]. While several minerals are remobilized from aboveground biomass
during senescence in switchgrass, some abundant minerals are not, such as Si, Ca, and Mg [2, 3, 6]. Feedstock quality requirements depend on the conversion platform [9, 10] and pyrolysis and other thermochemical platforms will benefit from feedstocks that contain high lignin and low levels of N and alkali metals [11]. Reducing minerals such as Si will lower ash content as well. A number of genes that contribute to root uptake of Ca and Mg [12, 13], and transporters for Si uptake and distribution [14-16] have been identified in grasses. Currently, no whole genome-scale annotation and transcriptomic information for mineral and related transporter genes are available in switchgrass, however, this data would be useful to correlate specific genes with mineral accumulation.

Increased understanding of genes that impact mineral acquisition, transport and recycling in switchgrass can be used to improve both the genetics and management of switchgrass as a high-yielding biomass crop. Mineral transporters will be key players in these processes, as transport into and out of cells and organelles are the molecular events that underlie cellular storage and whole-plant translocation or recycling of minerals. Mineral transporter families have been studied extensively in species such as Arabidopsis and rice, providing gene sequence data to predict identity and function of unknown transcripts from other species. In switchgrass, no molecular studies of transporter genes have been conducted. Understanding of the interactions in uptake and remobilization between different minerals over the course of a growing season is limited for switchgrass. The release of the switchgrass genome (PviDraft0.0) by the Joint Genomes Institute (www.phytozome.org)[17] has greatly facilitated the discovery and annotation of genes and gene families in switchgrass [23].
The long-term goal of our research is to develop and utilize genotyping and phenotyping tools that can significantly enhance the breeding of switchgrass (*Panicum virgatum* L.) as a sustainable bioenergy crop for marginal crop lands [10]. Our objectives in this study were to use next-generation sequencing data to discover, annotate and quantify expression of switchgrass genes that are potentially involved with mineral transport in switchgrass. Here, we have combined bioinformatics and real-time qRT-PCR to classify transporter gene families in switchgrass and to identify specific genes that show altered expression over the growing season. We also used mineral analysis to quantify seasonal concentration changes in crown tissues.

3.3 Materials and Methods

3.3.1 Gene discovery

Known mineral and nutrient transporters in *Arabidopsis thaliana* were used to identify putative homologs in *Panicum virgatum*, *Sorghum bicolor*, and *Setaria italica* based on protein similarity using BLASTp [18, 19] and the respective reference genomes for each plant ([www.phytozome.org](http://www.phytozome.org)) [17, 20, 21]. A maximum e-value of 1 x 10\(^{-25}\) and minimum alignment of 50% were used as thresholds in filtering the BLASTp results for putative homolog identification.

3.3.2 Phylogenetic analyses

Cladograms were generated for genes of selected Arabidopsis and putative switchgrass transporter families. Sequences were analyzed for phylogenetic relationships using Phylogeny.fr [22].

Publically available 454 transcriptome sequencing datasets were used to generate expression profiles for identified putative mineral and nutrient transporters in switchgrass
as described previously [23]. 454 reads were aligned to the draft switchgrass transcriptome using Bowtie2 [24] and gene counts calculated using HTSeq-Count version 0.5.1p2 [https://pypi.python.org/pypi/HTSeq]. Expression counts were normalized through conversion from raw counts to RPKM (reads per kilo base exon per million mapped reads).

3.3.3 Heat maps and clustering

Heat maps were generated using estimated 454 expression data and two-way hierarchical clustering with JMP 9.0 (SAS Institute Inc., Cary, NC). RPKM expression values were converted to standardized values (z-scores) for each gene, and hierarchical clustering using Ward’s method was performed to yield heat maps and clusters of coexpressing transporters.

3.3.4 Plant material

Crowns and rhizomes were collected, cleaned and flash-frozen from field-established plants of cv Summer, as described earlier [25]. At each harvest date tissues were obtained from three individual plants. Flash-frozen tissues were stored at -80 °C until needed. Tissues were ground in a cryogenic grinder [25]. Aliquots (0.1 g) of ground materials were used for isolating RNA as previously described [26]. RNA samples were subsequently purified using RNeasy columns (Qiagen; Valencia, CA, USA) according to manufacturer’s instructions.

3.3.5 qRT-PCR and primers

DNase treated RNA samples were used to synthesize first strand cDNA by using SuperScript III reverse transcriptase (Invitrogen; Carlsbad, CA, USA) and random primers according to the manufacturer’s protocol. qRT-PCR reactions were set up in a
total volume of 15 uL using 7.5 uL master mix (Bio-Rad), 0.2 uL cDNA template, 0.75 uL primers, and 6.55 uL 18 MΩ water and conducted on a BioRAD CFX Connect, Real Time PCR instrument. Each reaction was performed in quadruplicate and the experimental design blocked plate with amplicon (a single amplicon per plate). Primers were designed using Primer3Plus [27]. Data was efficiency corrected using LinRegPCR [28, 29], and geNORM was used to screen for effective normalization genes and calculate relative quantities for each gene of interest [30]. Primers used are shown in Table A1.

3.3.6 Mineral analyses

Tissues were dried at 60 °C for at least 72 h and weighed. Samples (typically 25-50 mg) were digested as described previously [31]. In brief, samples were digested with 3 ml of concentrated HNO₃ (VWR, West Chester, PA, USA, Trace metal grade) at room temperature overnight, then at 100 °C for 1.5 h, followed by addition of 2 ml of 30% H₂O₂ (Fisher Scientific, Fair Lawn, NJ, USA) and digestion for 1 h each at steps of 125°C, 150°C, 165°C, and finally were heated to dryness at 180 °C. Dried samples were then resuspended in 5 ml of 1% HNO₃, and minerals were quantified by inductively coupled plasma mass spectrometry (ICP-MS) at the University of Nebraska Redox Biology Center Spectroscopy and Biophysics Core Facility.

3.3.7 Statistical analyses

Transcript levels were investigated by utilizing cDNA that originated from three individual genotypes (biological replicates) at each time point with four technical replicates per genotype. The cDNA was not pooled prior to qRT-PCR analysis. Thus, for any given harvest date, there were 12 total reactions that were conducted which
included both true biological as well as technical replicates. The GeNORM program (Hellemans et al. 2007) within the qbase+ software package was used to analyze reference genes in order to find suitably stable ones with a M-value below 1.5. In this way, Pavirv00026367m (a ubiquitin protein ligase) was selected as the stable reference gene for generation of relative quantities. The relative quantities were then statistically analyzed using PROC GLM in SAS (SAS Institute, Cary, NC) and Tukey's multiple comparison procedure was utilized to conduct pairwise comparisons of different harvest dates. Familywise error rate was controlled at $\alpha = 0.05$.

Data for the mineral analyses were subjected to single-factor ANOVA analysis of each mineral, error bars are standard deviations are from 3 biological replicates, with 2 technical replicates each. Of the sixteen minerals analyzed, the ten minerals showed statistically significant variation for at least two time points. P-values for the mineral analysis were calculated by Single Factor ANOVAs (in Excel).

3.4 Results and Discussion

Little is known about the identities of specific genes that contribute to remobilization of minerals from senescing tissues. Some genes are known to be important for remobilization and/or translocation of minerals from source to sink tissues in Arabidopsis, for example $YSL1$, $YSL3$ and $OPT3$ for iron, zinc, and copper [1, 32, 33], $NRT1.7$ for nitrate [34], $Sultr1;3$ for sulfate [35], and $Phl1;5$ for phosphate [36]. These genes have usually been discovered by analysis of mutants. A transcriptomic approach can reveal new insights to help understand nutrient deficiency signaling pathways [31, 37, 38]. Likewise, transcriptomic studies in Arabidopsis during senescence have identified many transporters, transcription factors, and other senescence associated genes.
that are up or down regulated [39-42]. However, their correlation with specific changes in minerals or N remobilization is still incomplete. Our overall goal in this study was to identify and classify switchgrass transcripts into mineral transporter gene families and quantify their expression over the life cycle in different tissues as a first step to finding correlations between gene expression and mineral translocation through tissues. This will allow focused future studies to pinpoint the specific roles of individual genes during plant development.

3.4.1 Discovery of potential switchgrass transporters

In mining the switchgrass genome for mineral transporter gene family members we found a total of 520 genes belonging to 40 different classes in the current annotation of the switchgrass genome. The number of switchgrass genes was approximately twice as many as identified in the annotated genomes of *Sorghum bicolor* (274) and *Setaria italica* (281) (Table 1). Our results indicate that for the most part, the switchgrass genome (tetraploid, A and B genomes) contained about twice as many genes in each class (Table 1), although some exceptions were noticed. For example, four putative copper transporters (*COPT*) [43] were identified in the switchgrass genome, as compared to 1 each in sorghum and *Setaria*. Likewise, 7 potassium transporters (*HAKs*) [44] were found in *Sorghum* and *Setaria*, whereas 9 putative *HAKs* were identified in the switchgrass genome. As anticipated, the switchgrass genome contained large numbers (>10 genes per genome) for many classes of transporters including those for nitrate, phosphate, S, K, Mg, and putative peptide/nitrate transporters.

3.4.2 Differential regulation of transporter genes in switchgrass tissues over development
Both tissue and temporal specificity in the expression of putative transporter genes in switchgrass was observed. Transcriptome datasets generated for greenhouse grown switchgrass cv Alamo (Figure 1) were mined for the relative abundance of transcripts for transporters shown in Table 1. There appeared to be both tissue and stage specific expression for many transporter genes at the three stages of harvest, early vegetative, shoot elongation and reproductive (Figure 1). Most transporter transcripts had different apparent abundances over plant developmental stages for the roots and shoots, and in flowers at reproductive stage (Figure 1). In roots, a cluster of transporters with high transcript counts were observed at the early vegetative stage of harvest (cluster A1). Several genes associated with this cluster were downregulated at the shoot elongation stage, and a larger cluster of transcripts were upregulated in roots at the shoot elongation stage (Figure 1; cluster A2). At the reproductive stage of plant growth, a new set of transporter genes was more abundant in the roots (cluster A6), and there was an apparent downregulation of many of the genes present in greater abundance at the early vegetative and shoot elongation stages of plant growth. A few genes appeared to be upregulated at the early vegetative stage in shoots as compared to roots. At the shoot elongation stage of plant development, most transporter genes were less abundant in shoots as compared to roots, and also less abundant than in shoots at the early vegetative stage (Figure 1). Interestingly, transcripts for a cluster of transporter genes had higher abundance in reproductive stage shoots (cluster A5). These included genes that appeared to be primarily expressed in shoots and some that overlapped with roots and reproductive structures. Reproductive tissues contained greater levels of transcripts for a cluster of transporter genes that were less abundant in roots or shoots (Figure 1, cluster A4).
Figure 1. Two-Way clustered expression profile map of genes of switchgrass nutrient transporters present in different tissues of cv Alamo plants grown in a greenhouse. Stages of plant development are early vegetative (EV), stem elongation (SE), and reproductive (RP) are as described in these datasets. Yellow indicates high abundance, gray is intermediate and black is low or negligible abundance. The clusters are labeled as A1 through A7. The numbers in parentheses under each cluster indicate the total number of transporter identified within each cluster. The appropriate SRA identification numbers for these individual NGS files are SRX057826, SRX057827, SRX057828, SRX057829, SRX057830, SRX057831, and SRX057834.
We next evaluated expression profiles of several transporter gene families in field grown crown and rhizome tissues from cv Summer plants at different stages during a growing season. Of the 520 total mineral transporter genes identified in the switchgrass genome, transcripts for 401 mineral transporter genes were detected in the crown and rhizome datasets (see Table 1). As observed for the greenhouse grown cv Alamo datasets, some gene clusters were up- or downregulated at certain harvest dates (Figure 2). Some transcripts that were abundant early in the growing season (spring green up; May cluster C5) were less abundant later, suggesting that these genes are important for rapid growth in the spring. Enrichment of specific transporter classes (GO-terms) was not observed in these clusters. Other subsets of transporter genes were strongly upregulated during the periods of active shoot and rhizome growth (June and July, clusters C2 and C3), suggesting that genes in this cluster are important as plants continue to grow and progress to the reproductive development stage. At the July harvest plants were heading, with inflorescences visible at the top of the shoots. Although some of these genes were apparently being transcribed at continued high rates, a new cluster of transporter genes was upregulated at the August harvest date (Figure 2, cluster C1), when the plants were nearing physiological maturity, suggesting that these genes could be important for moving minerals to developing seeds. In crowns and rhizomes obtained from plants after a leaf killing frost in October, many of the transporter genes that were upregulated at the earlier harvest dates had decreased, whereas a new cluster of genes had increased transcript counts (cluster C4). This is a particularly interesting pattern, as these transporters are likely to be important for mineral storage or translocation of minerals to perennial storage tissues such as roots and rhizomes.
Figure 2. Two-Way clustered expression profile map of genes of switchgrass nutrient transporters present in crown and rhizome tissues of field grown cv Summer plants harvested at different times during the growing season. Yellow indicates high abundance, gray is intermediate and black is low or negligible abundance. The clusters are labeled as C1 through C5. The numbers in parentheses under each cluster indicate the total number of transporter identified within each cluster. The appropriate SRA identification numbers for these individual NGS files are SRX257007, SRX257030, SRX257031, SRX102934, and SRX257032.
Comparison of the gene members for each cluster in the two 454 transporter datasets showed significant overlap. Approximately 40% of the genes expressed early in the growing season in crown and rhizomes (clusters C5 and C3) are also found expressed in roots during the early vegetative and shoot elongation stages in Alamo (clusters A1, A2, and A3). Similarly, approximately 30% of the genes expressed during the reproductive period in crowns and rhizomes (clusters C2 and C1) are expressed in Alamo roots during the reproductive stage (clusters A6 and A7). These metadata analyses from both greenhouse grown cv Alamo and field grown cv Summer plants indicated that nutrient transporters were transcriptionally controlled at the tissue level and expression was influenced by the developmental stage of the plant. It will be interesting and useful to compare the gene expression data to changes in mineral concentrations over the seasonal growth and senescence of switchgrass.

3.4.3 Expression profiles of selected mineral transporter gene families

Phylogenetic relationships and expression levels for different classes of transporter genes were analyzed in crown and rhizome datasets. *HAK/KUP/KT* genes encode $K^+/H^+$ symporters [45] and are associated with the uptake of $K^+$ into roots and efflux from vacuoles [46]. *KUP* genes are involved with a number of different aspects of plant development [44], and expression of *KUP* genes throughout the plant [45] indicates roles in many tissue and cell types. A total of 33 *KUP* genes were found in our scan of the switchgrass genome, and transcripts ascribable to 29 of these genes were expressed in the crowns and rhizomes of field grown cv Summer plants (Figure 3A). Switchgrass and *Arabidopsis thaliana* annotated *KUP* sequences were separated into six clades. Two genes that share similarities to *AtKUP4*, *Pavirv00030241* and *Pavirv00010539*, were
Figure 3. Cladogram of the phylogenetic relationships between switchgrass transporter genes to *Arabidopsis thaliana* orthologs and their expression profiles in crowns and rhizomes of field grown cv Summer plants. (A) Potassium transporters (KUP) (B) Magnesium transporters (MRS). Other details are as described for Figure 2.
most highly expressed in the August harvest. Transcripts for switchgrass KUPs falling in the clade with AtKUP2 were all overrepresented at the onset of spring growth. For the other KUPs, gene expression within clades was variable, although patterns associated with specific harvest dates were evident. For example, all the switchgrass KUPs with sequence similarities to AtKUP7/13 proteins were overrepresented in the August or November harvests, whereas transcript abundances for the larger clade of switchgrass KUPs with protein sequence similarities to AtKUP5/10 were more variable (Figure 3A).

Magnesium is transported by members of the MRS2/MGT family, with expression in Arabidopsis noted in roots, leaves and senescing leaves, flowers, and pollen [12], and subcellular localizations in plasma membrane, tonoplast, mitochondria, ER and chloroplast. Most of the Mg in leaves is associated with ribosomes involved in protein synthesis, with the majority of the remaining fraction associated with chlorophyll [13]. As such, it is not surprising that transcripts of several MRS2 magnesium transporters were most abundant during the active shoot growth phase (May-July) (Figure 3B). However, four and two genes were highly expressed in the August and October harvests, respectively, suggesting that Mg or related mineral transport was active in the crowns and rhizomes at a time when the above ground parts of the plants were senescing (August) or fully senescent (October). These changes could arise potentially from Mg transported from roots and/or sequestration of Mg into different cellular compartments of the rhizome. It is conceivable that tiller buds and other meristematic tissues present on these rhizomes are metabolic sinks, and transport processes are associated with the continued delivery of nutrients to these critical organs.

Recycling of N in switchgrass is a major factor in sustainability and
environmental impacts of production, and a number of studies have addressed this issue from production and genetic perspectives (reviewed by [47]. Nitrate is a major source of N for plant roots, which is taken up by in part by nitrate transporters of the NRT family. Once nitrate is inside the plant, NRT transporters are also involved in xylem loading and unloading, phloem loading, and storage in vacuoles [48, 49] and are essential for the translocation of plant defense compounds to the seeds [50]. Transcripts were detected for 22 putative switchgrass nitrogen transporters from a total of 35 identified in the switchgrass genome (see Table 1). The Arabidopsis NRT2.4 gene is expressed in both the shoots and roots of nitrogen-starved plants [51] and functions as a high affinity N transporter. Five switchgrass proteins with strong homology (e values of 0 to 6 x 10^{-170}) to Arabidopsis NRT2.4 were identified. Of these, transcripts for Pavirv00019393 (Figure 4A) were most abundant in the July harvest when the plants had reached anthesis, whereas Pavirv00068021 was overexpressed at the August harvest, when seeds were nearing physiological maturity. A majority of the other switchgrass NRTs (14) were most abundantly expressed during the active phase of shoot and rhizome growth (May-July harvests; Figure 4A). The other 6 NRTs were overexpressed in the August and October harvests. Pavirv00010339 had higher expression in crowns and rhizomes of cv Summer plants at the October harvest date, and appears to be orthologous to the Arabidopsis NRT1.6 and 1.7 proteins. NRT1.7 is a low-affinity nitrate transporter involved in source to sink mobilization of nitrate via the phloem [34]. In a similar manner, Pavirv00039672, with homology to the Arabidopsis NRT1.5/NRT1.8 transporters, is also upregulated in the October harvest (Fig 4A). Arabidopsis NRT1.5 modulates the allocation of nitrate to the roots to mediate stress responses in concert with NRT1.8 and other proteins [52].
Figure 4. Cladogram of the phylogenetic relationships between switchgrass transporter genes to Arabidopsis thaliana orthologs and their expression profiles in crowns and rhizomes of field grown cv Summer plants. (A) Nitrate transporters (NRT) (B) Ammonium transporters (AMT). Other details are as described for Figure 2.
Ammonium is an additional source of N for plant growth. Ammonium is taken into roots by AMT1 or AMT2 family transporters [53]. Ammonium is also generated in tissues by reduction of nitrite or breakdown of amino compounds, some of which are remobilized during leaf senescence, and may need to be transported into chloroplasts for reassimilation. In Arabidopsis [53], and poplar [54] AMT genes are expressed in roots, shoots, and flower structures. Switchgrass ammonium transporters were separable into four clades based on Arabidopsis (AtAMT) protein sequences (Figure 4B). Of the 13 AMT sequences in the switchgrass genome, transcripts for 7 genes were identified in the crown and rhizome. Five were more abundantly expressed during the active growing phases (May-July). Two genes overexpressed in crowns and rhizomes at the October harvest, *Pavirv00064059* and *Pavirv00023098* were orthologs of AtAMT2, which is a high affinity ammonium transporter in both shoots and roots [55].

