

University of Nebraska - Lincoln

## DigitalCommons@University of Nebraska - Lincoln

---

Publications from USDA-ARS / UNL Faculty

U.S. Department of Agriculture: Agricultural  
Research Service, Lincoln, Nebraska

---

2012

### Next-Generation Sequencing of Crown and Rhizome Transcriptome from an Upland, Tetraploid Switchgrass

Nathan A. Palmer

USDA, [nathan.palmer@ars.usda.gov](mailto:nathan.palmer@ars.usda.gov)

Aaron J. Saathoff

USDA

Jaehyoung Kim

University of Nebraska-Lincoln

Andrew Benson

University of Nebraska-Lincoln, [abenson1@unl.edu](mailto:abenson1@unl.edu)

Christian M. Tobias

USDA-ARS

*See next page for additional authors*

Follow this and additional works at: <https://digitalcommons.unl.edu/usdaarsfacpub>

---

Palmer, Nathan A.; Saathoff, Aaron J.; Kim, Jaehyoung; Benson, Andrew; Tobias, Christian M.; Twigg, Paul G.; Vogel, Kenneth P.; Madhavan, Soundararajan; and Sarath, Gautam, "Next-Generation Sequencing of Crown and Rhizome Transcriptome from an Upland, Tetraploid Switchgrass" (2012). *Publications from USDA-ARS / UNL Faculty*. 1953.

<https://digitalcommons.unl.edu/usdaarsfacpub/1953>

This Article is brought to you for free and open access by the U.S. Department of Agriculture: Agricultural Research Service, Lincoln, Nebraska at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Publications from USDA-ARS / UNL Faculty by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

---

**Authors**

Nathan A. Palmer, Aaron J. Saathoff, Jaehyoung Kim, Andrew Benson, Christian M. Tobias, Paul G. Twigg, Kenneth P. Vogel, Soundararajan Madhavan, and Gautam Sarath

# Next-Generation Sequencing of Crown and Rhizome Transcriptome from an Upland, Tetraploid Switchgrass

Nathan A. Palmer · Aaron J. Saathoff ·  
Jaehyoung Kim · Andrew Benson ·  
Christian M. Tobias · Paul Twigg · Kenneth P. Vogel ·  
Soundararajan Madhavan · Gautam Sarath

Published online: 13 December 2011  
© Springer Science+Business Media, LLC. (outside the USA) 2011

**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 ~65% of the contig sequences and ~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.

**Electronic supplementary material** The online version of this article (doi:10.1007/s12155-011-9171-1) contains supplementary material, which is available to authorized users.

N. A. Palmer · A. J. Saathoff · K. P. Vogel · G. Sarath  
Grain, Forage and Bioenergy Research Unit, USDA Central-East  
Biomass Regional Center,  
Lincoln, NE 68583-0937, USA

J. Kim · A. Benson  
Center for Applied Genomics and Ecology, Department of Food  
Science and Technology, University of Nebraska at Lincoln,  
Lincoln, NE 68583, USA

C. M. Tobias  
Genomics and Gene Discovery Research Unit, Western Regional  
Research Center, USDA-ARS,  
Albany, CA 94710, USA

P. Twigg  
Biology Department, University of Nebraska at Kearney,  
Kearney, NE 68849, USA

S. Madhavan  
Department of Biochemistry, University of Nebraska at Lincoln,  
Lincoln, NE 68588-0664, USA

N. A. Palmer · A. J. Saathoff · K. P. Vogel · G. Sarath  
Department of Agronomy and Horticulture,  
University of Nebraska at Lincoln,  
Lincoln, NE 68583, USA

G. Sarath (✉)  
Grain, Forage and Bioenergy Research Unit,  
USDA-ARS,  
137 Keim Hall,  
Lincoln, NE 68583-0937, USA  
e-mail: Gautam.Sarath@ars.usda.gov

**Keywords** 454 FLX pyrosequencing · Bioenergy · Crowns and rhizomes · Switchgrass · Transcriptome · Transcription factors

## Introduction

Increasingly, there is interest in the 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 square meter) 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 hardiness 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 publicly 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 USA. 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 a 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].

## Materials and Methods

### Plant Material

Stands of switchgrass cv Summer had been established in the field near Mead, NE, USA, 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 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 h of harvest. Adherent soil was removed using toothbrushes. 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  $-80^{\circ}\text{C}$  until use. Crowns and rhizomes were fine-milled either by hand or using a cryogenic grinder (6870 Freezer Mill (Spex Sample Prep, Metuchen, NJ, USA)). Pulverized plant material was used to extract RNA.

### RNA Extraction and cDNA Library Generation

Total RNA was extracted from all switchgrass tissues using the modified Trizol (Invitrogen, Carlsbad, CA, USA) protocol of Tobias et al. [15]. 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  $\mu$ l of RNase-free water with RNaseOUT ribonuclease inhibitor added. The pellets from both aliquots were resuspended sequentially in the same 50  $\mu$ l of water with heating at 60°C for 5 min each. Any undissolved pellet material was discarded. From these samples, mRNA was isolated using the FastTrack MAG Maxi isolation kit and 100  $\mu$ l of magnetic beads as directed (Invitrogen, Carlsbad, CA, USA). The mRNA was quantitated using a Nanodrop spectrophotometer (Thermo Fisher, Waltham, MA). Synthesis of cDNA was performed using the high-yield protocol of QuantiTect Whole Transcriptome kit (Qiagen, Valencia, CA, USA) with 100 ng of mRNA as the starting material. The cDNA was purified from the reaction mixture using QIAamp DNA Blood Mini Kit (Qiagen, Valencia, CA, USA). 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/ $\mu$ l (total cDNA of 40–50  $\mu$ g provided for 454 sequencing).