In Arabidopsis, phosphate transporters of the PHT family are classified into 4 subfamilies [56] all of which had homologs in switchgrass. *PHT1* genes are primarily expressed in roots, where they are thought to take up phosphate from soil or mycorrhizal fungi, but are also expressed in leaves and pollen. *PHT2;1* is expressed primarily in the leaves and is thought to transport phosphate into leaves [57]. PHT3 proteins are found in a variety of cellular membranes including mitochondria [58]. *PHT4* genes are mainly targeted to the plastids or Golgi [59] and are expressed in both roots and leaves. In rice [56] and Arabidopsis [60] *PHT* genes were expressed in numerous tissue types during the life cycle. Transcripts for ~71 % of the total *PHT* genes, in the switchgrass genome were found in the crown and rhizome of field grown switchgrass. Expression of the *PHTs* predominantly tracked with active growth phases of the plant (May-July harvests),
although a smaller subset was upregulated in tissues harvested near physiological maturity (August, Figure 5A). Many of these genes clustered with Arabidopsis protein sequences belonging to the PHT3-2, 4-1 and 4-4 genes. These phosphate transporters are thought to be involved in a number of plant processes, involving the shuttling of P\textsubscript{i} across plant compartments. In crowns and rhizomes, this could involve both the acquisition and transport of phosphate from the soil to the developing shoots during the growing season as well as potentially in the redistribution of P\textsubscript{i} at the end of the growing season. Notably, transcripts for Pavirv00062983 (orthologous to AtPHT4-4) were upregulated in tissues obtained after a killing frost (October; Figure 5A). AtPHT4 has been implicated in the movement of Pi between the cytosol and plastids [59].

Silicon is important for resistance to abiotic and biotic stress in grasses [61, 62]. About 90\% of Si taken up by roots is transported into shoots [63]. Lsi1 and Lsi2 are required for uptake [14, 64], and Lsi6 is required for Si transport throughout leaves [15] and through stem nodes [16]. To move from roots to leaves, Si would have to pass through crown tissue. Five genes putatively code for silicon transporters in switchgrass, and transcripts were detected for all five genes in the crown and rhizome tissues (Figure 5B). In contrast to other transporters, silicon transporters genes were essentially downregulated as the growing season progressed. Expression patterns of these putative switchgrass Si transporters are consistent with the movement of Si from the soil to the shoots. The downregulation of all of these Si-transporter genes possibly resulted from death of the shoots (October harvest, Figure 5B).

We also analyzed the expression patterns of YSL and ZIP families of metal micronutrient transporters (Figure 6A; B). Yellow stripe-like (YSL) genes are related to
Figure 5. Cladogram of the phylogenetic relationships between switchgrass transporter genes to Arabidopsis thaliana (PHT) and barley (Hordeum vulgare; LSi) orthologs and their expression profiles in crowns and rhizomes of field grown cv Summer plants. (A) Phosphate transporters (PHT) (B) Silicon transporters (LSi). Other details are as described for Figure 2.
the Yellow Stripe gene that encodes an Fe(III)-phytosiderophore uptake protein in maize roots [65]. However, instead of functioning in uptake, YSL proteins carry out transport of nicotianamine-metal complexes within the plant, including Cu, Fe, Zn, and Mn [32, 66, 67]. Transcripts for all but two of 15 putative YSLs were found in the total crown and rhizome transcriptome dataset (Figure 6A). The expression patterns of these genes were similar to those for the NRTs and PHTs, in that certain genes were upregulated at different harvest dates, suggesting both developmental and tissue regulation. Notably, four of the YSL genes were at higher levels at the August harvests when minerals were being remobilized from shoots. Two Arabidopsis YSL genes have been implicated in remobilization of Cu, Fe, and Zn [1, 32]. The Pavirv00003688 transcript that is most closely related to Arabidopsis YSL2 appeared to be strongly upregulated in the crown and rhizomes harvested in October (Figure 6A). Arabidopsis YSL2 probably functions in the lateral movement of Fe, Cu, and Zn in tissues [68, 69]. Some of the switchgrass YSL paralogs showed a bimodal expression pattern, with higher transcript abundance in May and August harvests. The orthologous Arabidopsis genes AtYSL4 and YSL6 are key for iron homeostasis during plastid ontogeny, thereby modulating plant responses to iron availability [70].

Plant zinc/divalent metal transporters (ZIPs) are members of a relatively large group of related genes that participate in metal transport and homeostasis [71], including Fe [72, 73], Zn [74, 75], and Mn [75], with transport across plasma or vacuolar membranes [75, 76]. Transcripts were detected for nineteen of 27 switchgrass ZIPs in the crown and rhizome transcriptomes (Figure 6B). Two ZIP genes that are orthologous to Arabidopsis ZIP2 and ZIP7 were overexpressed in the July harvests. AtZIP2 is localized...
Figure 6. Cladogram of the phylogenetic relationships between switchgrass transporter genes to *Arabidopsis thaliana* orthologs and their expression profiles in crowns and rhizomes of field grown cv Summer plants. (A) Yellowstripe-like transporters (YSL) (B) Zinc transporters (ZIP). Other details are as described for Figure 2.
on the root plasma membrane and is thought to aid in the loading of Mn and Zn into the xylem [75]. Since above ground growth in plants is active in July, it is possible that the switchgrass orthologs fill a similar role. Transcripts for nine ZIPs were more abundant in the early (predominately May) harvest. Among these were three genes orthologous to the Arabidopsis ZIPs 4, 9, and 12, which are induced by Zn deprivation [77]. A related switchgrass gene (Pavirv00015988) however, was overexpressed in crowns and rhizomes at the end of the growing season (October; Figure 6B). Four switchgrass zinc transporter (ZTP) genes with homology to the Arabidopsis ZTP29 were upregulated in crown and rhizome tissues later in the growing season (August and October). AtZTP29 is localized to the ER and induced in roots in response to salt stress, and is thought to play a role in the unfolded protein response [78].

The distinct temporal changes in transcript abundances in the 454 dataset were validated using qRT-PCR for arbitrarily selected genes that in the 454 datasets (See Figure 3-6) were at higher abundance at a specific harvest date. Six transporter genes and one reference gene were analyzed by qRT-PCR using RNA from field grown plants (Figure 7). In 5 out of 6 genes, transcript abundance by qRT-PCR agreed with the 454 expression datasets. Transcript abundance determined by the two methods corresponded closely for four genes (Figure 7 Panels B, D, E, and F), where a simple regression of transcript counts at each harvest date to the relative quantities of abundance by qRT-PCR yielded $R^2$ values between 0.69 to 0.97. For two others (Figure 7 Panels A and C), the $R^2$ values were 0.37 (A) and 0.21 (C). In the case of the YSL gene (Figure 7 Panel A), the highest abundance for both datasets was in May, but the abundances observed in the August and November harvest dates were different, resulting in the lower correlation
Figure 7. Correspondence between expression profiles observed in NGS datasets to qRT-PCR data for size select switchgrass transporter genes. (A) YSL; (B) HKT; (C) PHT; (D) TPK; (E) KUP; (F) MRS. Colored bar at the top of each panel is the expression profile for each gene observed in the 454NGSdataset (see Figure2). Yellow is high expression and black is low or negligible expression. Gray bars in each panel show how the relative expression (±SD) for each individual gene as determined by qRT-PCR. Different letters above each bar were significantly different expression at $P < 0.05$. See text for more details.
coefficient. The PHT gene Pavirv00039095 (Figure 7 Panel C) had a weak correlation coefficient between the 454 dataset (highest in July) and the qRT-PCR dataset (highest in November). Taken together, these data support the findings presented in Figure 3-6.

3.4.4 Mineral dynamics in crown and rhizome

Mineral concentrations were analyzed in crown and rhizome tissues. Of the sixteen minerals analyzed, As, Fe, Na, Ni, S and Se levels did not change significantly across harvests. The remaining ten minerals with significant difference between any two harvest dates are shown in Figure 8. Excluding P and Ca, all the other minerals had lowest concentrations in crowns and rhizomes harvested in November. For K, Mg, Mn, Zn, Cu, Cd and Co, maximal levels were detected at the August harvest, when the shoots were at an advanced stage of senescence. Highest levels of Mo were found in rhizomes harvested in June. A bimodal pattern of mineral concentrations was seen for K, Ca, P, Mg, and Mn. For K, Mg and Mn, maximal concentrations were observed in tissues harvested in May (Mn) and June (period of active growth) and in August (shoot senescence). For Ca and P maximal amounts were observed for the June and November harvests respectively (Figure 8). These fluctuating concentrations may reflect the passage of minerals through the crowns seasonally. Cu, K, and Zn are minerals known to be remobilized, and were lowest during shoot dormancy when recycled minerals would be stored in roots. Alternatively, since large fluctuations were not observed, mineral concentrations may reflect demands of the tissues themselves for growth and metabolism.

3.5 Conclusions and Future Directions

This research represents a first step in the characterization of mineral transporter genes and associating their expression in a perennial grass. As more data for mineral
Figure 8. Mineral concentrations in crowns and rhizomes of field grown cv Summer plants across harvest dates. Each panel shows the change in mineral concentration for a specific mineral. Analysis for each time point consisted of two technical replicates from each of three biological replicates ($n = 6$). Average abundances with standard deviation error bars are shown in this figure. $P$-values were calculated for each mineral using Single-Factor ANOVA. (A) Potassium; (B) Calcium; (C) Phosphorus; (D) Magnesium; (E) Manganese; (F) Zinc; (G) Copper; (H) Cadmium; (I) Cobalt; (J) Molybdenum.
dynamics becomes available, a clear picture of the genes needed for translocation to shoots in spring and to storage tissues in the fall will emerge. We have identified a number of mineral transporter genes with seasonal expression patterns that give clues to the biology of crown and rhizome tissues as a gateway between shoot growth and mineral storage and uptake. Likewise, this tissue is also a recipient of nutrients remobilized from senescing shoots. We anticipate that as additional transcriptomic and mineral datasets from other switchgrass tissues become available, these resources will be a valuable tool for plant breeders to improve production and sustainability of switchgrass.

3.6 Conflict of Interest Statement

This research was conducted in the absence of any commercial or financial relationships that could be construed as a conflict of interest.

3.7 Acknowledgements

We thank Dr. Javier Seravalli for performing the mineral analysis, and Dr. Jaehyoung Kim for performing pyrosequencing analyses. This work was supported in part by USDA-NIFA Grant Number 2011-67009-30096, The Office of Science (BER), U. S. Department of Energy Grant Number DE-AI02-09ER64829, and by the USDA-ARS CRIS project 5440-21000-030-00D. The U.S. Department of Agriculture, Agricultural Research Service, is an equal opportunity/affirmative action employer and all agency services are available without discrimination. Mention of commercial products and organizations in this manuscript is solely to provide specific information. It does not constitute endorsement by USDA-ARS over other products and organizations not mentioned. All sequences used in this work available from the NCBI. The PviDraft
0.0 sequence data used in this study were produced by the US Department of Energy Joint Genome Institute.
Table 1. Bioinformatic analysis of switchgrass genome for mineral transporter classes.

<table>
<thead>
<tr>
<th>Class</th>
<th>Descriptions</th>
<th>In Ref</th>
<th>In Pvi0</th>
<th>In Sb</th>
<th>In Si</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACA</td>
<td>Ca2+-transporting ATPase</td>
<td>8</td>
<td>22</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>AKT/KAT</td>
<td>Potassium channel</td>
<td>9</td>
<td>18</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>AMT</td>
<td>ammonium transporter</td>
<td>6</td>
<td>13</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>CAX</td>
<td>cation/proton exchanger</td>
<td>7</td>
<td>16</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>CHX</td>
<td>cation/H(+) antiporter</td>
<td>28</td>
<td>26</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td>COPT</td>
<td>copper transporter</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FRO</td>
<td>ferric reduction oxidase</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FRU</td>
<td>ER-LIKE IRON DEFICIENCY-INDUCED TF</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>HKT</td>
<td>sodium transporter</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>HMA</td>
<td>heavy metal ATPase</td>
<td>8</td>
<td>18</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>IREG</td>
<td>iron regulated protein</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>IRT</td>
<td>Fe(2+) transport protein</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>KEA</td>
<td>K(+) efflux antiporter</td>
<td>6</td>
<td>9</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>KUP</td>
<td>potassium transporter</td>
<td>13</td>
<td>33</td>
<td>22</td>
<td>21</td>
</tr>
<tr>
<td>LSI</td>
<td>silicon transporter (from Hv)</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>MHX</td>
<td>magnesium/proton exchanger</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MOT</td>
<td>molybdate transporter</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>MRS</td>
<td>magnesium transporter</td>
<td>11</td>
<td>20</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>MTP</td>
<td>metal tolerance protein</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>NAS</td>
<td>nicotianamine synthase 1</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>NAXT</td>
<td>nitrate excretion transporter</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>NHX</td>
<td>sodium/hydrogen exchanger</td>
<td>8</td>
<td>15</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>NRAMP</td>
<td>metal transporter</td>
<td>6</td>
<td>16</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>NRT</td>
<td>nitrate transporter</td>
<td>16</td>
<td>35</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td>NTRm</td>
<td>misc NTR-class (major facilitator)</td>
<td>16</td>
<td>32</td>
<td>15</td>
<td>17</td>
</tr>
<tr>
<td>OPT</td>
<td>oligopeptide transporter</td>
<td>9</td>
<td>15</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>PHO</td>
<td>phosphate transporter</td>
<td>11</td>
<td>9</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>PHT</td>
<td>inorganic phosphate transporter</td>
<td>19</td>
<td>45</td>
<td>22</td>
<td>24</td>
</tr>
<tr>
<td>PNT</td>
<td>putative peptide/nitrate transporters</td>
<td>23</td>
<td>48</td>
<td>25</td>
<td>23</td>
</tr>
<tr>
<td>PTR</td>
<td>peptide transporter</td>
<td>5</td>
<td>11</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>SULTR</td>
<td>sulfate transporter</td>
<td>12</td>
<td>23</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>TPC</td>
<td>two pore calcium channel protein</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>TPK</td>
<td>Ca-activated outward-rectifying K channel</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>VIT</td>
<td>vacuolar iron transporter</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>YSL</td>
<td>metal-nicotianamine transporter</td>
<td>8</td>
<td>15</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>ZIP</td>
<td>zinc transporter</td>
<td>13</td>
<td>27</td>
<td>14</td>
<td>11</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>277</strong></td>
<td><strong>520</strong></td>
<td><strong>274</strong></td>
<td><strong>281</strong></td>
</tr>
</tbody>
</table>
3.8 Works Cited


47. Schwartz C, Amasino R: **Nitrogen recycling and flowering time in perennial bioenergy crops.** *Frontiers in plant science* 2013, **4**:76.


52. Chen CZ, Lv XF, Li JY, Yi HY, Gong JM: **Arabidopsis NRT1.5 is another essential component in the regulation of nitrate reallocation and stress tolerance.** *Plant Physiol* 2012, **159**(4):1582-1590.


Chapter 4

Contrasting metabolism in perenniating structures of upland and lowland
switchgrass plants late in the growing season

Note: The results described in this chapter have been published and all text is modified
from the original version. Supplemental data can be found online at the journal’s
website.

The citation is: Palmer NA, Saathoff AJ, Tobias CM, Twigg P, Xia YN, Vogel KP,
Madhavan S, Sattler SE, Sarath G: Contrasting Metabolism in Perenniating
Structures of Upland and Lowland Switchgrass Plants Late in the Growing Season.
Plos One 2014, 9(8).
4.1 Abstract

**Background:** Switchgrass (*Panicum virgatum* L.) is being developed as a bioenergy crop for many temperate regions of the world. One way to increase biomass yields is to move southern adapted lowland cultivars to more northern latitudes. However, many southerly adapted switchgrass germplasm can suffer significant winter kill in northerly climes.

**Materials and Methods:** Here, we have applied next-generation sequencing in combination with biochemical analyses to query the metabolism of crowns and rhizomes obtained from two contrasting switchgrass cultivars. Crowns and rhizomes from field grown lowland (cv Kanlow) and upland (cv Summer) switchgrass cultivars were collected from three randomly selected post-flowering plants. Summer plants were senescing, whereas Kanlow plants were not at this harvest date.