#### 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, The University of Nebraska—Lincoln, Lincoln, NE, USA.

Briefly, 10 ng 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  $\mu$ l of the EB buffer (QIAGEN, Valencia, CA, USA) and DNA profile and quantification were measured by running 1  $\mu$ l of the samples on Agilent Bioanalyzer 2100 (Santa Clara, CA, USA) using a RNA Pico 6000 chip. The final sst DNA library was quantified using Qubit (Invitrogen, Carlsbad, CA, USA) and was diluted to a normalized concentration of  $1 \times 10^8$  molecules/ $\mu$ l 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 are shown in Table 1. Sequence files are available in the NCBI Sequence Read Archives under study: SRP009076; and Runs: SRR358964; SRR358965; and SRR358966.

#### Bioinformatics and De Novo Transcriptome Assembly

454 GS FLX titanium sequence data were assembled using Roche's GS De Novo Assembler (gsAssembler) software,

**Table 1** Assembly of 454 data

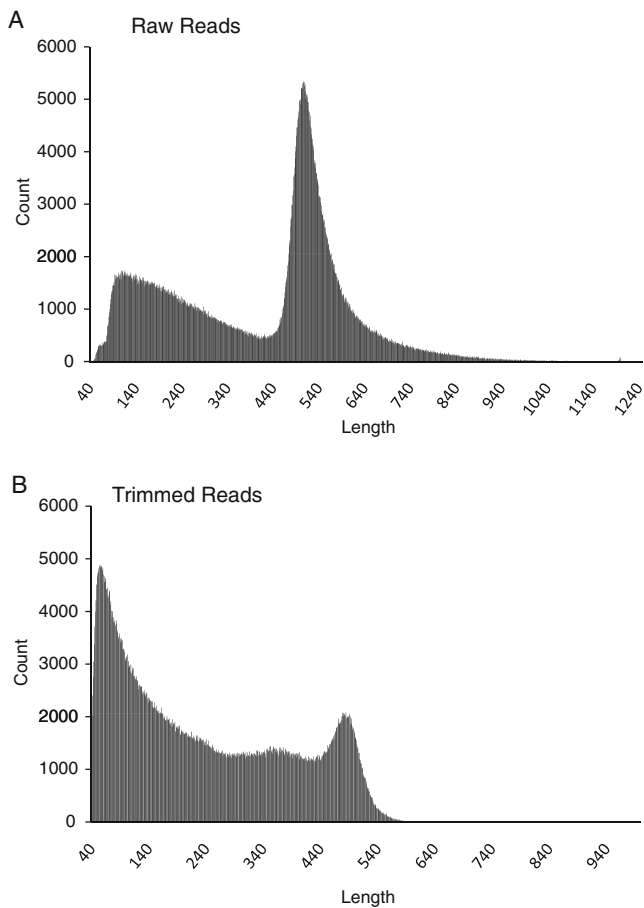
Total reads	928,820
Total bases	216,450,730
Aligned reads	641,443
Aligned bases	146,297,845
Inferred read error	2.23%
Singletons	170,312
Isogroups	12,548
Isogroups with one contig	7,288
Contigs	27,687

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.

#### 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 (Fig. 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 bimodal distribution was still evident (Fig. 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. A total of 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



**Fig. 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 50 bp were excluded from further analyses

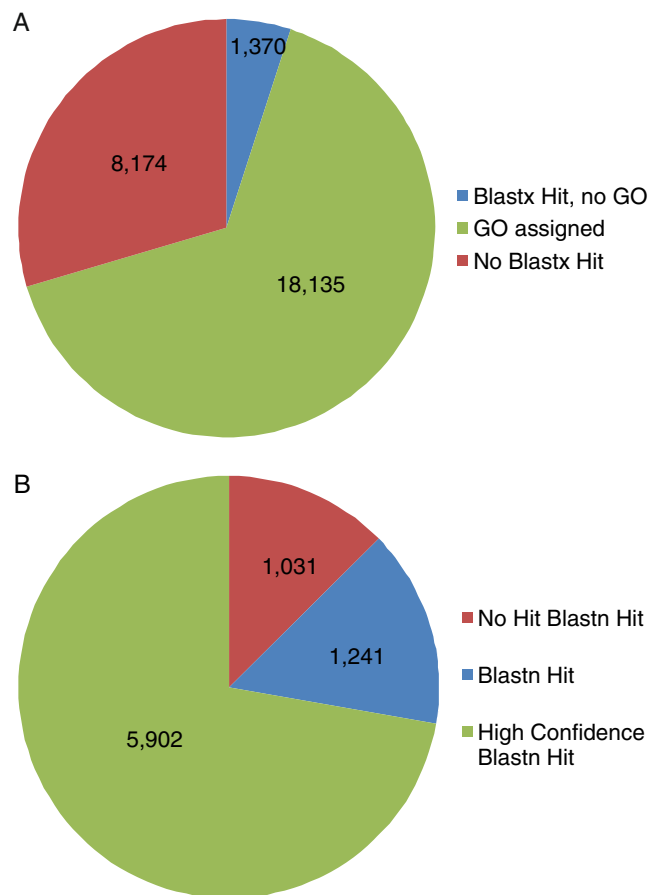
assemble into contigs using the Newbler 2.3 program and were categorized as singletons. The other aligned reads could be placed into 12,548 isogroups (gene models) and 27,687 contigs (Table 1). The average contig length was 722 bp and the median was 568 bp.

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 an  $e$ -value threshold of  $1 \times 10^{-7}$ . 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.

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/](http://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 (GO) 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 blast hit (Fig. 2a). The top 50 most abundant contigs are shown in Table S1 of “Electronic supplementary material”. 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 (Fig. 2b). Analysis of these 8,174 contigs indicated that approximately 87% of the remaining contig sequences matched other sequences

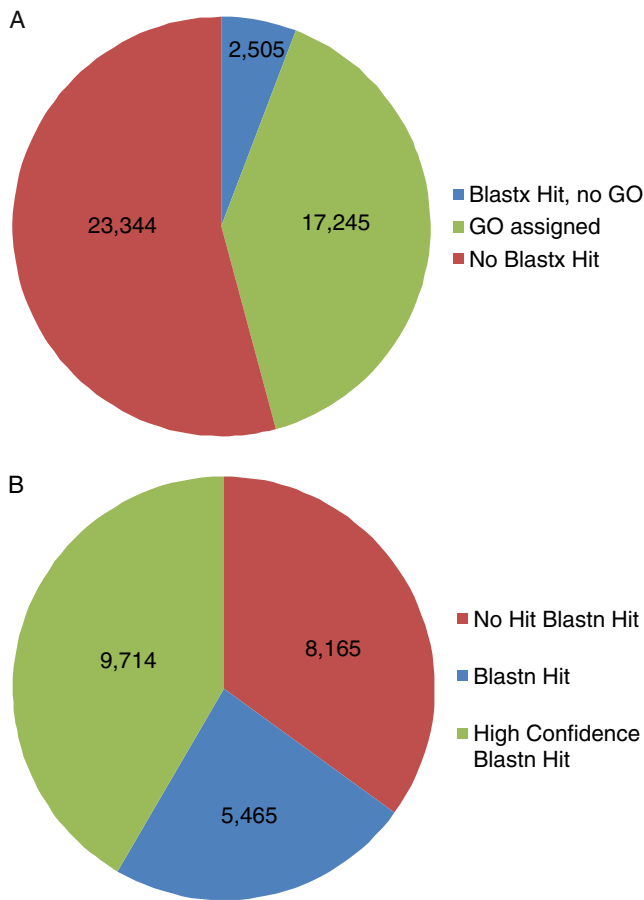


**Fig. 2** Pie chart showing results of Blast2GO alignments for contig sequences. **a** Distribution of sequences with GO terms assigned (green), Blast hit but no GO terms assigned at an  $e$ -value of  $1 \times 10^{-3}$  (blue), and no Blast hits at an  $e$ -value of  $1 \times 10^{-3}$  (red). **b** Reanalysis of the contig sequences with no Blast2GO assignments by Blastn, high-confidence blast hits with an  $e$ -value of  $10^{-30}$  or lower (green), hits with an  $e$ -value between  $10^{-3}$  and  $10^{-30}$  (blue), and no blast hits (red). Numbers in each section are the total number of contigs assigned to each category



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 (Fig. 3a, b). 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 (Fig. 3a). The singleton sequences (23,344) without a blastx hit were compared against the NCBI EST database using the blastn algorithm (Fig. 3b). A majority (65%) of the queried sequences displayed a match to existing EST sequences at an  $e$ -value of  $<1 \times 10^{-3}$  with 36% of the sequences having an EST match at the  $e$ -value of  $1 \times 10^{-30}$  cutoff. A total of 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,



**Fig. 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 the contig sequences with no Blast2GO assignments by Blastn, high confidence blast hits with an  $e$ -value cutoff of  $10^{-30}$  or lower (green), hits with an  $e$ -value cutoff between  $10^{-3}$  and  $10^{-30}$  (blue), and no blast hits (red). Numbers in each section are the total number of singleton sequences assigned to each category

56% displayed a match to EST sequences at an  $e$ -value threshold of  $1 \times 10^{-3}$ . 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  $<1 \times 10^{-25}$  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 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 Fig. 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 an  $e$ -value of  $1 \times 10^{-3}$ . As might be expected, increasing the stringency from an  $e$ -value of  $1 \times 10^{-3}$  to  $1 \times 10^{-50}$  resulted in a loss of 39% in the number of annotated sequences, with similar

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

Species	Blast hits <sup>a</sup>		No blast hits	
	Contigs	Singletons	Contigs	Singletons
<i>Sorghum bicolor</i>	17,553	11,523	10,134	31,571
<i>Brachypodium distachyon</i>	14,631	7,382	13,056	35,712
<i>Panicum virgatum</i> (EST)	22,488	16,710	5,199	26,384

<sup>a</sup> An  $e$ -value threshold of  $1 \times 10^{-15}$  was used for these analyses

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 the imperfect assembly of sequences introducing frame-shifts, 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 (Fig. 4a). There were fewer hits to other plant species, including a range of dicots (Fig. 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 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 (*Acyrtosiphon pisum*). Matches to a gram-negative bacterium (*Acinetobacter junii*) were also present in the singleton sequences (Fig. 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 was analyzed (Fig. 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

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

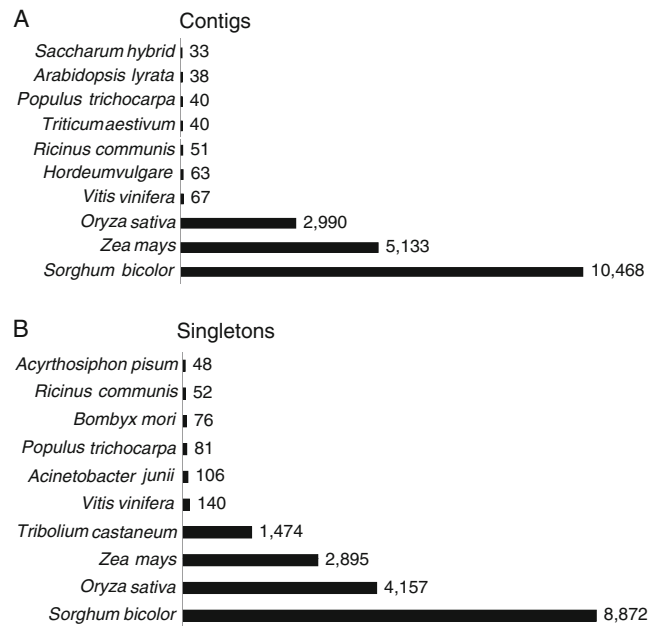
$e$ -value	Sequences <sup>a</sup>		Annotations <sup>b</sup>	EC	Number of sequences with EC <sup>c</sup>
	Number	Percent (%)			
$1 \times 10^{-3}$	16,708	92.1	72,772	8,221	6,291
$1 \times 10^{-10}$	15,979	88.1	69,473	7,912	6,079
$1 \times 10^{-15}$	15,102	83.3	65,704	7,527	5,781
$1 \times 10^{-25}$	13,484	74.4	58,683	6,849	5,273
$1 \times 10^{-50}$	10,223	56.4	44,388	5,400	4,167