**Results:** Principal component analysis (PCA) differentiated between both the Summer and Kanlow transcriptomes and metabolomes. Significant differences in transcript abundances were detected for 8,050 genes, including transcription factors such as WRKYS and those associated with phenylpropanoid biosynthesis. Gene-set enrichment analyses showed that a number of pathways were differentially up-regulated in the two populations. For both populations, protein levels and enzyme activities agreed well with transcript abundances for genes involved in the phenylpropanoid pathway that were up-regulated in Kanlow crowns and rhizomes. The combination of these datasets suggests that dormancy-related mechanisms had been triggered in the crowns and rhizomes of the Summer plants, whereas the crowns and rhizomes of Kanlow plants had yet to enter dormancy.
**Conclusions:** Delayed establishment of dormancy at more northerly latitudes could be one factor that reduces winter-survival in the high-yielding Kanlow plants. Understanding the cellular signatures that accompany the transition to dormancy can be used in the future to select plants with improved winter hardiness.
4.2 Introduction

Switchgrass (*Panicum virgatum* L.), a perennial C$_4$-grass native to the continental USA, is being developed as a major biomass feedstock for use in temperate regions [1]. Latitudinal adaptation has resulted in ecotypes with distinguishable genetic make-up [2] and differential responses to photoperiod [3]. Moving switchgrass germplasm more than one hardness zone north or south from its native adaptation zone or latitude generally results in significantly reduced biomass yields. This is caused by early flowering in the northern adapted plants when moved to southern latitudes and reduced winter survival of southern plants when grown in northern latitudes because of delayed flowering and senescence [3].

For switchgrass, perenniality and sustainability appear to be strongly interlinked and are significantly influenced by photoperiod [3, 4]. Photoperiodic cues likely influence the orderly developmental transition from spring emergence of shoots through the senescence of the aerial parts of the plant and the imposition of dormancy in the below-ground parts of the plant post seed ripening [5]. As with other perennial warm-season grasses, the above ground parts of the switchgrass plant senesce at the end of the growing season, whereas the below-ground tissues, (comprised of the crowns, rhizomes and roots), transition to a dormant state. Tillers buds originating from the crowns and rhizomes are the source of the eventual regrowth of the above ground portions of the plant in the following growing season. Little is currently known at a molecular or cellular level on these aspects of switchgrass biology; however, if these processes can be understood it will be possible to improve both breeding and management of this crop. This knowledge can be used not only to improve the sustainability of switchgrass and
other temperate perennial biomass grasses, but also potential perennial grain crops [6] which will face similar trade-offs of maximizing grain yield while maintaining crown and rhizome vitality. Seasonal senescence and dormancy has been studied in a number of other species [7-10], and these studies offer a scaffold to interpret data obtained from understudied species such as switchgrass.

This study focused on two tetraploid cultivars, namely, “Summer” that is adapted to the northern Great Plains of the USA and a contrasting cultivar “Kanlow” that is adapted to the southern latitudes of the USA. The cultivar Summer was developed at Brookings, South Dakota (~44° N), USA from a native collection that originated from southeast Nebraska, USA and was first released in 1963 [11]. The cultivar Kanlow was also first released in 1963 from Manhattan, Kansas (~39° N), USA and was developed from 200 plants originally sourced from a lowland site near Wetumka, Oklahoma, USA. Kanlow plants display adaptation to wet sites [11]. There is about a 26 cm differential in average rainfall between SE Nebraska (origin of cv Summer, drier) and Wetumka, OK (origin of Kanlow, wetter). Although we cannot rule out environmental-specific transcriptional differences between the two populations, our intention was to focus on understanding any stage specific differences. Little, if any, selection pressure was applied during the generation of the cultivars, and as with most switchgrass cultivars, both Summer and Kanlow are synthetic populations that display diversity for many plant characteristics including height, leaf morphology, phenology and genotype as has been described in other publications [1-3, 5, 12]. Tetraploid switchgrasses exhibit higher yields than octoploid switchgrasses, and are providing germplasm for the development of improved plant materials for the biofuel sector [13, 14]. Within the tetraploid
switchgrasses, the lowland cultivars such as Kanlow out-yield upland tetraploids such as Summer, but suffer from significant winter-kill in more northerly latitudes limiting their potential to be grown in these regions [1, 3]. The aims of this study were to utilize RNA-seq with some biochemical analyses to develop knowledge into the metabolism of crown and rhizome tissues near the end of the growing season. The combination of these unbiased datasets obtained from the crowns and rhizomes from field-grown plants suggests that dormancy-related mechanisms had been triggered in the crowns and rhizomes of the Summer plants, whereas the crowns and rhizomes levels had yet to enter dormancy. Recently, RNA-seq has been used to map transcripts and discover single nucleotide polymorphisms (SNPs) in several different populations of switchgrass [15]. In this study the authors were able to efficiently map a large number of sequencing reads and identify over one million SNPs. A similar approach has been used in our study to understand the transcriptomes in the two contrasting switchgrass cultivars.

4.3 **Materials and Methods**

4.3.1 **Plant materials, growth conditions, and selection of harvest date**

Stands of switchgrass cv Summer and cv Kanlow were established in the fields of the University of Nebraska-Agricultural Research Division near Mead, NE (~41°N) in 2009, using seedlings raised in the greenhouse and transplanted to the fields. Each plot (1m × 1.2 m) contained 12 closely-spaced plants to mimic sward densities. Plots were in a field that measured approximately 35 m × 30 m [16]. Agronomic management was as described earlier [17]. Plants were not irrigated and depended on ambient rainfall and stored soil moisture for growth. Climatic conditions for a few weeks prior to, and after collection are shown in Figure S1 in File S1. Daytime temperatures were generally
greater than 20°C for the ~4 weeks prior to harvest and low temperatures were generally 10° to 15°C. Three continuous days of rainfall were recorded three days prior to harvest. Soils were moist at harvest (Figure S1 in File S1).

At this location, Summer plants reach 50% anthesis generally by early July and reach physiological maturity by early-mid September. In contrast, for Kanlow plants, 50% anthesis is reached by mid-August and physiological maturity in late September/October. We did not control for plant size (although Kanlow plants were always larger than Summer plants), or measure above-ground yields. Individual plants planted in 2009 were still distinct in 2010, and the crowns and rhizomes were collected from three individual plants of each cultivar as described earlier [16] on September 29, 2010. Data were not collected on the above ground material except to visually verify seed developmental stage and if there were obvious signs of plant senescence (leaf yellowing, tiller death, etc.). Such signs of senescence were present in Summer plants, but were not observed for the Kanlow plants. This developmental state was selected as a starting point to understand the changes in the below-ground metabolism of switchgrass that impacted transition into winter dormancy. At harvest, tissues were flash frozen in liquid nitrogen and later pulverized cryogenically in order to obtain high yields of total RNA [16].

4.3.2 RNA-Seq

Total RNA was extracted from tissues with TRIzol Reagent (Invitrogen Corp, Carlsbad, CA), followed by further purification by Qiagen RNeasy Mini Kit (Qiagen, Valencia, CA). Sequencing libraries were prepared from poly(A)+RNA using the Illumina mRNA-Seq Sample Prep Kit according to the manufacturer’s instructions.
Libraries were then sequenced on the Illumina Genome Analyzer IIx using two 36-cycle sequencing kits to read 75 nucleotides of sequence from a single end of each insert, by the v8 protocol (Illumina Inc, San Diego, CA). Each sample was analyzed on a single lane. The purity, integrity, and profile of extracted total RNA was verified on an Agilent 2100 BioAnalyzer (Agilent Technologies, Inc., Santa Clara, USA), and for concentration using a NanoDrop instrument (Thermo Scientific, Waltham, USA). Once generated, quality of the libraries was verified by gel electrophoresis and on a NanoDrop spectrophotometer. The PhiX virus DNA library was used as a control in a sequencing lane on each flowcell run to monitor sequencing operation. Obtained sequences were trimmed for adapter sequences and verified for quality using the software available with the instrument.

4.3.3 Mapping and differential gene expression analysis

Illumina reads were mapped to the draft switchgrass genome available at www.phytozome.net [18], using splice-junction aware Tophat2 (version 2.0.11) [19], with Bowtie2 (version 2.2.3) [20] alignment. Default parameters were used and reads flagged as having multiple alignments were not included in gene expression. Gene expression counts were calculated using featureCounts [21], a script included as part of the Subread (version 1.4.4) [22] analysis package, and the gene annotation file provided with the draft genome release. Differential expression analysis was conducted using the program edgeR (version 3.4.2) [23] in R [24]. Very low expression genes were filtered out of the data set prior to analysis by requiring 1 count per million (cpm) in at least three of the six samples. An FDR threshold of 0.05 was used to determine differentially expressed genes. Variance stabilizing transformation values, calculated using the
DESeq2 (version 1.2.8) [25] package in R, were used for gene expression quantities in further downstream analyses.

4.3.4 Genome functional annotation

Draft genome nucleotide transcript sequences were functionally annotated by sequence similarity using the program Blast2GO [26] similarly to our earlier work [16]. Blast2GO annotations with an e-value threshold of $1 \times 10^{-15}$ or lower resulting from a Blastx search against the NCBI non-redundant protein database were used in all the subsequent analyses, including Enzyme Commission (EC) (http://www.chem.qmul.ac.uk/iubmb/enzyme/) and KEGG pathways [27, 28].

4.3.5 Gene set enrichment analysis

Gene Set Enrichment Analysis was done using GSEA from the Broad Institute [29]. Gene sets were created from Blast2GO generated KEGG pathway assignments.

4.3.6 Metabolite analyses

Aliquots (100 ± 2 mg) of ground crown and rhizome tissue were extracted for metabolites, generally following a previously described protocol [30]. Metabolites were separated and identified (where possible) by GCMS [31] using the Agilent Fiehn GC/MS Metabolomics RTL library as the basis for instrument operation and metabolite identification. Metabolite identification was accomplished using the Fiehn libraries in Chemstation and in the supplied version of AMDIS. Metabolites were extracted from 100 mg of tissues with 350 µL 100% methanol and 20 µL of 0.2 mg mL$^{-1}$ ribitol in water, as an internal standard. Samples were then heated in a 70°C heat block for 15 minutes. After heating, 1 volume (370 µL) DI water (18 MΩ) was added and each tube was vigorously mixed and then centrifuged at 14,000 RPM for 10 min. The supernatants
were placed into new 1.5 mL tubes, and polar and nonpolar metabolites separated by two chloroform washes (300 µL each time). The final upper phase (methanol and water) was carefully pipetted into a new microfuge tube and 25 µL aliquots from each tube were then placed into 2 mL glass vials. Next, 5 µL of 0.4 mg mL\(^{-1}\) myristic-d\(_{27}\) acid (#366889 Sigma-Aldrich Co.) in 100% methanol and 20 µL of 0.2 mg mL\(^{-1}\) docosane (#134457 Sigma-Aldrich Co.) in dichloromethane were added to each vial; these were used as a retention locking compound and internal standard, respectively. The vials were evaporated to dryness under vacuum (Labconco Inc.). The dried samples were derivatized using 30 µL of 40 mg mL\(^{-1}\) methoxyamine hydrochloride at 37°C for three hours followed by 30 min. of trimethylsilylation at 37°C using 90 µL \(N\)-methyl-\(N\)-(trimethylsilyl)trifluoroacetamide (Thermo Scientific Inc.). The derivatized sample was placed into a new 2 mL glass vial with an insert due to the small volumes. The GC column was a 30-meter, 0.250 mm I.D. HP-5MS column (Agilent Technologies, Inc.). The mass selective detector was auto-tuned regularly with perfluorotributylamine as recommended by the manufacturer and operated over a \(m/Q\) scan range of 50 – 600. Temperature settings were 250°C for the injector and MS source and 290°C for the transfer line. The initial oven temperature was 60°C, which was increased at a rate of 10°C per minute to 325°C. Helium was the carrier gas and the flow rate was 1.1 mL min\(^{-1}\). The injector was operated in splitless mode. Since only major metabolites were of interest, features that had a mean major ion count of less than 1.0×10\(^5\) across both genotypes were discarded. Also, features that were missing in more than 10 of the 27 GC-MS runs per cultivar were flagged; features that were flagged this way across both cultivars were eliminated from the data set since they were not reliably detected across
the biological and technical replicates. After this filtering, a total of 418 features remained in the data set, although not all of these features could be reliably identified by Chemstation or AMDIS.

4.3.7 Enzyme, protein, and immunoassays

Total soluble protein was extracted by sonicating approximately 100 ± 2 mg of pulverized crown and rhizome tissue with 600 µL of Tris-HCl buffer, pH 8.0 containing a plant protease cocktail (Sigma P9599) with 5 mM DTT for cinnamyl alcohol dehydrogenase (CAD) or with 300 µL of Tris-HCl buffer, pH 8.0, without DTT for ascorbate peroxidase (AscPx) and caffeoyl-O-methyl transferase (COMT). For phenylalanine ammonia lyase (PAL), tissues were extracted using 0.1 M borate buffer, pH 8.8. Crude homogenate was desalted using a spin column (Thermo Scientific 89889, 2 mL columns) and used directly for enzyme assays. Protein content was measured using a dye-binding assay (Pierce 660 nm Protein assay kit) using BSA as a standard. Gel electrophoresis, protein transfer and immunodetection were as described previously [32]. Chemiluminescence was detected using a BioRAD Chemidoc system. Enzyme activities were analyzed as described previously [32-34].

4.3.8 Statistical analyses

Statistical analysis of metabolite data was conducted using SAS 9.3 (SAS Institute, Inc., Cary, NC) and JMP 9.0 (SAS Institute Inc., Cary, NC). Pairwise comparisons across the two cultivars used the Wilcoxon rank-sum test in PROC NPAR1WAY, and the raw p-values were adjusted in PROC MULTTEST to correct for multiple comparisons using the false discovery rate [35]. JMP 9.0 was used to conduct principal component analysis and generate the color map. ANOVA testing was
performed on other datasets using statistical routines available in EXCEL.

4.4 Results

4.4.1 80% of all reads map to the switchgrass draft genome

The mapping summary for the three biological replicates obtained from field grown Summer and Kanlow plants are shown in Table S1 in File S1. Each sample yielded approximately 24-27 million reads, of which greater than 77-83% were mapped to the 0.0 early-release version of the switchgrass genome. Functional annotation of the PviDraft0 genome by Blast2GO resulted in the annotation of 45,487 gene models at an e-value threshold of \(1 \times 10^{-15}\) or lower (Table S2 in File S1).

4.4.2 Transcriptomes of Summer and Kanlow crowns and rhizomes are significantly different

EdgeR analysis identified a total of 8,050 genes that were differentially expressed (File S2) between Kanlow and Summer crown and rhizomes at an FDR <0.05. A principal component analysis (PCA) of this expression dataset effectively separated the three Kanlow samples from the three Summer samples (Figure 1A). Component 1 accounted for 43% of the variance and separated the two cultivars. Component 2 accounted for approximately 18% of the variance and separated the three Summer samples. A Venn diagram of the total number of genes found in the Summer and Kanlow datasets is shown in Figure 1B. In total transcripts were detected for 36,572 genes. Of this total ~84% (30,771 genes) were common to both populations and 2,303 and 3,498 genes were unique to the Summer and Kanlow datasets respectively (Figure 1B). About 41% of the differentially expressed genes (DEGs) had two-fold or greater average expression in Kanlow, and ~26% had two-fold or greater expression in Summer (Figure
A more detailed evaluation of the DEGs was performed to identify changes in transcript abundance for three plant-specific classes of transcription factors, namely MYBs, NACs and WRKYs which are known to have multiple functions in cellular processes [36-40]. A scan of the switchgrass genome based on Pfam PF03106 (WRKY), PF00249 (MYB) and PF02365 (NAC) domains [41], identified 178 WRKYs (~89 WRKYs per diploid genome); 324 MYBs (~162 MYBs per diploid genome); and 239 NACs (~120 NACs per diploid genome). Similar searches for WRKY members in other species gave the following results: Arabidopsis – 72; rice – 94 and 82 in Brachypodium. Using a similar approach for MYBs and NACs in other species, we found: Arabidopsis – 256/112 (MYBs/NACs); rice – 233/140 (MYBs/NACs), and Brachypodium – 200/100 (MYBs/NACs). Based on the relative numbers of these genes found in model plants, we expect that approximately 80-90% of the orthologous switchgrass genes were identified by this search. A total of 199 MYBs (30 DEG) and 108 NACs (20 DEG) were found in the expression datasets. In all 119 WRKYs (~67% of the all WRKY genes identified in the genome) were found in the Summer and Kanlow datasets. Of these 119 genes, a total of 12 were differentially expressed at an FDR<0.05 (Figure 1C). A database search indicated that orthologs to four of the WRKY genes upregulated in Kanlow (identified by a star symbol in Figure 1C) were involved in responses to pathogens. In contrast, the ortholog to one WRKY gene upregulated in Summer and has been demonstrated to be an important mediator of cellular responses to senescence (inverted triangle).
Figure 1. Transcriptomes of Kanlow and Summer plants are different, and show differential enrichment in WRKY genes. (A) PCA of transcriptomes of the three biological replicates obtained from each ecotype. Kanlow (orange triangle) were differentiated from Summer (cyan squares) by the first component. (B) Venn diagram of numbers of common and highly expressed genes in Summer (log$_2$FC < -1) and Kanlow (log$_2$FC > 1), including DEGs in the Kanlow (orange) and Summer (cyan) datasets. (C) Mean counts for twelve transcripts identified as WRKY transcription factors. Stars above orange bars (Kanlow) identify putative switchgrass orthologs of WRKY factors involved in defense response in other systems. The inverted triangle over a cyan bar identifies a WRKY ortholog to an Arabidopsis gene involved in senescence that is upregulated in Summer crowns and rhizomes.
4.4.3 Metabolite analysis differentiates Kanlow from Summer crowns and rhizomes

Principal component analysis of a broad-based polar metabolite profiling of Kanlow and Summer crown and rhizome extracts using GC-MS revealed that Summer and Kanlow extracts were clearly differentiated by the first principal component, which explained 27% of the variance (Figure 2A). The second component differentiated among the individual genotypes within each cultivar and explained an additional 13% of the variance. Concentration ratios, approximated by major ion count, were coupled with differences that had an FDR <0.05 to generate a volcano plot of the metabolite data (Figure 2B). Concentration ratios between the cultivars were considered meaningful if the averaged ion counts for a metabolite had a ratio greater than 1.75 or less than 1.75⁻¹. The stricter criteria resulted in 66 metabolites with sufficient FDR values and ion area ratios in Summer and 47 metabolites in Kanlow. Hierarchical two-way clustering was used to generate an initial color map of the overall metabolite profile (Figure S2 in File S1). Following this analysis, we generated the heat map shown in Figure 2C by using one-way hierarchical clustering based on cultivar, and manually reordered metabolites into more biologically related groupings which could be indicative of differences in their metabolism (Figure 2C A-G). In Summer plants, levels of several amino acids (Figure 2C group A), amine derivatives (group B) and some sugar alcohols (group C) were significantly higher. In contrast, erythrose-4-phosphate, maltotriose, sucrose and trehalose were detected in greater levels in Kanlow crowns and rhizomes. Further significant differences (based on an FDR <0.05) in the metabolism between Kanlow and Summer rhizomes were evident in the observed levels of several organic acids such as
Figure 2. Metabolite profiling reveals ecotype specific differences in crown and rhizome tissues. (A) PCA of overall metabolite profiles observed by GCMS for each of three individual genotypes within each cultivar. Kanlow tissue extracts (orange triangles) are separated from Summer tissue extracts (cyan squares) by the first component. Two-way error bars are shown for each plant that was based on nine separate GCMS runs for each sample. (B) Volcano plot showing the log_{10}FDR versus the log_{1.75} fold change in peak area for major ions for all metabolites detected by GCMS. Significant differences were observed for several metabolites between the two cultivars (Kanlow orange triangles) and Summer (cyan squares). Grey circles are metabolites that failed to show sufficient differences in ion area of had an FDR value of >0.05. (C) Heat map shows marked differences in tissue abundances of selected metabolites in Kanlow and Summer crowns and rhizomes. Data are the average of triplicate injections from each of three separate extractions from the three biological replicates of each cultivar. Scale is from high abundance (yellow) to low abundance (black) for each metabolite. Data were subjected to one-way hierarchical clustering based on cultivar, and metabolites were manually reordered into more biologically related groupings (A-G).
glucuronic acid, a precursor of xylans [42], and a range of secondary metabolites, including phenolic acids (Figure 2C, group D,E). Summer tissues also appeared to be in a more oxidized state as compared to Kanlow crowns and rhizomes as they contained significantly higher levels of dehydroascorbic acid, gluconic acid lactone, and lower concentrations of α-tocopherol, (Figure 2C, group D, F).