EC number of enzyme codes recovered at each  $e$ -value threshold

<sup>a</sup> Total number of annotated sequences and percentage of total input sequences with a match

<sup>b</sup> Total number of annotations across all GO terms

<sup>c</sup> Total number of sequences with an enzyme code as identified by GO analysis

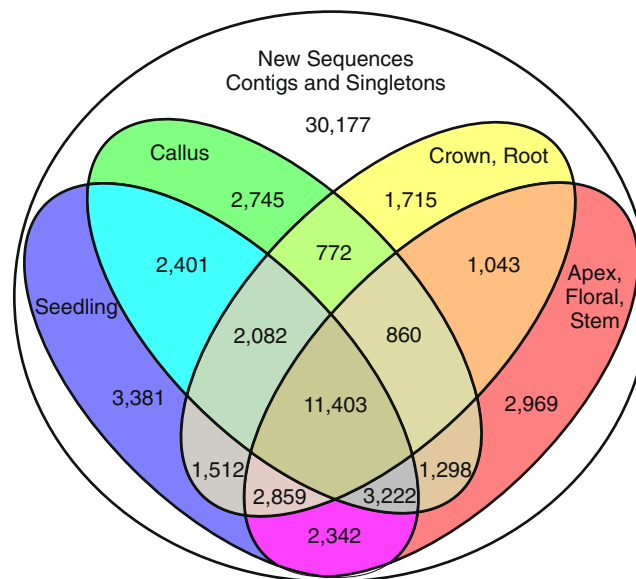


**Fig. 4** Top ten matches of switchgrass crown and rhizome sequences to other species. **a** Contigs. **b** Singletons. An  $e$ -value cutoff of  $10^{-3}$  or lower was used to designate matches to the species identified on the left axis

seeds. A total of 11,043 sequences were common to all of the tissue/stage-specific ESTs queried (Fig. 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/underrepresentation within broad and narrow GO terms.

At the “Biological Processes, Level 2 Terms,” there were some variations between 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 (Fig. 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/





**Fig. 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 publicly available databases. The assembled 454 sequences were compared against each library by Blastn with an *e*-value of  $10^{-25}$  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

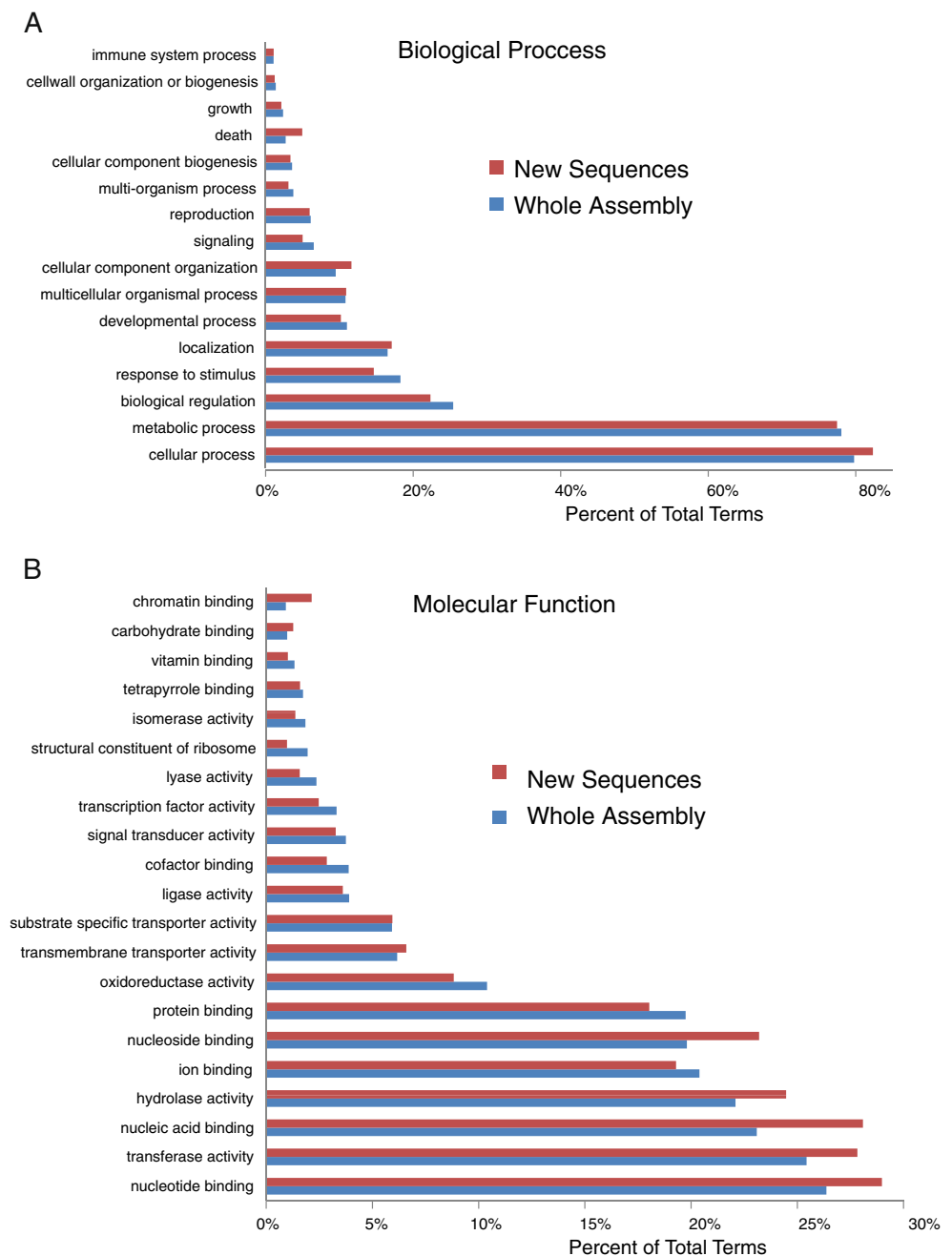
nucleic acid and chromatin binding as compared to the whole assembly (Fig. 6b). These new sequences also appeared to be enriched in transcripts coding for hydrolases and transferases as compared to the whole assembly. There was a slight decrease in transcripts coding for proteins assigned into the cofactor binding, structural constituent of ribosomes, and transcription factor activity categories (Fig. 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 a 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 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 five class III peroxidases present in the “new” sequences showed that they belonged to four 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], and 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

**Fig. 6** Comparison of the percent of sequences associated with a specific Blast2GO term for the whole assembly (*blue*) and the “new” sequences (*red*) (see Fig. 5). **a** Biological processes terms and **b** molecular processes terms



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; Fig. 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

**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)

C or S	Protein family	Predicted ortholog and group <sup>a</sup>
S	Ascorbate peroxidase	Cytosolic
C	Bacterial-induced peroxidase	Os02g0237000, group IV.4
S	Catalase	Catalase
C	Class III peroxidase	PviPrx-19, monocot-specific group V
C	Class III peroxidase	Os09g0471100, group IV
C	Class III peroxidase	Rice peroxidase 124, group III
S	Class III peroxidase	Os03g0762400, group VI
S	Peroxidasin homolog	Fatty acid dioxygenase
S	Thioredoxin peroxidase	2-cys peroxiredoxin

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

<sup>a</sup> Predicted rice ortholog based on Passardi et al. [24]

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) (Fig. 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 probably still occurring in the crowns and rhizomes of these plants.

### Discussion

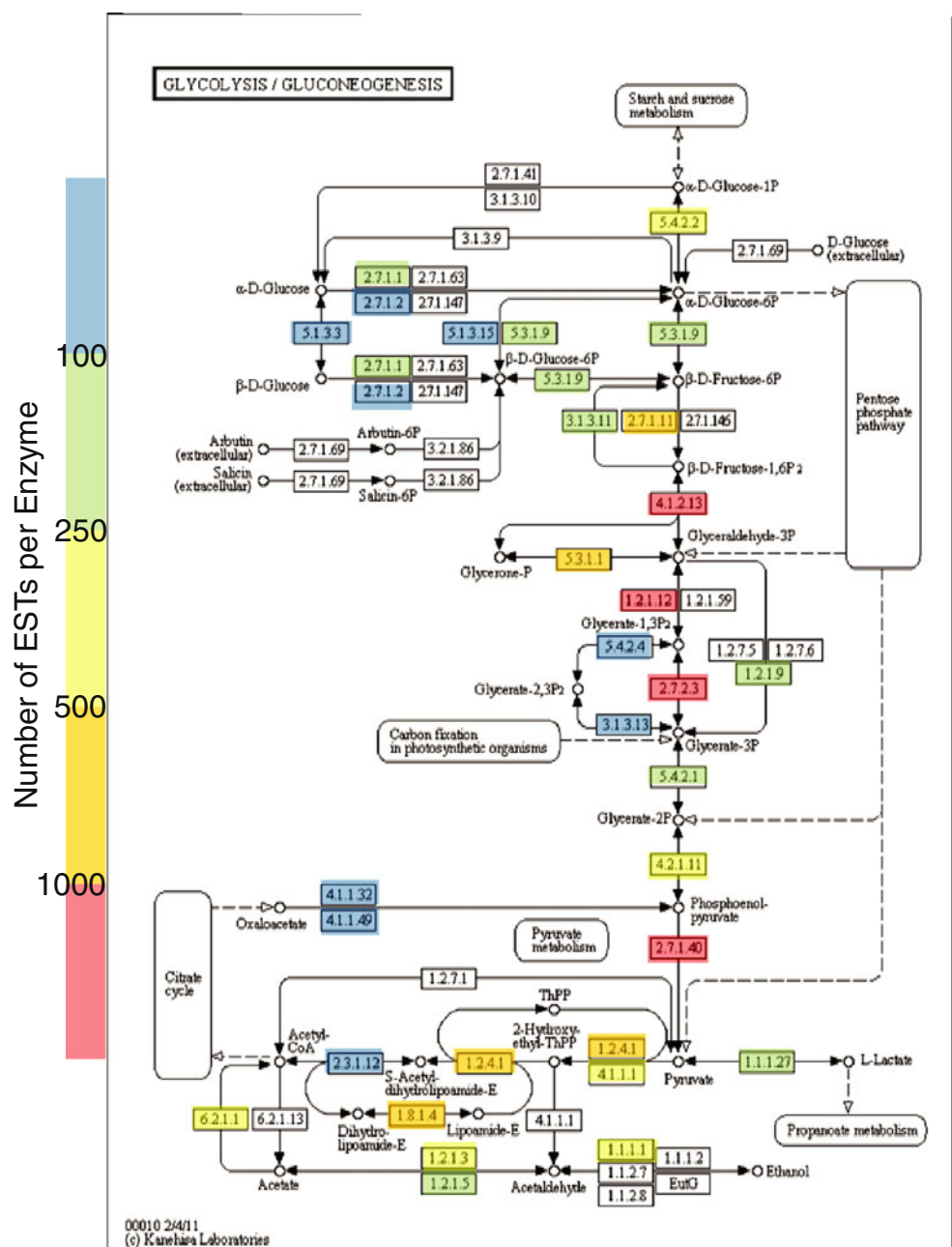
Perenniality in switchgrass is likely to be controlled by several mechanisms that ultimately impact the physiological status of the below-ground components of the plant [25]. These below-ground structures include the roots, rhizomes, and crowns. It can be anticipated that over a growing cycle there will be significant, but cyclical, changes in the