4.4.4 Nexus between gene-set enrichment and metabolite analyses

As a prelude to gene-set enrichment analysis (GSEA), 124 gene sets (Table S3 in File S1) were generated for the predicted transcriptome available in the PviDraft0 using Blast2GO and the Kyoto Encyclopedia of Genes and Genomes (KEGG; http://www.genome.jp/kegg/). These gene sets were then used through GSEA to identify specific pathways differentially regulated in the crown and rhizome tissues of Kanlow and Summer plants. A total of 24 metabolic pathways (FDR <0.2) were different between the two ecotypes (Table S4 in File S1). In Kanlow tissues, there was an apparent upregulation of phenylpropanoid biosynthesis (cell wall accretion), starch, sucrose and fatty acid metabolism. Within Summer plants, there was enrichment (FDR <0.001) in transcripts for enzymes involved in diterpenoid biosynthesis, and a somewhat weaker upregulation for transcripts for enzymes associated with the biosynthesis of unsaturated fatty acids (FDR <0.193). A number of metabolic pathways associated with the degradation and scavenging of carbon compounds were also enriched in Summer crowns and rhizomes (Table S4 in File S1), providing some consistency with metabolite data.

In Kanlow tissues, there was significant enrichment in the transcripts coding for enzymes in the phenylpropanoid (lignin biosynthesis) pathway as compared to Summer
tissues (Figure 3). Key metabolites caffeic acid and ferulic acid were enriched in Kanlow tissues. In Summer crowns and rhizomes, levels of Phe and Tyr, precursors for the phenylpropanoid pathway were significantly greater, consistent with the transcriptomic data.

4.4.5 Enzyme activity, protein levels, and enrichment of key transcription factors confirm GSEA of phenylpropanoid pathway

The phenylpropanoid pathway was targeted to understand if the changes documented for transcript abundance (Figure 3) were also evident at the protein level. Transcripts for several genes associated with the phenylpropanoid pathway were significantly enhanced in Kanlow crowns and rhizomes as compared to the Summer plants. These included PAL, 4-coumarate-CoA ligase (4-CL), CAD, COMT and caffeoyl CoA 3-O-methyltransferase (CCoAOMT) (Figure 4A). It should be noted that transcript abundance for a called gene in Figure 4A is representative of all sequences assigned to this class of gene by Blast2GO, and is likely to contain transcripts for one or more related genes within a given gene family. Abundances varied over an order of magnitude, suggesting that differential regulatory mechanisms might exist for these individual genes. Transcript levels for actin were not significantly different (Figure 4A). AscPx transcripts were slightly elevated in Summer tissues as compared to Kanlow. AscPx is involved in detoxification of ROS in cells. Next, antibodies generated to select enzymes in the lignin biosynthesis pathway were utilized and showed that apparent protein levels for PAL, 4-CL, CAD, COMT and CCoAOMT were lower in all three biological replicates of Summer crown and rhizome extracts as compared to Kanlow extracts (Figure 4B). Levels for s-adenosyl methionine synthetase (SAMS), an enzyme that produces
Figure 3. Transcripts for enzymes and several metabolites associated with the phenylpropanoid pathway are significantly enhanced in Kanlow crowns and rhizomes. Orange boxes identify individual enzymes upregulated in Kanlow. Boxes with EC numbers and relative enrichment of transcripts (Kanlow/Summer, K/S) are shown. Boxes with numbers indicate metabolites associated with this pathway identified by GCMS. Cyan (higher levels in Summer tissues), orange (higher in Kanlow tissues) and grey boxes (not significantly different).
S-adenosyl methionine, which is a substrate for COMT and CCoAOMT, was also depressed in Summer crowns and rhizomes. Signal intensity for actin (used as a loading control) was similar for the Kanlow extracts and two of the Summer extracts, and diminished in one Summer extract. However, all the other proteins probed were essentially identical to the levels detected in the two other Summer plants (Figure 4B). The consistent differences observed in the transcript abundances and protein levels were also mirrored in the activities of PAL, CAD and COMT which were significantly lower or not detected in extracts from Summer tissues as compared to extracts prepared from the Kanlow crowns and rhizomes. AscPx activities were greater in Summer plants as compared to the Kanlow plants (Figure 4C) in agreement with the transcript and protein data. These data suggested that transcription factors known to affect the phenylpropanoid pathway [43] could be similarly regulated in these plants at this harvest date. The expected relationships of the key transcription factors involved in secondary cell wall deposition and lignification, adapted from [43], is shown in Figure 4D. Individual regulatory genes are color coded in boxes and the cellular pathways they impact are shown in ovals. Arrows connect the putative relationships between these gene products in model systems and to their respective cellular processes (ovals, Figure 4D). Although MYB 26 has been depicted as a master regulator in Figure 4D, its overall role is not as well established as for the other regulatory genes [43]. The XND1 gene (pink box) is a negative regulator of xylogenesis, and inhibits programmed cell death (PCD) and related processes in Arabidopsis [44]. The best switchgrass orthologs to these individual transcription factors (color coded identically in Figure 4D and 4E) which positively regulate secondary cell growth and/or lignification exhibited greater ratios of transcript
Figure 4. Transcript, protein and enzyme activities confirm downregulation of phenylpropanoid pathway and upregulation of ascorbate peroxidase in Summer crowns and rhizomes. (A) Transcript abundances for all transcripts identified as belonging to a specific gene by Blast2GO. PAL, phenylalanine ammonia lyase; 4CL, 4-coumarate-CoA ligase; CAD, cinnamyl alcohol dehydrogenase; COMT, caffeoyl-O-methyltransferase; CCoAOMT, caffeoyl-CoA 3O-methyltransferase; SAMS, s-adenosylmethionine synthetase; AscPX, ascorbate peroxidase. Orange bars = Kanlow; Cyan bars = Summer. (B) Immunblots of crown and rhizome extracts separated by SDS-PAGE. Extracts from three biological replicates for each cultivar are shown. Other designations as described for panel A. (C) Enzyme activities of PAL, COMT, AscPx and CAD. Orange bars = Kanlow; Cyan bars = Summer. Other designations as described for panel A. (D) A map of potential relationships of key transcription factors that are known to play a major role in plant cell wall developmental processes. Each specific factor is shown in a different colored box. Figure is based on work described by Zhao and Dixon [43]. (E) Best switchgrass orthologs to transcription factors shown in (D). Boxes are colored identically to Panel D above. The switchgrass loci are identified and the FDR <0.05 values are shown in the last column. For the last column, boxes in orange color (transcripts significantly upregulated in Kanlow), cyan color (significantly upregulated in Summer), uncolored, not significant.
abundance (Kanlow/Summer; K/S). Most of these genes were significantly upregulated (FDR <0.05) in Kanlow crowns and rhizomes (highlighted in orange). In contrast, transcripts for the XND1 ortholog were significantly greater in Summer tissues (Figure 4E).

4.5 Discussion

In this study the dynamic aspects of the crown and rhizome transcriptome and metabolome in switchgrass has been explored for the first time. Analysis at this single harvest date revealed interesting insights into the metabolism of the crowns and rhizomes obtained from lowland Kanlow and upland Summer plants which differ in their photoperiod maturity response at the field site used for this experiment [3]. The work reported here complements several recent articles have expanded the available molecular resources for switchgrass and have begun the process of understanding switchgrass cellular processed within the context of its genomic complexity [16, 45-50]

The use of NGS coupled to bioinformatic analyses allowed approximately 80% of all reads to be effectively mapped to the switchgrass genome. Subsequent bioinformatic analyses indicated that these transcripts arose from 36,572 genes after filtering, which is ~56% of the total loci identified so far in the switchgrass genome, suggesting that the mapping and identification processes were relatively robust. It is likely that potentially interesting features could have been missed due to the mapping of reads to the unfinished contig-based switchgrass genome assembly. Thus, mapping of transcripts to incomplete genes, gene sequences with no discernible domains or near the end of contigs would have been excluded from our dataset. However, these data provided adequate depth to begin
an understanding of the metabolism of the crowns and rhizomes from the two contrasting switchgrass populations used in this study.

RNA-Seq of the transcriptomes provided a first broad-scale means to directly assess the metabolic status of these plants. PCA showed that the transcriptomes were clearly differentiated by the first component, indicating that between population (cultivar) differences were greater than within population differences. Both switchgrass cultivars are a heterogenous collection of genotypes [11] and the data for Summer plants would support this fact. The greater variation observed between the Summer plants could be indicative of differences in the timing of aerial senescence in these genotypes and its effect on crown and rhizome metabolism. In contrast, the transcriptomes of the Kanlow plants appeared to be more tightly clustered. Indeed the metabolomes also reflected this analysis. The two types of relatively non-biased large-scale approaches (RNA-Seq and GCMS) yielded similar conclusions with regard to the divergent metabolism in the Summer plants as compared to the Kanlow plants at this specific harvest date.

Environmental conditions prior to the harvest date indicated that the field site had received adequate rainfall and had not experienced freezing temperatures minimizing abiotic stresses on these plants. Although these data are supportive of the notion that the differences in metabolism observed in the crowns and rhizomes were more reflective of developmental events, it would be difficult to rule out biotic and abiotic stresses these plants may have experienced during growth and during sample handling.

Differential enrichment of transcripts, metabolites and gene-sets were observed between the two populations. Analysis of transcription factors such as WRKYs [51-53] which impact an extensive array of plant cellular processes was undertaken.
Approximately ~67% of the all WRKYs tentatively identified in the switchgrass genome were represented in the total transcriptomic datasets, indicative of their potential role in switchgrass crowns and rhizomes. Within this larger group, of the nine WRKY genes significantly upregulated in Kanlow plants, four genes appeared to be orthologs to Arabidopsis WRKYs with a role in the defense response to pathogens [54-56]. In contrast, the WRKY factor enriched in Summer tissues have been implicated as a negative regulator of leaf senescence [55]. Although the actual roles of individual switchgrass WRKYs remain to be explored, they could be reflective of other changes observed at the metabolite and GSEA levels. Summer crowns and rhizomes were enriched in some oxidized compounds and appeared to contained lower levels of α-tocopherol [57], and phenolic acids that could potentially function as antioxidants [58, 59]. Although these data show a link between the transcriptomic and metabolomic datasets, much work will be needed to unequivocally solve these interrelationships.

Transcriptomic and metabolomic data were consistent with the hypothesis that Kanlow tissues were actively utilizing carbon in contrast to Summer crowns and rhizomes. Sucrose and trehalose levels as well the GSEA results showing the enrichment of sucrose-starch metabolism pathways in Kanlow crowns were consistent with this hypothesis. Also other downstream biosynthetic pathways that rely on carbon skeletons derived from sugars were also enriched in Kanlow tissues. Most notably, there was a strong enrichment in pathways linked to cell wall accretion and growth occurring in Kanlow plants relative to the Summer plants. Data shown in Figure 4 effectively linked transcript and metabolite datasets to the underlying protein activities, at least for a portion of the lignin biosynthesis pathway which showed higher activity in Kanlow tissues.
Certain phenolic metabolites, namely cinnamic acid, ferulic acid and caffeic acid that are associated with monolignol synthesis [31, 60], were also more enriched in Kanlow consistent with enhanced flow of carbon through this pathway. In contrast, the amino acid precursors for this pathway, Phe and Tyr, were significantly enriched in Summer crowns and rhizomes which suggested that down-regulation of the phenylpropanoid pathway enzymes could have led to the increased accumulation of these amino acids. The significant up-regulation in transcript abundance for the switchgrass genes identified as potential orthologs of master transcription factors that regulate secondary cell wall biogenesis and lignification, based on [43], in Kanlow plants was striking and validated the other protein and metabolite findings. In contrast, transcripts for the putative switchgrass NAC gene XND1 were significantly greater in Summer tissues, suggestive of reduction in xylogenesis, and possibly in growth of tiller initials and rhizomes. In Arabidopsis, XND1 acts as a repressor of xylogenesis and programmed cell death (PCD) [44].

A number of KEGG pathways were found in Summer crowns and rhizomes that were enriched for transcripts, although not all enzymes were populated, unlike the more clear-cut observations on the phenylpropanoid pathway in Kanlow tissues. In trying to establish the possible connections between these diverse metabolic pathways, it became apparent that most possessed a potential common metabolite in acetyl-CoA (Figure 5). Potentially, scavenging of carbon compounds to form acetyl-CoA could represent a means to meet energy demands needed for maintaining cellular health in the possible absence of photosynthate delivery from the senescent shoots in Summer plants, which are consistent with results obtained for leafy spurge [61]. Acetyl-CoA serves as a key link
Figure 5. Acetyl-CoA appears to be a central hub connecting diverse pathways upregulated in Summer crowns and rhizomes. Cyan squares are KEGG pathways found to be upregulated in Summer plants by GSEA, and cyan circles are metabolites that were elevated in Summer plants relative to Kanlow plants. Edges connecting pathways (squares) to metabolites (circle) indicate that the given metabolite is found in the connected pathway. Pathways with direct connections in KEGG are also connected by edges. Acetyl-CoA (diamond) could be a potential linker molecule among these diverse pathways and is suggested to be a possible metabolic hub in Summer crowns and rhizomes entering dormancy.
between cellular metabolism, energetics and the acetylation of proteins in animal cells [62]. Similar mechanisms appear to be operating in plants [63-65], suggesting that a deeper understanding of the flux of carbon through a acetyl-CoA hub could provide new insights into the metabolism of crowns and rhizomes accompanying the seasonal growth habit of switchgrass and related perennial plants. Unfortunately, the metabolite analysis protocol used in this study cannot detect acetyl-CoA. Future use of other platforms such as LC-ms/ms might yield information on these more unstable metabolites.

4.6 Conclusions

Data presented in this manuscript showed that differences existed in the metabolite and transcript profiles in the crowns and rhizomes obtained from field-grown switchgrass plants belonging to two populations with divergent photoperiod responses. Crowns and rhizomes obtained from the more southern adapted cultivar Kanlow appeared to be in a “growth-mode” with enrichment in gene-sets associated with biosynthesis and accretion of tissues. In contrast, the northern adapted cultivar Summer appeared to have entered a more quiescent state with greater enrichment of transcripts and metabolites favoring channeling of carbon skeletons through an acetyl-CoA hub. Our findings provide an initial framework to understand differences in crown and rhizome metabolism which could be exploited to enhance the latitudinal adaptation of diverse switchgrass populations.

4.7 Acknowledgements

We thank Dr. Jean-Jack Reithoven and Mei Chen at the University of Nebraska-Lincoln Genomics Core Facility for preparing, sequencing and validating sequence data. We also thank Steve Masterson and Patrick Callahan for excellent technical support. The
4.8 Supporting Information

**Table S1 in File S1.** Mapping summary for switchgrass crown and rhizome samples analyzed by HTS

**Table S2 in File S1.** Annotation of all transcripts using Blast2GO

**Table S3 in File S1.** List of 124 KEGG Pathways populated by Blast2GO analysis of the switchgrass draft genome.

**Table S4 in File S1:** Gene Set Enrichment Analysis Results. KEGG pathways with an FDR <0.2 are shown. Transcript observed for individual enzymes in a given pathway are tabulated (enzymes).

**Figure S1 in File S1.** Temperature and rainfall at the field site prior to and post-harvest.

**Figure S2 in File S1.** Heat map of all major ions detected by GCMS using two-way hierarchical clustering. Yellow = high, Black = low. Data are for three independent extractions from three plants.

**File S2.** Excel file with list of differentially expressed genes (DEGs)
4.9 Works Cited


Chapter 5

Switchgrass (*Panicum virgatum* L.) flag leaf transcriptomes reveal molecular signatures of leaf development, senescence, and mineral dynamics

Note: The results described in this chapter have been published and all text is modified from the original version. Supplemental data can be found online at the journal’s website.


5.1 Abstract

Switchgrass flag leaves can be expected to be a source of carbon to the plant, and its senescence is likely to impact the remobilization of nutrients from the shoots to the rhizomes. However, many genes have not been assigned a function in specific stages of leaf development. Here we characterized gene expression in flag leaves over their development. By merging changes in leaf chlorophyll and the expression of genes for chlorophyll biosynthesis and degradation, a four phase molecular roadmap for switchgrass flag leaf ontogeny was developed. Genes associated with early leaf development were up-regulated in phase 1. Phase 2 leaves had increased expression of genes for chlorophyll biosynthesis, and those needed for full leaf function. Phase 3 coincided with the most active phase for leaf C and N assimilation. Phase 4 was associated with the onset of senescence, as observed by declining leaf chlorophyll content, a significant up-regulation in transcripts coding for enzymes involved with chlorophyll degradation, and in a large number of senescence-associated genes. Of considerable interest were switchgrass NAC transcription factors with significantly higher expression in senescing flag leaves. Two of these transcription factors were closely related to a wheat NAC gene that impacts mineral remobilization. The third switchgrass NAC factor was orthologous to an Arabidopsis gene with a known role in leaf senescence. Other genes coding for nitrogen and mineral utilization, including ureide, ammonium, nitrate, and molybdenum transporters, shared expression profiles that were significantly co-regulated with the expression profiles of the three NAC transcription factors. These data provide a good starting point to link shoot senescence to the onset of dormancy in field-grown switchgrass.
5.2 Introduction

Switchgrass (*Panicum virgatum* L.) is a perennial C₄ species that is likely to be grown as a source of biomass for the biofuel sector [1]. Perenniality resides in the below-ground rhizome and crown tissues that are the primary sources of tiller buds. Dormant tiller buds elongate in spring to produce above-ground biomass. Each tiller consists of several phytomers, each comprised of nodes, internodes and leaves [2], and most tillers will become reproductive [3]. The last leaf to be produced on a flowering tiller is the flag leaf, which subtends the inflorescence. Leaves are the dominant source of photosynthates. During a growing season leaves go through several stages of development, with dynamic physiological and metabolic activities being reflected by underlying gene expression. An increased understanding of the coordination of gene expression and leaf development could lead to insights into the biology of switchgrass plants, and lead to new breeding targets for crop improvement.

Flowering and seed production are generally accompanied by tiller senescence and associated with the onset of dormancy in the rhizomes. It is likely that flag leaves will supply nutrients to both the seeds and the rhizomes (sinks). Initiation of senescence in flag leaves could provide signals that trigger remobilization of nutrients from the shoots, and the initiation of dormancy mechanisms in the rhizomes, but these signals are unknown. Switchgrass can remobilize substantial amounts of N and other nutrients to the crowns and rhizomes, especially towards the end of a growing season [4-6]. Improving remobilization of N and minerals from the above-ground biomass to the below-ground tissues is important to switchgrass breeding programs, since biomass with lower N and minerals significantly improves sustainability of production and conversion of biomass to
fuels in thermochemical platforms [7, 8]. We currently lack an understanding of the genes involved in regulating and carrying out nutrient remobilization during senescence, and improved knowledge of genes preferentially expressed in senescing tissues during leaf senescence could lead to new insights.

In cereal crops, flag leaves provide a large portion of the nutrients needed for grain fill [9], and different aspects of flag leaf senescence have been studied in several major cereal crops (for example [10-12]. In a large scale transcriptomic analysis of the genes involved in maize internode and leaf senescence [13], a large number of pathways were common or different between these two tissues. Commonalities and differences in the progression of leaf senescence programs between maize and Arabidopsis [13] were uncovered by using a database of plant senescence-associated genes (SAGs) [14]. While these studies and others in the model dicot Arabidopsis [15, 16] have been productive in developing a framework to understand leaf senescence, they may not be reflective of all plant species, since differences exist between plant species, and especially between annuals and perennials [17-19].

In this study, we evaluated molecular changes accompanying leaf development and senescence through global transcriptional profiling in field grown switchgrass plants. By computational analysis of expression of over 40,000 transcripts over developmental time, a number of key switchgrass genes and networks that impact leaf growth, development, senescence and remobilization of nutrients have been found. This work could provide needed information about potential mechanisms that could integrate changes in flag leaf metabolism to the onset of dormancy in the rhizomes, and provide molecular targets for phenotyping plants with improved traits [20].
5.3 Experimental Methods

5.3.1 Field layout, replication, and sample collection

Field plantings of cv Summer switchgrass were established in small (1 m x 1.2 m) plots using seedlings raised in a green house in June 2009 at the experimental farms of the University of Nebraska, near Mead, NE. Twelve plants were planted in each plot to mimic sward conditions at maturity. Switchgrass fields were not-irrigated and managed as described elsewhere [21]. There were thirty replicated small plots in this field. Flag leaves were harvested from plants during the 2012 growing season, at which time, plants were fully established, but individual plants could still be identified within each small plot. Plots and plants were randomly selected to capture maximum diversity at each harvest date. Three biological replicate pools of flag leaves harvested from ten separate plants each were collected at five different harvest dates (a total of 15 samples from ~100 individual plants) corresponding to different developmental states of the plant: July 3 (> 90 % headed), July 27 (>80 % anthesis), August 16 (>90 % seed set), August 31 (> 70 % mature seed), and September 19 (> 90 % senescence onset). Collected tissues were immediately flash-frozen with liquid nitrogen and stored at -80 C. Leaf samples were later ground in a cryogenic grinder [22] prior to analysis.

5.3.2 Chlorophyll quantification

Chlorophyll was extracted in 80% acetone in water and quantified using the method described by Porra et al [23].

5.3.3 RNA extraction

Total RNA was extracted from 100 mg of frozen plant tissue using TRIzol reagent following manufacturer protocol (Invitrogen, Carlsbad, CA, USA). RNA was
cleaned up and residual DNA was removed using the RNeasy MinElute Cleanup Kit (Qiagen, Valencia, CA, USA).

5.3.4 RNA sequencing

One ug of total RNA was reverse transcribed and converted to sequencing libraries using the TruSeq RNAseq Library kit per manufacturer’s suggestion (Illumina Inc, San Diego, CA). Utilizing unique indexes from the library kit, individual samples were diluted and to a concentration of 10 nM and multiplexed at five samples per lane. Single read 100 bp sequencing was performed on the Illumina HiSeq2000 system. Following sequencing, fastq files were used for mapping.

5.3.5 Mapping and expression counting

HiSeq2000 100bp reads (an average of 45 million reads / sample) were mapped to the switchgrass draft genome (Pvi0, www.phytozome.org [24]) using Bowtie2 [25] with “sensitive-local” settings. On average, 93% of the reads mapped to the draft genome with 78% of the reads mapping to annotated gene regions. Gene counts were calculated using the primary transcript annotation file released with Pvi0 and the program featureCounts [26].

5.3.6 Differential expression analysis

Prior to differential expression testing, genes that did not have an average expression level of 50 counts in at least one time point were removed from the dataset. This was an arbitrary cut-off, approximately equivalent to 1 read per million, and used to minimize the potential for overestimating differences between harvest dates arising from genes with low expression. Differential expression analysis was done using DESeq2 [27] in R [28]. Pairwise comparisons of all five timepoints (10 total comparisons) were
carried out to generate a list of differentially expressed genes (DEGs) for the entire
dataset using an FDR cutoff of 0.05.

5.3.7 Gene set creation

Metabolic pathways for the switchgrass draft genome were built based on existing
pathways found in the Kyote Encyclopedia of Genes and Genomes (KEGG) [29, 30].
Sorghum proteins annotated at each node in pathways of interest were used to find the
best switchgrass matches using blastp [31]. Genes involved in histone (de)methylation
and (de)acetylation were identified following the approach used by Cigliano et al., [32]
using already identified chromatin modification genes annotated in ChromDB
(www.chromdb.org). Nutrient transporters in the switchgrass draft genome were
identified as previously described [33].

Transcription factors were identified based on PFAM annotations included with
the draft genome release and the rules for classification of Transcription Factor Families
as detailed in the Plant Transcription Factor Database v3.0 (plntfdb.bio.uni-
potsdam.de/v3.0) [34].

Previously identified senescence associated genes (SAGs) from Arabidopsis [14, 16] and rice [35] were used to identify putative SAGs in switchgrass. Orthologs were
identified in the switchgrass draft genome by matching the closest Arabidopsis and rice
genes included with the genome annotation with the above SAG lists. The switchgrass
orthologs were then filtered for genes differentially expressed with peak expression in the
4th or 5th collection time point to yield a set of putative switchgrass SAGs.

5.3.8 Pathway Studios analysis
Log2 transformed expression data was used in the program Pathway Studio 9.0 [36], to identify over-represented ontologies. To facilitate this analysis, switchgrass sequences were mapped to the most similar Arabidopsis gene using BlastX, and the resulting gene function associations were used for gene set and sub-network analysis using default settings.

5.3.9 Statistical analysis

Principal component analysis (PCA) was done using the “prcomp “ function in R [28]. Heatmaps were created by hierarchical clustering in JMP® Version 9.0 (SAS Institute Inc, Cary, NC, 1989-2007) using Ward’s method on standardized gene expression values.

5.4 Results

5.4.1 Changes in gene expression define developmental stages

On average approximately 45.6 M reads were obtained for each sample, of which almost 93 % could be mapped to the switchgrass genome version 0.0 available at www.phytozome.org [24] and over 78 % were mapped to specific genes (Table S1). An analysis of all unique transcripts with average normalized read counts >50 in at least one harvest date are shown (Fig.1a). Of these genes, 27,214 were common to all datasets, and only a small fraction (<0.04 %) were uniquely associated with a specific harvest date. Numbers of shared genes were fewer between more distant harvest dates (Fig.1a), suggesting that underlying changes in leaf physiology were responsible for the changing expression profiles.

We next asked whether transcriptomes at each harvest date could be used as molecular signatures of the stage of development. At each harvest date, the
Figure 1. Gene expression profiles differentiate flag leaf transcriptomes. (a) Venn diagram of gene expression at different harvest dates. (b) PCA of transcriptomic dataset.
transcriptomes were effectively separated by PCA (Fig. 1b), suggesting that these differences arose from underlying developmental status, and to some extent from environmental conditions associated with the time of harvest (supplementary Fig. S1). The first component accounted for 27.9% of the variation and separated the transcriptomes from the anthesis (7/27), early seed fill (8/16), and physiological maturity (9/19) harvest dates. PC2 accounted for 19.2% of the variation and separated the transcriptomes present in flag leaves harvested at the heading (7/3) and late seed fill stage (8/31) of plant development. Within each harvest date, greatest variation between the biological replicates was observed in the transcriptomes for the earliest (7/3) and last (9/19) harvest, plausibly attributable to variation in plant development within switchgrass populations [37].

Chlorophyll levels in switchgrass flag leaves were low at the first harvest, increased to a maximum near the early seed-set stage, and declined thereafter, with the lowest chlorophyll levels at the last harvest date (Fig. 2a). In general, transcript abundance for the six gene families coding for proteins involved in chlorophyll biosynthesis were highest in flag leaves harvested at anthesis (7/27) and least abundant in senescing leaves harvested in September (9/19). In contrast, three gene families coding for switchgrass protein orthologs involved in chlorophyll catabolism were strongly up-regulated at the last harvest date (Fig. 2b). The loss of chlorophylls in flag leaves harvested at this date combined with the gene expression data indicated that senescence had been initiated in these plants sometime after seed filling had started.

Combining the information shown in Figs. 2a and b, we developed a road-map of switchgrass flag leaf development to assist in interpreting the large transcriptomic
Figure 2. Road map of switchgrass flag leaf development. (a) Total leaf chlorophyll content. (b) Expression profiles of genes coding for proteins involved in chlorophyll biosynthesis and degradation. Overview of chlorophyll biosynthesis (blue box) and chlorophyll degradation (red box). Z score heatmap of relative total expression. Cyan is low expression, magenta is high expression. (c) Road map of flag leaf development categorized in four phases based on data shown in panels a and b. Blue line = chlorophyll biosynthesis; brown line = chlorophyll degradation. The identities of each gene family shown in this figure are given in supplementary Excel file 1.
datasets. Essentially, phase 1 (light green bar, Fig. 2c) was associated with leaf expansion. Phase 2 (dark green bar, Fig. 2c) was associated with a substantial up-regulation of the genes associated with chlorophyll biosynthesis (blue line, Fig. 2c), likely coinciding with the maturation of the flag leaves from sinks to sources. Phase 3 (yellow bar, Fig. 2c) was associated with a decline in transcript abundances for the genes associated with chlorophyll biosynthesis and a small increase in the transcripts of genes associated with chlorophyll degradation. Phase 4 (brown bar, Fig. 2c) was associated with a large increase in transcripts coding for proteins involved in chlorophyll degradation, consistent with the onset of leaf senescence.

### 5.4.2 C and N assimilation genes are substantially up-regulated in phase 2

Expression of genes coding for proteins associated with the C4 pathway and CO2 and N assimilation were analyzed to determine if they followed the phases outlined in Fig. 2c. Mapped reads for individual genes within a gene family annotated as coding for the same protein were tabulated, and counts for all the genes within a given gene family were totaled and used to generate color heat maps (Fig. 3). These data are separate from the differentially expressed gene (DEG) analyses discussed later in the text.

Genes linked to C4 photosynthesis were all highly expressed at phase 2 (Fig. 3a), except for phosphoenolpyruvate carboxylase kinase genes (PEPCK) which codes for a kinase that modulates phosphoenolpyruvate carboxylase (PEPC) activities through phosphorylation [38]. However, all of these genes were downregulated in leaves at successive harvests. This pattern of C4-related gene expression in switchgrass flag leaves suggests that phase 3 is one where leaves are fully mature and functioning efficiently as sources for fixed C. Genes involved in the C assimilatory pathway
Figure 3. Changes in the expression profiles of genes coding for enzymes involved in key functions in switchgrass flag leaves. (a) C4 pathway. (b) Calvin cycle. (c) Nitrogen assimilation. (d) Ascorbate and glutathione metabolism. Cyan is low expression, magenta is high expression. The identities of each gene family shown in this figure are given in supplementary Excel file 1.
(Fig. 3b) and in N metabolism (Fig. 3c) appeared to be regulated similarly to genes associated with CO₂ assimilation, consistent with the developmental patterns suggested above. An exception was NADH-DEPENDENT GLUTAMATE DEHYDROGENASE (GDHb, Fig. 3C) for which transcript abundance was highest in senescing flag leaves.

5.4.3 Genes associated with ascorbate and glutathione metabolism are differentially regulated

Expression of genes involved with ascorbate biosynthesis [39] was generally greatest during phases 2 and 3 (Fig. 3d), and appears to be linked with the period of growth and function as a carbon source. There was, however, an increase in the abundance of transcripts coding ascorbate peroxidases (AscPrx) at later harvests. In contrast, transcripts coding for monodehydroascorbate reductase (mDHAR) had a bimodal expression profile coincident with phase 2 and phase 4 stages. Transcripts for genes coding for dehydroascorbate reductase (DHAR) were relatively constant over the first 4 harvests, before declining in senescing leaves.

Glutathione is another important redox intermediate in cells and serves multiple roles during plant development [40]. Transcripts for genes coding for glutathione biosynthesis were up-regulated at the phase 2 and 4 (Fig. 3d). Genes coding for enzymes related to glutathione catabolism were most abundantly expressed during phase 4 (Fig. 3d).

5.4.4 Gene set enrichment analysis identifies possible regulatory factors controlling leaf development

Gene set enrichment analysis was performed sequentially for all pair-wise comparisons (supplemental Table S2). Physiological processes involved in flavonoid
biosynthesis (Kaempferol glucoside biosynthesis and Quercetin glucoside biosynthesis), cytokinin biosynthesis, and sequence-specific DNA binding transcription factor activity were preferentially expressed in the early developmental stages, while oxidative stress-associated ontologies (heme binding, oxygen binding, and monooxygenase activity) were preferentially expressed at later developmental stages. Several ontologies associated with the earlier time points (heme binding, oxygen binding, and monooxygenase activity) showed a reversal of expression patterns when the late August time point was compared to the final harvest supporting data shown in Figures 2 and 3.

Sub-network analysis generally indicated an increase in jasmonic acid signaling during senescence. There was also some indication that targets of auxin transport had a higher expression prior to senescence. Genes for proteins that regulate phosphate import were preferentially expressed during senescence, suggesting that remobilization of phosphate was occurring (see Supplementary file 2).

Among the more interesting associations observed in the sub-network analysis was a strong differential expression of targets for miRNA156 (down-regulated during senescence) and miRNA164 (up-regulated during senescence as early as late August) (see Supplementary file 2).

5.4.5 Clustering of differentially expressed genes support developmental timeline of flag leaves

A 2-way hierarchical clustering of all differentially-expressed genes (DEGs) from the entire transcriptome dataset with an FDR <0.05 is shown in Fig. 4. The three biological replicates at each harvest date were quite similar in gene expression profiles and clustered together, consistent with the overall PCA analyses. The samples from the
Figure 4. Two-way clustering of differentially expressed genes (DEGs) across all harvested samples. At each harvest date, three biological replicates were analyzed. The bars on the left show the phases that were categorized in Fig. 2c. Seven major clusters labeled C1–C7 are so indicated. The numbers of DEGs within each cluster are in brackets. Cyan is low expression, majenta is high expression. The identity of each gene is provided in supplementary Excel file 1.
second sampling date form an outlier in both the PCA and differential hierarchal clustering analysis. It is noteworthy that this sampling date occurred shortly after the only significant rainfall of the season (Figure S1). Indeed, many of the DEGs in this sample were indicative of a relaxed drought stress response, including higher expression of genes with ontologies associated with photosynthesis and carbon metabolism, along with lower expression of ABA-associated genes (Table S2).

Seven major clusters of significantly up-regulated genes were found across all harvest dates. Of these, one cluster of ~1,619 genes (cluster 4) had a distinct bimodal expression profile. These genes were specifically up-regulated in flag leaves at anthesis (7/27) and at physiological maturity (9/19). However, it is impossible to determine if these differences were related specifically to the developmental state of these plants, or to the precipitation event just prior to harvest (Fig. S1). Many of the genes associated with clusters 1 and 3 code for proteins involved with growth and biosynthetic processes and have been described earlier (phases 1 and 2; Figures 2 and 3). Cluster 2 genes (phase 3) were significantly up-regulated at the third harvest time, and appear to have been in response to environmental conditions. There was no major precipitation event between the second and third harvest dates (Fig. S1), and this cluster contained several heat shock protein orthologs, suggesting that leaves were stressed. The remaining clusters were associated with senescence (clusters 4, 5, 6, 7, phase 4).

We next undertook a deeper analysis of DEGs that could be ascribed to specific functions. Expression heat maps of genes coding for glutathione-S-transferases, transcription factors, transporters, and proteases that belong to large multigenic families are shown in supplementary Figs. S2-S5. Expression profiles of several other classes of
genes including the senescence-associated genes (SAGs) are discussed below.