**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)

C or S	TF family	Predicted ortholog		Function of orthologs	References
S	AP-2 domain containing	RAV 1 and 2	(Atha)	Interactions with CONSTANS and FT	[33]
S	Antagonist of e2f	E2L	(Atha)	Represses E2F activation of genes	[34]
C	Auxin response	ARF-7	(Slyc)	Auxin and gibberellin signaling	[35]
S	bHLH family			Involved in rooting/root hairs	[36]
C	bHLH Family	HEC-1	(Atha)	Female flower development	[37]
S	basic leucine zipper	Rolled leaf	(Zmay)	Organ development	[38]
S	basic-leucine zipper	LIGULELESS-2 (Zmay)		Demarcates boundary of leaf and sheath	[39]
S	bZip	ABI-5	(Atha)	Abscisic acid signaling	[40]
S	bZip	bZIP 42	(Atha)	Organ formation	[41]
C	CBF/DREB-Like	CBF-7	(Atha)	Cold hardiness	[42]
S	CCAAT-binding factor	LEC1	(Zmay)	Embryo/fatty acid biosynthesis	[43]
S	Chromatin remodeling	SWI3D	(Atha)	Interactions with SWI/SNF complex	[44]
S	E2F	E2F	(Atha)	Control of cell cycle	[34]
C	Heat shock	SPL7	(Osat)	Control of leaf spots	[45]
S	Hox-family	BHL4	(Atha)	Leaf margin organization	[46]
S	Jumonji domain	ELF 6	(Atha)	Histone H3K4 demethylation	[47]
S	Jumonji-domain	REF6	(Atha)	Histone H3K27 demethylation	[48]
S	Myb family	DUO-1	(Atha)	Male germline formation	[49]
S	Myb r2r3 family	MYB101	(Atha)	Hormone signaling	[50]
S	NAM-superfamily	NAC 8	(Atha)	Suppressor of gamma response 1	[51]
S	NIN-Like	NLP7	(Atha)	Nitrate sensing	[52]
S	NF-X1 type	NFXL1	(Atha)	Protection under stress	[53]
C	Tudor/PWWP/MBT domain-containing protein	HUA2	(Atha)	Flowering/shoot morphology	[54]
C/S	WRKYs	Several members		ABA, UV responses biotic and abiotic stresses	[55]

C contig, S singleton, TF transcription factor, *Atha Arabidopsis thaliana*, *Osat O. sativa*, *Slyc Solanum lycopersicum*, *Zmay Z. mays*

**Fig. 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 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)

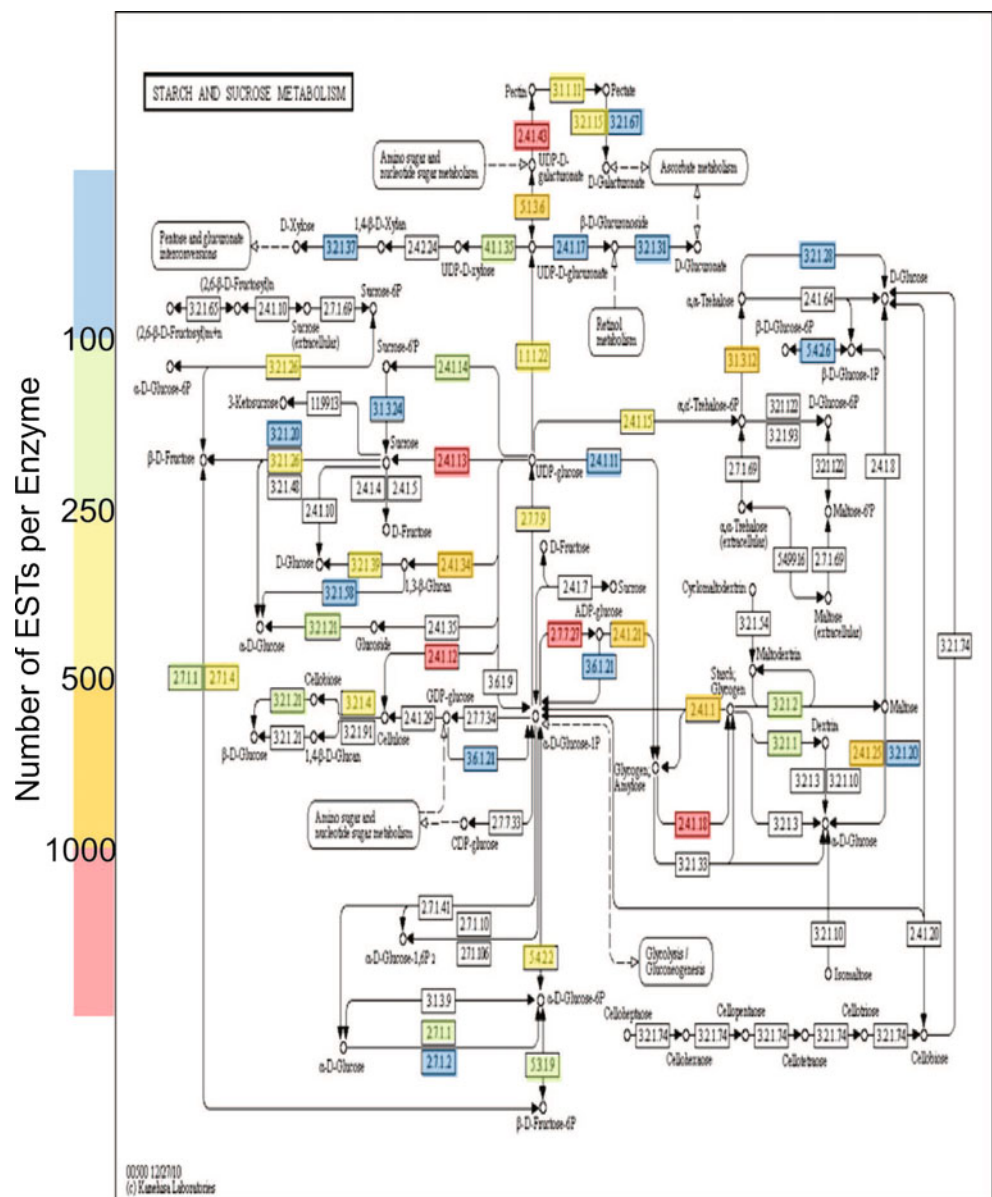


physiology of these organs underpinned by significant changes in the transcriptome. These changes will include nutrient remobilization and regeneration of new shoots (tillers) at the onset of green-up in early spring, increased tissue accretion over the growing season, transition to the slowing of developmental processes in the fall, and followed by a potentially 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

**Fig. 8** KEGG map for starch and sugar metabolism populated with transcripts coding for specific enzymes in the pathway. Other details are as described for Fig. 7



annotated genomes [12, 14, 27, 28] (Table S2 of “[Electronic supplementary material](#)”). 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 Fig. 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 h 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.

**Acknowledgements** We thank Steve Masterson and Patrick Callahan for excellent technical support. This work was supported by the Office of Science (BER), US Department of Energy Grant Number DE-AI02-09ER64829, and USDA-ARS CRIS project 5440-21000-028-00D. The US 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.

## References

- McLaughlin SB, Kszos LA (2005) Development of switchgrass (*Panicum virgatum*) as a bioenergy feedstock in the United States. *Biomass Bioenergy* 28(6):515–535
- Schmer MR, Vogel KP, Mitchell RB, Perrin RK (2008) Net energy of cellulosic ethanol from switchgrass. *Proc Natl Acad Sci U S A* 105(2):464–469
- Perlack RD, Wright LL, Turhollow A, Graham RL, Stokes BJ, Erblich DC (2005) Biomass as feedstock for a bioenergy and bio-products industry: the technical feasibility of a billion-ton annual supply. Oak Ridge National Laboratory, Oak Ridge
- Pedersen JF, Funnell DL, Vogel KP (2005) Impact of reduced lignin on plant fitness. *Crop Sci* 45:812–819
- Casler MD, Buxton DR, Vogel KP (2002) Genetic modification of lignin concentration affects fitness of perennial herbaceous plants. *Theor Appl Genet* 104(1):127–131
- Vogel KP, Hopkins AA, Moore KJ, Johnson KD, Carlson IT (2002) Winter survival in switchgrass populations bred for high IVDMD. *Crop Sci* 42(6):1857–1862
- Okada M, Lanzatella C, Saha MC, Bouton J, Wu R, Tobias CM (2010) Complete switchgrass genetic maps reveal subgenome collinearity, preferential pairing and multilocus interactions. *Genetics* 185(3):745–760. doi:10.1534/genetics.110.113910
- Gedye K, Gonzalez-Hernandez J, Ban Y, Ge X, Thimmapuram J, Sun F et al (2010) Investigation of the transcriptome of prairie cord grass, a new cellulosic biomass crop. *Plant Gen* 3(2):69–80. doi:10.3835/plantgenome2010.06.0012
- Huang SW, Guo SG, Zheng Y, Joung JG, Liu SQ, Zhang ZH et al (2010) Transcriptome sequencing and comparative analysis of cucumber flowers with different sex types. *BMC Genomics* 11. doi:10.1186/1471-2164-11-384
- Luo HM, Li Y, Sun C, Wu QO, Song JY, Sun YZ et al (2010) Comparison of 454-ESTs from *Huperzia serrata* and *Phlegmariurus carinatus* reveals putative genes involved in lycopodium alkaloid biosynthesis and developmental regulation. *BMC Plant Biol* 10. doi:10.1186/1471-2229-10-209
- Srivastava AC, Palanichelvam K, Ma J, Steele J, Blancaflor EB, Tang Y (2010) Collection and analysis of expressed sequence tags derived from laser capture microdissected switchgrass (*Panicum virgatum* L. Alamo) vascular tissues. *Bioenergy Res* 3(3):278–294
- Tranel PJ, Riggins CW, Peng YH, Stewart CN (2010) Characterization of de novo transcriptome for waterhemp (*Amaranthus tuberculatus*) using GS-FLX 454 pyrosequencing and its application for studies of herbicide target-site genes. *Pest Manag Sci* 66(10):1042–1052. doi:10.1002/ps.2006
- Der JP, Barker MS, Wickett NJ, dePamphilis CW, Wolf PG (2011) De novo characterization of the gametophyte transcriptome in bracken fern. *Pteridium aquilinum*. *BMC Genomics* 12. doi:10.1186/1471-2164-12-99
- Swarbreck SM, Lindquist EA, Ackerly DD, Andersen GL (2011) Analysis of leaf and root transcriptomes of soil-grown *Avena barbata* plants. *Plant Cell Physiol* 52(2):317–332. doi:10.1093/pcp/pcq188
- Tobias CM, Twigg P, Hayden DM, Vogel KP, Mitchell RM, Lazo GR et al (2005) Analysis of expressed sequence tags and the identification of associated short tandem repeats in switchgrass. *Theor Appl Genet* 111(5):956–964. doi:10.1007/s00122-005-0030-3
- Tobias CM, Sarath G, Twigg P, Lindquist E, Pangilinan J, Penning BW et al (2008) Comparative genomics in switchgrass using 61 585 high-quality expressed sequence tags. *Plant Genome* 1(2):111–124
- Vogel KP (2004) Switchgrass. In: Moser LE, Sollenberger L, Burson B (eds) Warm-season (C4) grasses. ASA-CSSA-SSSA Monograph No. 45. ASA-CSSA-SSSA, Madison, pp 561–588