5.4.6 Genes involved in epigenetic regulation were differentially expressed between sampling dates

Reversible histone modification is a critical component of the epigenetic control of cellular process, and plant genomes contain an extensive set of genes that code for these enzymes. Well characterized histone marks include acetylation and methylation among many others [41]. Several genes coding for proteins classified as histone acetyltransferases (HAT, HAC and HAG - Figure 5A) [32] were found to be differentially expressed, however their relative transcript abundances were variable, and in general low. In contrast, the genes categorized as coding for histone deacetylases (HDA; SRT; Fig. 5A) were generally abundant during phase 2 and during leaf senescence (phase 4). One paralog, HDA8 (Pavirv00041899) was abundantly expressed in senescent leaves (Fig. 5A).

Many arginine-methyl transferases (PRMT) and SET-domain group lysine methyltransferase genes were found to be differentially expressed (Fig. 5B). Interestingly, four histone demethylases belonging to the Jumonji-family of demethylases [42], (JMJ – Fig. 5B), with high transcript abundance were expressed at early or late periods of flag leaf development. Several genes paralogous to the ones shown in Fig. 5 had high abundance across all harvest dates (Table S3), suggesting that these genes were needed to maintain proper cellular functions. The relative contributions of individual genes to transcriptional events at specific loci are not known.
Figure 5. Clustering of DEGs associated with histone modification. (a) Switchgrass genes coding for enzymes involved in histone acetylation and deacetylation. (b) Switchgrass genes coding for enzymes involved in histone methylation and demethylation. The relative expression value (number of mapped reads within each class across all time points) is shown as a single bar on the left (yellow is low counts, red is high counts). The overall expression across harvest dates are shown as colored rectangles. Cyan is no/low expression, magenta is high expression. The identity of each gene is provided in supplementary Excel file 1.
5.4.7 Transport processes are enhanced during senescence

A large number of mineral transporters were differentially expressed during switchgrass flag leaf development (Fig. 6; S4). In general, two distinct profiles of expression were observed; genes that were most abundant during the period of leaf expansion and transition to a source (phase 2), and genes that were most abundant during senescence (phase 4). Since both expansion and senescence stages of leaf development are likely to require transporters to mobilize minerals into (phase 2) or out of the leaf (phase 4), such a profile of transporter gene expression was anticipated.

Transcripts for a Si transporter was most abundant during active leaf metabolism, and decreased when leaves transitioned into senescence (LSi1, Fig. 6a), in a manner similar to that observed for a switchgrass crown and rhizome dataset [33]. Putative genes coding for ammonia (AMT) and nitrate (NRT) transporters with highest overall total transcript abundance were up-regulated during phase 2 in flag leaves (AMT1, NRT1, 2 and 3 Fig. 6a). However AMT3 and 4, and NRT 4, 5, 6, 10, and 12 were up-regulated during phase 4. Transcripts for NRTs1, 2, 3, and 13 were high during phase 2, whereas, transcripts for NRT12 were most abundant in senescing leaves (Fig. 6a).

Transcripts for nine genes encoding switchgrass potassium transporters (Palmer et al. 2014; labeled KUP; Fig. 6a) were present in the DEGs. Of these, three were most abundantly expressed during phase 2 of flag leaf development. Reads for the other six genes were most elevated in senescing leaves. Similar patterns were observed for the sulfur transporters (SULTR, Fig. 6a).

Four genes classified as sucrose transporters (SUT, Fig. 6a) were part of the DEGs and were highly expressed during phase 2 of switchgrass flag leaf development.
Figure 6. Two-way clustering of DEGs belonging to different nutrient transporter classes across harvest dates. (a) Silicon, ammonia, nitrate, potassium, sulfate, and sugar transporters. (b) Molybdenum, phosphate, and zinc/copper/iron transporters. Other details are as described for Fig. 5.
Expression profiles of the *SUTs* was consistent with the substantial up-regulation observed for a number of other biosynthetic genes regulating photosynthesis, and carbon and nitrogen assimilation observed during phase 2 as has been described earlier (see Figs. 2-4). Three *MOT* genes (molybendum transporters) [33] were most abundantly expressed during phase 4 (Fig. 6b).

We also analyzed the putative phosphate transporters (PHO and PHT) and zinc/divalent metal transporters (ZIPs) [43, 44] (Fig. 6b). Four *PHOs* were most abundant during phase 1, and the gene *PHO3*, which was most abundant, was highly expressed at phase 2 of flag leaf development. Inorganic phosphate requirements are expected to be high during leaf maturation and subsequent transition into a source leaf. Most of the *PHT* genes with high expression (red/orange boxes; Figure 6b) were most abundant during phase 2 or phase 3 of flag leaf development where leaf biosynthetic activities appear to peak. Several other genes with lower expression (yellow; Fig. 6b) were significantly more abundant in senescing leaves, suggesting a role in phosphate remobilization (also see GSEA data, TableS2). Eight ZIP genes were found to be differently expressed during switchgrass flag leaf development (Fig. 6B).

5.4.8 **Senescence and growth processes are differential through seasonal development**

Leaf senescence is accompanied by the increase in transcript abundances for a host of genes that are referred to as senescence-associated genes or *SAGs* [14, 45]. A large number of putative switchgrass *SAGs* were significantly up-regulated at the last harvest date (Fig. 7), although smaller clusters were more abundantly expressed at earlier time points. As an example, transcripts of switchgrass *AUXIN RESPONSE FACTOR 2*
Figure 7. Clustering of differentially expressed switchgrass SAGs. Cyan is low expression, majenta is high expression. The identities of each gene family shown in this figure are given in supplementary Excel file 1.
(ARF2) [46] orthologs *SAGs* 29, 47 and 55 (Fig. 7) were higher at the third harvest and either declined (*SAG*47 and 55) or increased with time (*SAG* 29; Fig. 7). A small cohort of genes provisionally identified as *SAGs* were first up-regulated in flag leaves harvested at the early seed development stage (8/16; Fig. 7), which coincided with maximal leaf chlorophyll content (see Fig. 2).

5.4.9 Organic acids: the carbon currency from the breakdown of lipids during senescence

Breakdown of fatty acids via the β-oxidation pathway is one of the important processes that occurs in senescent tissues to recapture carbon [47, 48]. Genes associated with the β-oxidation and related pathways were significantly up-regulated during phase 4 (Fig. 8). We found 230 genes annotated as lipases that were differentially expressed, of which 88 had maximal expression during phase 4 (Fig. S6).

In contrast to Arabidopsis, transcripts for all of the genes needed for a functional glyoxylate cycle were strongly up-regulated in senescing switchgrass leaves (Fig. 8, orange). The two key glyoxylate cycle enzymes, isocitrate lyase (ICL) and malate synthase (MS) [47, 49, 50], were found in greatest abundance in senescing (phase 4) flag leaves (Fig. 8). Conversion of fatty acids into organic acids requires reactions in glyoxysomes, cytoplasm, and mitochondria (Fig. 8). To a large extent, transcripts coding for all of the key enzymes localized in the different cellular compartments were significantly more abundant during phase 4 (Fig. 8).

5.4.10 Roles for switchgrass NAC transcription factors in remobilization

DEG analysis of several different classes of transcription factors are shown in Fig. S3 and for all the NAC genes in Fig. S7. NAC transcription factors are members of
Figure 8. Lipids are targeted to the β-oxidation pathway in senescing flag leaves. A simplified pathway linking the different compartments involved in this process is shown. Expression levels of each enzyme (1–11) identified in the pathways are shown as a color map. Cyan is low expression, magenta is high expression. Identity of each ortholog is provided in supplementary Excel file 1.
a large family of related genes and proteins involved in virtually every aspect of plant development [51]. We found a total of 78 NACs in the DEGs (Fig. S7), of which 26 were most abundantly expressed at phase 1, six at phase 2, two at phase 3, and 39 at phase 4.

Using the wheat NAM-B1 protein sequence [52], orthologous proteins were identified in switchgrass, sorghum, foxtail millet, maize, Brachypodium, rice, and Arabidopsis (Fig. 9a). The amino acid sequence near the C-terminus within the red box (Fig. 9a) distinguishes NAM-B1 from other related NACs. Close orthologs were identified in all the grasses but not in Arabidopsis. For switchgrass, only two NACs contained this specific domain (Pavirv00068192 and Pavirv00065253), however a third closely related NAC did not have this specific sequence of amino acids (Pavirv00018572). A dendrogram of all these protein sequences (Fig. 9b) suggests that the Arabidopsis NAC025 and Os07g37920 were more related to each other, compared to the other grass sequences analyzed.

All three switchgrass NACs described above (Fig. 9b) had an expression profile that linked them to the senescence process (Fig. 9c) akin to what has been demonstrated for wheat NAM-B1[52]. Transcriptional profiles of some select genes co-expressed with the NACs described above (with an expression correlation of >0.9) included the Mo transporters, several hydrolases, and other genes coding for NACs, autophagy, and catalase 2 proteins (Fig. 9c).

5.5 Discussion

To determine whether gene expression profiles in flag leaves collected at different times throughout development corresponded with known markers of leaf development,
Figure 9. Identification of switchgrass NACs orthologous to *Triticum dicoccoides* NAM B1 and profiles of co-expressed genes. (a) The wheat NAM-B1 conserved domain is present in two switchgrass NAC genes (red box). (b) A dendrogram of the sequences used in this study. (c) Expression profiles of switchgrass genes with high (>0.9) correlation to the expression profiles of potential remobilization NACs shown in panel (a). Gene identities are provided in supplementary Excel file 1.
we tracked chlorophyll content as a known marker of leaf expansion and senescence onset [18, 53]. Chlorophyll content in flag leaves increased from the time of heading to the onset of seed set and declined thereafter, similar to patterns reported for flag leaves in cereal crops [10, 35, 54]. Chlorophyll biosynthesis and degradation are catalyzed by a unique set of enzymes [53, 55], and the first molecular signature for the initiation of leaf senescence is the onset of expression of genes coding for chlorophyll breakdown [53].

Consistent with these prior observations, transcript abundances were high for the genes coding for chlorophyll biosynthesis during the early part of switchgrass flag leaf development and declined thereafter; whereas, transcripts for genes coding for enzymes involved in chlorophyll degradation were significantly greater with the onset of leaf senescence. Overall, data on the transcript abundances for genes linked to C, N and redox balance were consistent with published work on other plant species such as maize and rice [13, 35, 56] and the suggested roadmap for switchgrass flag leaf development (see Fig. 2).

Pathway analyses also supported these overall findings, and indicated that a projected loss of cytokinin synthesis with the cessation of flag leaf growth, and the up-regulation of targets of the transcription factor TT2 during the onset of senescence, were as expected [57, 58]. A possible role for miRNA156 was also discerned through pathway analyses. MiRNA156 is a well-known positive regulator of flowering [59], and has also been implicated in regulation of senescence through down-regulation of ORESARA 1 [60]. These findings were consistent with the floral development activity during phase 2 and senescence during phase 4 described in Fig. 2. As small RNAs, these signals have the additional possibility of being systemic regulators of development [61] and might
participate in the circuits that link shoot senescence to the onset of rhizome dormancy towards the end of the growing season [8].

Although the early phases of flag leaf development are important to the efficient functioning of the leaf as a source of fixed carbon, the timing and onset of senescence are of considerable basic and applied interest [8]. In switchgrass, flag leaves appear to be functional after seeds have attained physiological maturity (this work), and thus could continue to serve as a source of photosynthate to the rhizomes towards the end of a growing season. Furthermore, it is likely that flag leaf senescence is intimately linked to the remobilization of nutrients from the shoots to the rhizomes contributing to the perenniality of the plant [8].

Among the host of metabolic changes that occur during the onset of leaf senescence are increase in ROS, the suppression of transcription of genes, and the remobilization of nutrients to sinks [18]. Increase or decrease in transcripts of genes in these diverse pathways were discerned in switchgrass flag leaves, especially as they entered phase 4 (see Fig. 2) of development. As examples, genes coding for enzymes involved in C and N assimilation and ascorbate biosynthesis were down-regulated, and other genes implicated in alleviating oxidative stress and substrate level generation of NAD(P)H, including several SAGs were up-regulated.

Lowered levels of ascorbate and higher ROS have been associated with the progression of leaf senescence [62-64], and increased abundance of isocitrate dehydrogenase (ICDH; Fig. 3) has been suggested to reduce oxidative stress by regenerating NADPH [65]. Substrate level regeneration of NADPH might be important in senescing switchgrass leaves when mitochondrial generation of reducing equivalents
could become compromised. Several switchgrass SAGs with potential roles during leaf senescence were also identified. Among these, \textit{SAG14-HEXOKINASE-1} and \textit{SAG28-ATAFH} are known to be involved in different aspects of cellular signaling involving sugars [66, 67] and signaling involving reactive oxygen species (ROS) [68] in other plants. \textit{SAG 18}, which codes for a cytosolic NADP-malic enzyme (NADP-ME), was also present in this cohort. NADP-ME is expected to provide NADPH and pyruvate in the cytoplasm. The expression profiles for the \textit{NADP-ME (SAG 2 and 18)} mirrored those of the \textit{HEXOKINASE1 SAG14} (Fig.7) suggesting that the transcriptional co-expression profiles of these genes could be linked to changes in the sugar, nitrogen, and redox metabolism expected with the onset and progression of leaf senescence [18]. Sugar sensing has been proposed to be coupled to ROS through an interplay between the cytoplasm and mitochondria [69], which in turn is linked to the production of ascorbate and ascorbate utilizing enzymes such as ascorbate peroxidases and other redox processes [69, 70]. \textit{SAGs 57 and 63}, which code for cytosolic ascorbate peroxidases were significantly up-regulated during phase 4.

There were significant increases in transcript abundance for glutamate dehydrogenase (GDH, Fig. 3) over time. In Arabidopsis, GDH isoforms are largely localized in the mitochondria present in phloem companion cells in leaves and roots [71]. They are also important to the catabolism and remobilization of amino acids during seed germination and senescence and are expected to be involved in signaling associated with C and N metabolism in plants [72-74]. We also observed a significant up-regulation of genes involved in the remobilization of lipids with the onset of senescence primarily through the action of lipases and β-oxidation. It is known that lipids are lost from
senescing switchgrass leaves [48] and up-regulation of genes coding for enzymes in the β-oxidation and a large number of lipases occurs in senescing Arabidopsis leaves [47]. However, gluconeogenesis did not appear to be occurring during leaf senescence in switchgrass, since transcripts for a key enzyme (PEPCK; Fig. 3) were low in senescing leaves and highest during the phase 1 and 2 of leaf development, in accordance to cucumber cotyledons and barley leaves [49]. Similar changes in the profiles for lipid degradation pathways have been reported in barley and sweet potato leaves [49, 50]. Our data also suggest that much of the organic acids are used as skeletons to synthesize amino acids (see Fig. 8), chiefly ASN, GLN and possibly GLU, for eventual export to sinks, probably the rhizomes.

Epigenetic control of plant development is well established [41], and transcripts for a large number of switchgrass genes coding for these histone modifying enzymes were differentially regulated during flag leaf development. As examples, genes encoding switchgrass Jumonji-type demethylases JMJ5 and JMJ6 (Fig. 5) were most highly expressed in senescing leaves. Switchgrass JMJ6 is orthologous to the Arabidopsis PKDM7B (AT4G20400) gene that codes for an enzyme that demethylates histone H3 lysine 4 and suppresses gene transcription [75].

Mineral acquisition, transport and remobilization is an integral part of plant development, and senescence often serves as a cue for redistribution of some, but not all, minerals from senescing tissues to sinks. Mineral concentration data from switchgrass crowns and rhizomes [33] and aerial biomass [5] harvested over the course of a growing season would indicate that there is remobilization of several minerals from the shoots to the rhizomes. Several genes coding for different classes of switchgrass mineral nutrient
transporters were significantly overexpressed during the phase 2 (period of active growth) or during phase 4 (onset of senescence). These included transporters for ammonia (AMT), K (KUP), S (SULTR), P (PHO) and Zn (ZIP) (Fig. 6). These nutrients are needed during active growth and can be remobilized during senescence, [44, 76-78], suggesting that these specific switchgrass genes are integral for efficient mineral mobilization.

There was a significant up-regulation in putative MOTs with the onset of leaf senescence. The switchgrass genome appears to code for at least 5 MOTs [33]. In plants, molybdenum is assimilated into a pterin to form the MoCo factor, which is present in nitrate reductase (NR), xanthine dehydrogenase (XDH), sulfite oxidase (SO), and aldehyde oxidase (AO) [79]. How Mo is remobilized during leaf senescence is not known. However, XDH appears to be important for scavenging purines into ureides during leaf senescence [80]. Of four switchgrass genes predicted to code for XDH (Pavirv00070280; Pavirv00060710; Pavirv00032623; Pavirv00030276), the first three exhibit increased transcript abundance during leaf senescence, in concert with the transcripts for three putative switchgrass ureide-transporters (Pavirv00005720; Pavirv00055942; Pavirv00023499), suggestive of a role in moving ureides (N and C) from senescing tissues to the crowns and rhizomes, as has been described for Arabidopsis [80]. Changes in the flux of C and N metabolites could link aerial senescence to the onset of dormancy in the crowns and rhizomes, since these metabolites can serve as signals integrating physiological processes [74, 81]. In Arabidopsis there are two MOTS, and MOT2 protein is involved in molybdenum transport and allocation within the plant.
Arabidopsis *MOT2* transcripts have been documented to be elevated in senescent tissues [82].

Plant developmental processes are controlled through the concerted action of transcription factors, and a number of these factors involved in senescence and affecting the remobilization of nutrients to sink tissues have been described [18, 83]. Although a large number of switchgrass genes coding for transcription factor families were identified and shown to be divergently expressed (Fig. S3; Fig. 9), there were three NAC genes that were up-regulated close to the onset of senescence. Two of these three NACs were orthologous to the Wheat NAM-B1 shown to impact nutrient remobilization [52] suggesting a similar role in switchgrass. The third switchgrass NAC which lacks the C-terminal NAM-B1 signature domain (Fig. 9) appears to be orthologous to Arabidopsis NAC 029 (*AT1G62300*), which has a known role in promoting leaf senescence [84] suggesting a similar role in switchgrass and reinforcing the evidence for NACs (*Pavirv00068192* and *Pavirv00065253*) to be involved in nutrient remobilization from senescing flag leaves. Unlike switchgrass and wheat, the rice and Arabidopsis orthologs to Wheat *NAM-B1* were most abundantly expressed in anthers [85, 86], and the orthologous rice NAC did not have a role in nutrient remobilization [85]. These data would indicate that some caution is needed before fully ascribing the role of these three switchgrass NACs as causally related to remobilization and senescence. Nevertheless, the expression data for the switchgrass NACs discussed above support their role in flag leaf senescence. Among the genes that displayed a significant co-expression profile with the NACs, were the *MOTs* (discussed above) and several *SAGs*, including genes coding for an autophagy protein and catalase. The autophagy ortholog in Arabidopsis
(AT3G19190; ATG2) is involved in the early stages of the biogenesis of the autophagosomes [87], and lowered expression of the catalase 2 ortholog in Arabidopsis (AT4G35090, CATALASE2) leads to premature senescence [88]. These data indicate that phase 4 switchgrass leaves were not fully senescent and required proteins such as CATALASE 2 to maintain cellular function and permit the remobilization of nutrients from the flag leaves to the rhizomes. Seeds were at physiological maturity at this harvest date, and unlikely to be active sinks, indicating a bulk of the remobilized nutrients was targeted to the below-ground tissues.

Our data provide a comprehensive, global transcriptomic inspection of switchgrass flag leaf development. The coordination of gene expression with known physiological processes during leaf development provides new insights into biochemistry, mineral metabolism, and epigenetic processes. Because we have identified specific transcripts, our data can be mined for other insights, and will inform future investigations. Overall, we identified several genes that could be important to signaling the onset of senescence and nutrient remobilization.

5.6 Acknowledgements

We thank Drs. James D. Eudy and Alok Dhar for sample preparation and analysis on the Illumina Hi-Seq 2000 instrument. We thank Dr. Aaron J. Saathoff for help with leaf collection. This work was supported in part by the Office of Science (BER), U. S. Department of Energy Grant Number DE-AL02-09ER64829, USDA-NIFA Grant Number 2011-67009-30096, and by the USDA-ARS CRIS project 5440-21000-030-00D. The U.S. Department of Agriculture, Agricultural Research Service, is an equal opportunity/affirmative action employer and all agency services are available without
discrimination. Mention of commercial products and organizations in this manuscript is solely to provide specific information. It does not constitute endorsement by USDA-ARS over other products and organizations not mentioned.

The University of Nebraska DNA Sequencing Core receives partial support from the NCRR (1S10RR027754-01, 5P20RR016469, RR018788-08) and the National Institute for General Medical Science (NIGMS) (8P20GM103427, GM103471-09). This publication’s contents are the sole responsibility of the authors and do not necessarily represent the official views of the NIH or NIGMS.
5.7 Works Cited


62. Srivalli S, Khanna-Chopra R: **Delayed wheat flag leaf senescence due to removal of spikelets is associated with increased activities of leaf antioxidant enzymes, reduced glutathione/oxidized glutathione ratio and oxidative damage to mitochondrial proteins.** *Plant Physiol Biochem* 2009, **47**(8):663-670.


64. Gallie DR: **The role of L-ascorbic acid recycling in responding to environmental stress and in promoting plant growth.** *J Exp Bot* 2013, **64**(2):433-443.


83. Ricachenevsky FK, Menguer PK, Sperotto RA: **kNACKing on heaven's door: how important are NAC transcription factors for leaf senescence and Fe/Zn remobilization to seeds?** *Frontiers in plant science* 2013, 4:226.

84. Guo Y, Gan S: **AtNAP, a NAC family transcription factor, has an important role in leaf senescence.** *Plant J* 2006, 46(4):601-612.


Chapter 6

Summary and Conclusions
6.1 Final Discussion

Results in this dissertation highlighted several molecular aspects underlying plant development in switchgrass. One goal was to develop an efficient bioinformatics pipeline to analyze high-throughput sequencing (HTS) data within a biological context. Once this pipeline became available, it was possible to study disparate plant developmental processes. At the time of the first series of experiments, a draft of the switchgrass genome was not available and Illumina sequencing was not a viable option. An emphasis at the start of my dissertation project was to begin a molecular understanding of below ground metabolism in switchgrass using Roche 454 sequencing technology and existing bioinformatic tools.

As a first step, 454 sequencing of crown and rhizome RNA obtained from the cultivar Summer was performed. The sequencing data was assembled to generate the first detailed transcriptome of crown and rhizome tissues obtained from field grown plants. These results provided a detailed inventory of gene expression profiles in perenniating switchgrass tissues prior to transitioning for winter dormancy [1]. Over 30,000 new sequences were discovered, representing a large addition to switchgrass sequence resources from a tissue central to switchgrass perenniality. Querying the metabolic pathways of glycolysis/gluconeogenesis and starch and sucrose metabolism revealed significant levels of transcripts for nearly all enzymes in both pathways, suggesting active carbohydrate metabolism in crown and rhizome tissues. This first window into the rhizome transcription suggested that it would be possible to mine this and other 454 sequenced datasets to develop greater understanding of processes occurring at the end of the growing season. Among these processes was the possible transfer of
minerals and other nutrients from the senescing shoots to the crowns and rhizomes. Ultimately, remobilization of nutrients and minerals from the aerial portions to the perenniating structures is of significance for maintaining plant health and in the sustainable production of biomass.

Plants contain a large and diverse array of transporters which are responsible for moving minerals and other nutrients within the plant. The expression patterns of individual transporter genes, often present within large gene families, can be tissue and stage specific. A bulk of the studies performed to date on transporters and transporter genes have been done on Arabidopsis (*Arabidopsis thaliana*; *At*) and rice (*Oryza sativa*), and nothing is known in switchgrass. However, there is data to indicate that there are genotypic differences exist in both the acquisition and remobilization of minerals (and other nutrients) in switchgrass [2, 3]. These data indicate that there could be both stage and genotypic differences in the expression profiles of transporter genes. Using bioinformatic tools, it was possible to classify most of the transporter genes present in the draft version of the switchgrass genome (Pvi0) into specific families.

Identification of mineral transporters in switchgrass and analysis of their expression patterns in various tissues across developmental stages highlighted genes potentially involved in mineral reallocation between perennial and aerial tissues [4]. Members of large transporter families such as KUP (potassium), MGS (magnesium), ZIP (zinc), NRT (nitrate), and PHT (phosphate) exhibited diverse expression patterns in crowns and rhizomes. These expression patterns correlated with time points from specific developmental stages of the plant. Such patterns suggest mineral reallocation roles for certain transporters. Minerals such as P, Cl, K, and S are known to be mobilized
out of aerial tissues in switchgrass at the end of the growing season, while Si, Ca, and Mg are not [3, 5-8]. High levels of alkali metals, principally K, can be detrimental to some biomass conversion methodologies [9]. The identification of transporters associated with these minerals, and other nutrients, provides targets for future work in fine tuning switchgrass biomass characteristics for more efficient downstream conversion into fuels or chemicals.

The release of the first draft of the switchgrass genome by the Joint Genomes Institute (www.phytozome.org) also coincided with significant improvements in the Illumina sequencing platforms. These two developments enabled the application of RNA-Seq methodologies in switchgrass, which was not a feasible option with the Roche 454 instrument. RNA-Seq analyses enable the relative quantification of global gene expression levels in a sample by counting the occurrences of specific gene sequences aligned to a reference (annotated genome). Advancements in Illumina sequencing platforms facilitated sequencing of biological replicate samples at high read depths, allowing application of more accurate statistical testing methods for differential gene expression. To utilize the power of HTS (RNA-Seq) with metabolite profiling, a study was undertaken to query the transcriptomes of switchgrass crowns and rhizomes obtained from contrasting tetraploid populations at a single developmental stage. The two populations analyzed were the upland tetraploid cv. Summer, adapted to the Upper Midwest, and the lowland tetraploid cv. Kanlow, adapted to the Southeast US.

RNA-Seq and metabolic comparisons of crown and rhizomes from upland (cv. Summer) and lowland (cv. Kanlow) plants emphasized the effects of latitudinal photoperiod adaptation on end of season developmental timing [10]. Source plants for
the cv. Summer originated in southeast Nebraska while source plants for cv. Kanlow originated in central Oklahoma [11]. Movement of cv. Kanlow plants further north into Nebraska resulted developmental delays due to different responses to day length (photoperiod). Metabolic, transcriptomic, enzymatic, and proteomic analyses suggested that Kanlow crowns and rhizomes were still in a state of active growth, whereas Summer crowns and rhizomes were redirecting metabolic activities in preparation for winter dormancy. A delay in the initiation of dormancy in Kanlow plants may be a significant factor behind observations of winter kill in Kanlow populations at the field sites used in this study (ARDC, near Ithaca, NE: 41.1603° N). A long term goal of this research is to understand the cellular mechanisms that control the transition to dormancy, impose dormancy, and directly influence the winter hardiness of the plant [12]. As described elsewhere in this dissertation, aerial senescence at the end of the growing season has a profound effect on the dormancy-related mechanism in the perenniating structures of the switchgrass plant. However, there is no knowledge of the signals that could originate from senescing shoots (primarily leaves). Therefore, an in depth transcriptomic analysis of switchgrass flag leaves was performed [13]. A molecular roadmap was established, dividing flag leaf development into four phases based on chlorophyll levels and metabolic enzyme expression levels. Phases 1 and 2 included up-regulated genes associated with early leaf development and chlorophyll biosynthesis respectively. Phase 3 up-regulated genes were functionally associated with carbon and nitrogen assimilation. Genes related to senescence were highly expressed during phase 4, including chlorophyll degradation enzymes and a large number of senescence-associated genes (SAGs). Specifically, three NAC (NAM, ATAF1/2, and CUC2) family transcription factors were
identified with high expression levels in senescing leaves: two NAC transcription factors closely related to wheat NAM-B1, which impacts mineral remobilization [14], and one NAC transcription factor orthologous to *AtNAC029*, which promotes senescence in Arabidopsis plants. Genes with similar expression profiles as these NACs included ammonium, nitrate, ureide, and molybdenum transporters. The overexpression of such transporters during flag leaf senescence suggest a possible link between aerial senescence and dormancy onset and represent targets for further investigation into this association.

In a companion study, recently submitted for peer review (titled “The WRKY transcription factor family and senescence in switchgrass” with authors Rinerson CI, Scully ED, Palmer NA, Donze-Reiner T, Rabara RC, Tripathi P, Shen QJ, Sattler SE, Rohila JS, Sarath G, and Rushton PJ), members of the WRKY transcription factor family (named for their characteristic protein domain sequence WRKYGQK) in switchgrass were identified and the relationship between individual WRKY gene expression and leaf senescence was investigated. 240 WRKY transcription factors were identified and named in the switchgrass genome and their expression patterns were queried in the switchgrass flag leave RNA-Seq dataset [13]. Differential expression and co-expression clustering analyses resulted in the division of 79 differentially expressed WRKYGQK into five co-expression modules which show clear leaf developmental stage associations (Figure 1). One module (ME3) contains 23 WRKYGQKs that demonstrate peak expression during senescence. Promoter analysis of these WRKYGQKs along with the other near 3,000 switchgrass genes in the co-expression module revealed the WRKYGQK target W box [15] element in the promoter of 1,536 genes, with 681 having 2-4 elements and 42 genes having 5-7 elements. The results suggest that WRKY genes may play a significant role
Figure 1. Co-expression analysis and clustering of switchgrass WRKYs.
Highlighted co-expression modules of identified WRKYs and other genes in switchgrass flag leaves. The representative expression profile (module eigengene) is shown for each module, along with the total number of genes (n) and the number of WRKYs (WRKY) present in each co-expression module.
in senescence processes and may be involved in the direct regulation of over half of the genes induced during senescence in switchgrass flag leaves. Delaying senescence of aerial tissues in switchgrass is one approach that may be used to increase overall biomass yield [12].

6.2 Ongoing Work

The results presented in this dissertation have provided a solid foundation for additional functional genomic studies on switchgrass developmental processes. One such study currently in progress is an extensive investigation of crown and rhizome transcriptomes and metabolomes from several different populations of field grown tetraploid switchgrass.

In the first study, replicate samples of cv. Summer crowns and rhizomes harvested at 11 time points across two years have been analyzed using RNA-Seq (based on HTS on an Illumina HiSeq2000 instrument) to query gene expression, and GCMS methods to query metabolites. Additionally, tissues were analyzed for starch, sucrose, and plant hormones (first year only). Preliminary results from these analyses show carbohydrate profiles similar to those observed by Anderson et al. [16] in leafy spurge (Figure 2). Starch levels increased in rhizomes throughout the growing season until the start of aerial senescence (September), at which point they began to decrease. The much greater levels of starch levels detected in 2011 as compared to 2010 is most likely a function of the overall increased amount of aerial biomass as plants became fully established [17, 18]. Sucrose levels were significantly increased in rhizomes harvested in September (aerial senescence largely completed) and November (dormancy established). Sucrose levels were essentially comparable between the different harvest years indicating that the
Figure 2. Starch and sucrose levels in Summer crowns and rhizomes. (A) Measured starch content in crowns and rhizomes of cv. Summer plants harvested across the growing season for two sequential years. (B) Measured sucrose content in crowns and rhizomes of cv. Summer plants. Error bars represent the Standard Error.

Production of sucrose was essentially rhizome specific at these time periods, unlike starch. The overall patterns starch accumulation and loss and sucrose biosynthesis supported earlier observations of high expression levels of genes in the starch and sucrose biosynthesis pathway [1, 10]. The levels of phytohormones, abscisic acid (ABA) and an active form of gibberellic acid (GA4), had contrasting patterns across the growing season.
Gibberellic acid promotes plant growth [19, 20] and GA4 was most abundant in July (period of active growth) and least abundant in May and November (periods of low or no growth) (Figure 3A). In contrast ABA, is a known regulator of dormancy-related processes [21], was present at low levels through the growing months and then increased significantly during September and November (Figure 3B). When compared as ratios, there was a distinct increase in the amount of ABA relative to GA in Summer rhizomes as the growing season progressed (Figure 3C), suggesting that this change could be one of the key signaling events for the onset of dormancy. A preliminary analysis of the RNA-Seq data indicated that over 200 ABA-responsive genes were significantly over-expressed during dormancy.

Among the group of genes regulated by ABA were those coding for enzymes catalyzing the biosynthesis of the raffinose family of oligosaccharides (RFOs). RFOs consist of galactose conjugated to myo-inositol, sucrose, or raffinose resulting in galactinol, raffinose, and stachyose respectively. All of these sugars serve as osmoprotectants and can act as scavengers of reactive oxygen species [22, 23]. Measurement of galactinol levels by GCMS revealed an abundance pattern similar to that of sucrose, with significant increases of galactinol levels in crowns and rhizomes in September and November harvests (Figure 4).

Completion of analyses on this Summer crown and rhizome dataset will yield a detailed picture of the transcriptional and metabolic events occurring during dormancy progression in a latitudinally well-adapted population. Subsequent analyses of samples from cultivar Kanlow and experimental strain Kanlow N1 Early Maturing, High Yield
Figure 3. ABA and GA4 levels in Summer crowns and rhizomes. (A) Measured gibberellic acid (GA4) levels in crowns and rhizomes of cv. Summer plants across the growing season. (B) Measured abscisic acid (ABA) levels in crowns and rhizomes of cv. Summer plants. (C) Log2 Fold Change of GA4 / ABA levels showing GA4 is significantly more abundant in May through August, while ABA becomes increasingly higher in September and November. Error bars represent the Standard Error.
collected simultaneously with Summer samples will enable a more detailed dissection of molecular events that are common or different to the three populations. They will also yield data on the transcriptional changes that occur in response to the selection pressure applied to the base Kanlow population to produce the Kanlow N1 Early Maturing, Yigh Yield plants.

Integrating the flag leaf RNA-Seq datasets [13] with the initial Summer crown and rhizome datasets enabled the generation of a three stage system model for senescence and dormancy progression (Figure 5). Stage one proceeds from vegetative growth through tiller maturity. Hallmarks of this stage consist of active bi-directional transport between above and below ground tissues, abundance of the phytohormone GA4 in crowns and rhizomes, and steadily increasing starch content in below ground tissues. Stage two is the transitional point when aerial tissues begin the senescence process. Key features of this stage are the reduction of GA4, an increase in ABA levels in crowns and rhizomes, and a shift in nutrient transport balance heavily weighted towards translocation from the aerial to the below ground tissues. Of specific interest is further investigation into metabolites transported during this stage as they may be key signals initiating dormancy processes in perenniating tissues. Stage three is the below ground dormancy stage, highlighted by high ABA levels and the beginning of a reduction in starch levels in crowns and rhizomes. Integration of existing RNA-Seq datasets, as of yet unanalyzed RNA-Seq datasets, and additional targeted experiments will enable the development of a more complete systems-based model that can account for the relationships between the progression of aerial senescence to below-ground winter dormancy in switchgrass.

The identification of key genes (for example *NAMB-1*-like *NAC, AtNAC29*-like
NAC, and *PviWRKY35*) and metabolites (for example galactinol and starch) that are causally associated with a specific stage of plant development will provide first set of targets to query differential developmental progression in switchgrass germplasm. It can be anticipated that changes in the levels of gene expression or target metabolites will be reflective of the adaptability of a specific genotype to the environment. Once these relationships are established, it will become possible to associate genetic architecture to a phenotype. These data can be utilized in breeding programs to more efficiently detect superior germplasm.

**Figure 4.** **Galactinol levels in Summer crowns and rhizomes.** Measured galactinol levels by GCMS in crowns and rhizomes of cv. Summer plants across the growing season for two sequential years. Error bars represent the Standard Error.
Figure 5. End of season system model. A system model highlighting three stages of developmental progression from aerial growth through below ground dormancy in switchgrass. Phytohormone (GA / ABA), transport processes, and starch and sucrose levels are key markers in this system.
6.3 **Works Cited**


Appendix A  Body of Work

7.1  Genetic background impacts soluble and cell wall-bound aromatics in brown midrib mutants of sorghum

Citation:

Abstract

Sorghum (Sorghum bicolor (L.) Moench) BMR-6 and BMR-12 encode cinnamylalcohol dehydrogenase and caffeic acid-O-methyltransferase, respectively. We have evaluated the impact of two bmr alleles, bmr-6 and bmr-12, respectively, on soluble and wall-bound aromatics in near isogenic, wild-type (WT), bmr-6, bmr-12 and double-mutant (DM; bmr-6 and bmr-12) plants in two genetic backgrounds, RTx430 and Wheatland. Immunoblots confirmed that COMT protein was essentially absent in bmr-12 and DM plants, but was present in bmr-6 and WT plants. In contrast, although CAD activity was not detected in bmr-6 and DM plants, proteins crossreacting to anti-CAD sera were found in stem extracts from all genotypes. In both sorghum backgrounds, WT plants had lowest amounts of free aromatics, higher levels of cell wall-bound pCA and FA esters and guaiacyl (G), syringyl (S), and p-hydroxyphenyl (H) lignins. Soluble aromatics and cell wall phenolic ester content in Wheatland DM plants resembled that of Wheatland bmr-6 plants, whereas in the RTx430 background, levels of these components in the DM plants more closely resembled those observed in bmr-12 plants. In both backgrounds, bmr-6 plants exhibited reduced levels of G, S, and H lignins relative to WT, and increased...
incorporation of G-indene into lignin. In bmr-12 plants, there was greater incorporation of G- and 5-hydroxyguaiacyl (5-OHG) lignin into cell walls. Histochemical staining of internode sections from Wheatland plants indicated that apparent lignification of cortical sclerenchyma and vascular bundle fibers was greatest and most uniform in WT plants. Relative staining intensity of these tissues was decreased in bmr-6, followed by bmr-12 plants. DM plants exhibited poor staining of cortical sclerenchyma and vascular bundle fibers.
7.2 A Nonsense Mutation in a Cinnamyl Alcohol Dehydrogenase Gene Is Responsible for the Sorghum brown midrib6 Phenotype

Citation:

Abstract:
brown midrib6 (bmr6) affects phenylpropanoid metabolism, resulting in reduced lignin concentrations and altered lignin composition in sorghum (Sorghum bicolor). Recently, bmr6 plants were shown to have limited cinnamyl alcohol dehydrogenase activity (CAD; EC 1.1.1.195), the enzyme that catalyzes the conversion of hydroxycinnamoyl aldehydes (monolignals) to monolignols. A candidate gene approach was taken to identify Bmr6. Two CAD genes (Sb02g024190 and Sb04g005950) were identified in the sorghum genome based on similarity to known CAD genes and through DNA sequencing a nonsense mutation was discovered in Sb04g005950 that results in a truncated protein lacking the NADPH-binding and C-terminal catalytic domains. Immunoblotting confirmed that the Bmr6 protein was absent in protein extracts from bmr6 plants. Phylogenetic analysis indicated that Bmr6 is a member of an evolutionarily conserved group of CAD proteins, which function in lignin biosynthesis. In addition, Bmr6 is distinct from the other CAD-like proteins in sorghum, including SbCAD4 (Sb02g024190). Although both Bmr6 and SbCAD4 are expressed in sorghum internodes, an examination of enzymatic activity of recombinant Bmr6 and SbCAD4 showed that Bmr6 had 1 to 2 orders of magnitude greater activity for monolignol substrates. Modeling
of Bmr6 and SbCAD4 protein structures showed differences in the amino acid composition of the active site that could explain the difference in enzyme activity. These differences include His-57, which is unique to Bmr6 and other grass CADs. In summary, Bmr6 encodes the major CAD protein involved in lignin synthesis in sorghum, and the bmr6 mutant is a null allele.
7.3 A Continuous, Quantitative Fluorescent Assay for Plant Caffeic Acid O-Methyltransferases

Citation:

Abstract
Plant caffeic acid O-methyltransferases (COMTs) use S-adenosylmethionine (ado-met), as a methyl donor to transmethylate their preferred (phenolic) substrates in vivo, and will generally utilize a range of phenolic compounds in vitro. Collazo et al. (Anal. Biochem. 2005, 342, 86-92) have published a discrete, end-point fluorescence assay to detect histone methyltransferases using S-adenosyl homocysteine hydrolase and adenosine deaminase as coupling enzymes and a thiol-specific fluorophore, Thioglo1, as the detecting reagent. Using this previous assay as a guide, we have developed and validated a facile, sensitive and real-time fluorescence assay for characterizing plant COMTs and in the process simplified the original assay as well by obviating the need for adenosine deaminase in the assay, and simultaneously converting an end-point assay into a continuous one. Our assay has been used to kinetically characterize recombinant sorghum COMT (Bmr-12) a key enzyme involved in cell wall lignification, and analyze COMT activity in maturing tillers from switchgrass plants. Data indicated that the calculated K(m) and V(max) values for the recombinant sorghum COMT using different substrates in the fluorescent assay were similar to published values for COMT enzymes from other plant species. Native COMT activity was greatest in internodes at the top of a tiller and
declined in the more basal internodes. This new assay should have broad applicability for characterizing COMTs and potentially other plant methyltransferases that utilize ado-met as a methyl donor.
7.4 Identification and Characterization of Four Missense Mutations in Brown midrib 12 (Bmr12), the Caffeic O-Methyltranferase (COMT) of Sorghum

Citation:

Abstract

Modifying lignin content and composition are targets to improve bioenergy crops for cellulosic conversion to biofuels. In sorghum and other C4 grasses, the brown midrib mutants have been shown to reduce lignin content and alter its composition. Bmr12 encodes the sorghum caffeic O-methyltransferase, which catalyzes the penultimate step in monolignol biosynthesis. From an EMS-mutagenized TILLING population, four bmr12 mutants were isolated. DNA sequencing identified the four missense mutations in the Bmr12 coding region, which changed evolutionarily conserved amino acids Ala71Val, Pro150Leu, Gly225Asp, and Gly325Ser. The previously characterized bmr12 mutants all contain premature stop codons. These newly identified mutants, along with the previously characterized bmr12-ref, represent the first allelic series of bmr12 mutants available in the same genetic background. The impacts of these newly identified mutations on protein accumulation, enzyme activity, Klason lignin content, lignin subunit composition, and saccharification yield were determined. Gly225Asp mutant greatly reduced protein accumulation, and Pro150Leu and Gly325Ser greatly impaired enzyme activity compared to wild type (WT). All four mutants significantly reduced Klason
lignin content and altered lignin composition resulting in a significantly reduced S/G ratio relative to WT, but the overall impact of these mutations was less severe than bmr12-ref. Except for Gly325Ser, which is a hypomorphic mutant, all mutants increased the saccharification yield relative to WT. These mutants represent new tools to decrease lignin content and S/G ratio, possibly leading toward the ability to tailor lignin content and composition in the bioenergy grass sorghum.
7.5 Next-Generation Sequencing of Crown and Rhizome Transcriptome from an Upland, Tetraploid Switchgrass

Citation:

Abstract
The crown and rhizome transcriptome of an upland tetraploid switchgrass cultivar cv Summer well adapted to the upper Midwest was investigated using the Roche 454-FLX pyrosequencing platform. Overall, approximately one million reads consisting of 216 million bases were assembled into 27,687 contigs and 43,094 singletons. Analyses of these sequences revealed minor contamination with non-plant sequences (< 0.5%), indicating that a majority were for transcripts coded by the switchgrass genome. Blast2Gos comparisons resulted in the annotation of similar to 65% of the contig sequences and similar to 40% of the singleton sequences. Contig sequences were mostly homologous to other plant sequences, dominated by matches to Sorghum bicolor genome. Singleton sequences, while displaying significant matches to S. bicolor, also contained sequences matching non-plant species. Comparisons of the 454 dataset to existing EST collections resulted in the identification of 30,177 new sequences. These new sequences coded for a number of different proteins and a selective analysis of two categories, namely, peroxidases and transcription factors, resulted in the identification of specific peroxidases and a number of low-abundance transcription factors expected to be involved in chromatin remodeling. KEGG maps for glycolysis and sugar metabolism
showed high levels of transcript coding for enzymes involved in primary metabolism. The assembly provided significant insights into the status of these tissues and broadly indicated that there was active metabolism taking place in the crown and rhizomes at post-anthesis, the seed maturation stage of plant development.
7.6 Towards uncovering the roles of switchgrass peroxidases in plant processes

Citation:


Abstract

Herbaceous perennial plants selected as potential biofuel feedstocks had been understudied at the genomic and functional genomic levels. Recent investments, primarily by the U.S. Department of Energy, have led to the development of a number of molecular resources for bioenergy grasses, such as the partially annotated genome for switchgrass (*Panicum virgatum* L.), and some related diploid species. In its current version, the switchgrass genome contains 65,878 gene models arising from the A and B genomes of this tetraploid grass. The availability of these gene sequences provides a framework to exploit transcriptomic data obtained from next-generation sequencing platforms to address questions of biological importance. One such question pertains to discovery of genes and proteins important for biotic and abiotic stress responses, and how these components might affect biomass quality and stress response in plants engineered for a specific end purpose. It can be expected that production of switchgrass on marginal lands will expose plants to diverse stresses, including herbivory by insects. Class III plant peroxidases have been implicated in many developmental responses such as lignification and in the adaptive responses of plants to insect feeding. Here, we have analyzed the class III peroxidases encoded by the switchgrass genome, and have mined available transcriptomic datasets to develop a first understanding of the expression profiles of the
class III peroxidases in different plant tissues. Lastly, we have identified switchgrass peroxidases that appear to be orthologs of enzymes shown to play key roles in lignification and plant defense responses to hemipterans.
7.7 Global changes in mineral transporters in tetraploid switchgrasses (Panicum virgatum L)

Citation:

Abstract
Switchgrass (Panicum virgatum L) is perennial, C-4 grass with great potential as a biofuel crop. An in-depth understanding of the mechanisms that control mineral uptake, distribution and remobilization will benefit sustainable production. Nutrients are mobilized from aerial portions to below-ground crowns and rhizomes as a natural accompaniment to above-ground senescence post seed-set. Mineral uptake and remobilization is dependent on transporters, however, little if any information is available about the specific transporters that are needed and how their relative expression changes over a growing season. Using well-defined classes of mineral transporters, we identified 520 genes belonging to 40 different transporter classes in the tetraploid switchgrass genome. Expression patterns were determined for many of these genes using publically available transcriptomic datasets obtained from both greenhouse and field grown plants. Certain transporters showed strong temporal patterns of expression in distinct developmental stages of the plant. Gene-expression was verified for selected transporters using gRT-PCR. By and large these analyses confirmed the developmental stage-specific expression of these genes. Mineral analyses indicated that K, Fe, Mg, Co, and As had a similar pattern of accumulation with apparent limited remobilization at the end of the
growing season. These initial analyses will serve as a foundation for more detailed examination of the nutrient biology of switchgrass.
7.8  Contrasting Metabolism in Perenniating Structures of Upland and Lowland Switchgrass Plants Late in the Growing Season

Citation:

Abstract

Background: Switchgrass (*Panicum virgatum* L.) is being developed as a bioenergy crop for many temperate regions of the world. One way to increase biomass yields is to move southern adapted lowland cultivars to more northern latitudes. However, many southerly adapted switchgrass germplasm can suffer significant winter kill in northerly climes.

Materials and Methods: Here, we have applied next-generation sequencing in combination with biochemical analyses to query the metabolism of crowns and rhizomes obtained from two contrasting switchgrass cultivars. Crowns and rhizomes from field-grown lowland (cv Kanlow) and upland (cv Summer) switchgrass cultivars were collected from three randomly selected post-flowering plants. Summer plants were senescing, whereas Kanlow plants were not at this harvest date.

Results: Principal component analysis (PCA) differentiated between both the Summer and Kanlow transcriptomes and metabolomes. Significant differences in transcript abundances were detected for 8,050 genes, including transcription factors such as WRKYs and those associated with phenylpropanoid biosynthesis. Gene-set enrichment analyses showed that a number of pathways were differentially up-regulated in the two populations. For both populations, protein levels and enzyme activities agreed well with
transcript abundances for genes involved in the phenylpropanoid pathway that were up-regulated in Kanlow crowns and rhizomes. The combination of these datasets suggests that dormancy-related mechanisms had been triggered in the crowns and rhizomes of the Summer plants, whereas the crowns and rhizomes of Kanlow plants had yet to enter dormancy.

Conclusions: Delayed establishment of dormancy at more northerly latitudes could be one factor that reduces winter-survival in the high-yielding Kanlow plants. Understanding the cellular signatures that accompany the transition to dormancy can be used in the future to select plants with improved winter hardiness.
Switchgrass (Panicum virgatum L) flag leaf transcriptomes reveal molecular signatures of leaf development, senescence, and mineral dynamics

Citation:

Abstract

Switchgrass flag leaves can be expected to be a source of carbon to the plant, and its senescence is likely to impact the remobilization of nutrients from the shoots to the rhizomes. However, many genes have not been assigned a function in specific stages of leaf development. Here, we characterized gene expression in flag leaves over their development. By merging changes in leaf chlorophyll and the expression of genes for chlorophyll biosynthesis and degradation, a four-phase molecular roadmap for switchgrass flag leaf ontogeny was developed. Genes associated with early leaf development were up-regulated in phase 1. Phase 2 leaves had increased expression of genes for chlorophyll biosynthesis and those needed for full leaf function. Phase 3 coincided with the most active phase for leaf C and N assimilation. Phase 4 was associated with the onset of senescence, as observed by declining leaf chlorophyll content, a significant up-regulation in transcripts coding for enzymes involved with chlorophyll degradation, and in a large number of senescence-associated genes. Of considerable interest were switchgrass NAC transcription factors with significantly higher expression in senescing flag leaves. Two of these transcription factors were
closely related to a wheat NAC gene that impacts mineral remobilization. The third switchgrass NAC factor was orthologous to an Arabidopsis gene with a known role in leaf senescence. Other genes coding for nitrogen and mineral utilization, including ureide, ammonium, nitrate, and molybdenum transporters, shared expression profiles that were significantly co-regulated with the expression profiles of the three NAC transcription factors. These data provide a good starting point to link shoot senescence to the onset of dormancy in field-grown switchgrass.
7.10 Characterization of Greenbug Feeding Behavior and Aphid (Hemiptera: Aphididae) Host Preference in Relation to Resistant and Susceptible Tetraploid Switchgrass Populations

Citation:

Abstract
Two choice studies were performed to evaluate greenbug, Schizaphis graminum (Rondani), and yellow sugarcane aphid, Sipha flava (Forbes), preference for two tetraploid switchgrass populations, Summer and Kanlow, and one experimental hybrid, K x S, derived by crossing Kanlow (male) x Summer (female) plants. Additionally, an assessment of S. graminum feeding behavior was performed on the same switchgrass populations, by using the electrical penetration graph (EPG) technique. Choice studies for S. flava indicated a lack of antixenosis, with no preference by aphids among any of the switchgrass populations at any time point. However, choice studies with S. graminum indicated a preference for the K x S plants at 24 h after aphid introduction. No obvious differences were observed for the leaf surfaces between the three populations. Feeding behavior studies for S. graminum on switchgrasses indicated no differences for the time to first probe or time to first sieve element phase among switchgrass populations. However, duration of sieve element phases for S. graminum was significantly less on Kanlow compared to K x S and Summer. S. graminum also had a significantly lower
potential phloem ingestion index (PPII) and few aphids showing sustained phloem ingestion on Kanlow as compared to K x S and Summer plants. These results suggest that resistance factors (chemical or mechanical) in Kanlow are located in the phloem tissue. At the whole leaf level, some differences were observed for a subset of polar metabolites, although Kanlow plants were significantly enriched for oxalic acid.
7.11 Transcriptional responses of tolerant and susceptible soybeans to soybean aphid (Aphis glycines Matsumura) herbivory

Citation:

Abstract
The soybean aphid, Aphis glycines Matsumura, was introduced in 2000 to North America and has become one of the most significant pests to soybean, Glycine max (L.) Merrill, production. Possible solutions to this problem are the use of resistant plants and the understanding of the genes involved in plant resistance. In this study, we sought to better understand the genes involved in the tolerance response of soybean plants to the soybean aphid, utilizing tolerant (KS4202) and susceptible (K-03-4686) plants. Studies were conducted under greenhouse conditions. Leaf samples of both tolerant and susceptible plants were collected at day 5 and day 15 after infestation and analyzed by sequencing-by-synthesis on an Illumina GA II X instrument. In the tolerant genotype, 3 and 36 genes were found to be differentially expressed in the infested plants compared to the control treatments at day 5 and day 15, respectively. A similar comparison in the susceptible genotype revealed 0 and 11 genes to be differentially expressed at day 5 and day 15, respectively. Predominately, genes related to plant defense, such as WRKY transcription factors, peroxidases, and cytochrome p450s, were up-regulated in the tolerant genotype 15 days post-infestation by aphids. In contrast, none of these genes were similarly up-
regulated in the susceptible plants, suggesting that consistent elevation of defense responses is important to plant tolerance. However, significant genotypic differences in global gene expression were also found when transcriptomes from control uninfested plants were compared at both day 5 and 15. qPCR validation of select genes confirmed our RNA-seq data. These comparisons indicate that potentially broader regulation of transcriptomes also contributes to the tolerance response and provides data that the tolerant genotype (KS4202) could be useful in soybean breeding programs trying to minimize production losses accruing from soybean aphid feeding.
7.12 Transcriptional Profiling of Resistant and Susceptible Buffalograsses in Response to Blissus occiduus (Hemiptera: Blissidae) Feeding


Abstract

Understanding plant resistance mechanisms at a molecular level would provide valuable insights into the biological pathways impacted by insect feeding, and help explain specific plant tolerance mechanisms. As a first step in this process, we conducted next-generation sequencing using RNA extracted from chinch bug-tolerant and -susceptible buffalograss genotypes at 7 and 14 d after chinch bug feeding. Sequence descriptions and gene ontology terms were assigned to 1,701 differentially expressed genes. Defense-related transcripts were differentially expressed within the chinch bug-tolerant buffalograss, Prestige, and susceptible buffalograss, 378. Interestingly, four peroxidase transcripts had higher basal expression in tolerant control plants compared with susceptible control plants. Defense-related transcripts, including two peroxidase genes, two catalase genes, several cytochrome P450 transcripts, a glutathione s-transferase, and a WRKY gene were upregulated within the Prestige transcriptome in response to chinch bug feeding. The majority of observed transcripts with oxidoreductase activity, including nine peroxidase genes and a catalase gene, were downregulated in 378 in response to initial chinch bug feeding. The observed difference in transcript expression between these
two buffalograss genotypes provides insight into the mechanism(s) of resistance, specifically buffalograss tolerance to chinch bug feeding.