18. Martinez-Reyna JM, Vogel KP (2008) Heterosis in switchgrass: spaced plants. *Crop Sci* 48(4):1312–1320. doi:10.2135/cropsci2007.12.0695
19. Vogel KP, Mitchell KB (2008) Heterosis in switchgrass: biomass yield in swards. *Crop Sci* 48(6):2159–2164. doi:10.2135/cropsci2008.02.0117
20. Casler MD, Vogel KP, Taliaferro CM, Ehlke NJ, Berdahl JD, Brummer EC et al (2007) Latitudinal and longitudinal adaptation of switchgrass populations. *Crop Sci* 47(6):2249–2260. doi:10.2135/cropsci2006.12.0780
21. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410. doi:10.1006/jmbi.1990.9999
22. Paterson AH, Bowers JE et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457(7229):551–556. doi:10.1038/Nature07723
23. Vogel JP, Garvin DF et al (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463(7282):763–768. doi:10.1038/Nature08747
24. Passardi F, Longet D, Penel C, Dunand C (2004) The class III peroxidase multigenic family in rice and its evolution in land plants. *Phytochemistry* 65(13):1879–1893
25. Vogel KP, Sarath G, Saathoff AJ, Mitchell RB (2010) Switchgrass. *Energy crops*. Royal Society of Chemistry, London. doi:10.1039/9781849732048-00275
26. Moore KJ, Moser LE, Vogel KP, Waller SS, Johnson BE, Pedersen JF (1991) Describing and quantifying growth stages of perennial forage grasses. *Agron J* 83(6):1073–1077
27. Jain M, Garg R, Patel RK, Tyagi AK (2011) De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Res* 18(1):53–63. doi:10.1093/dnares/dsq028
28. Logacheva MD, Kasianov AS, Vinogradov DV, Samigullin TH, Gelfand MS, Makeev VJ et al (2011) De novo sequencing and characterization of floral transcriptome in two species of buckwheat (*Fagopyrum*). *BMC Genomics* 12:30. doi:10.1186/1471-2164-12-30
29. Delannoy E, Jalloul A, Assigbetse K, Marmey P, Geiger JP, Lherminier J et al (2003) Activity of class III peroxidases in the defense of cotton to bacterial blight. *Mol Plant Microbe Interact* 16(11):1030–1038. doi:10.1094/MPMI.2003.16.11.1030
30. Jeong JH, Song HR, Ko JH, Jeong YM, Kwon YE, Seol JH et al (2009) Repression of FLOWERING LOCUS T chromatin by functionally redundant histone H3 lysine 4 demethylases in *Arabidopsis*. *PLoS One* 4(11):e8033. doi:10.1371/journal.pone.0080333
31. Lu F, Cui X, Zhang S, Jenuwein T, Cao X (2011) *Arabidopsis* REF6 is a histone H3 lysine 27 demethylase. *Nat Genet* 43(7):715–719. doi:10.1038/ng.854
32. Sarnowski TJ, Rios G, Jasik J, Swiezewski S, Kaczanowski S, Li Y et al (2005) SWI3 subunits of putative SWI/SNF chromatin-remodeling complexes play distinct roles during *Arabidopsis* development. *Plant Cell* 17(9):2454–2472. doi:10.1105/tpc.105.031203
33. Castillejo C, Pelaz S (2008) The balance between CONSTANS and TEMPRANILLO activities determines FT expression to trigger flowering. *Curr Biol* 18:1338–1343
34. Kosugi S, Ohashi Y (2002) E2Ls, E2F-like repressors of *Arabidopsis* that bind to E2F sites in a monomeric form. *J Biol Chem* 277:16553–16558
35. de Jong M, Wolters-Arts M, Garcia-Martinez JL, Mariani C, Vriezen WH (2011) The *Solanum lycopersicum* AUXIN RESPONSE FACTOR 7 (SlARF7) mediates cross-talk between auxin and gibberellin signalling during tomato fruit set and development. *J Exp Bot* 62:617–626
36. Menand B, Yi K, Jouannic S, Hoffmann L, Ryan E et al (2007) An ancient mechanism controls the development of cells with a rooting function in land plants. *Science* 316:1477–1480
37. Gremski K, Ditta G, Yanofsky MF (2007) The HECATE genes regulate female reproductive tract development in *Arabidopsis thaliana*. *Development* 134:3593–3601
38. Prigge MJ, Otsuga D, Alonso JM, Ecker JR, Drews GN et al (2005) Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in *Arabidopsis* development. *Plant Cell* 17:61–76
39. Walsh J, Waters CA, Freeling M (1998) The maize gene *liguleless2* encodes a basic leucine zipper protein involved in the establishment of the leaf blade–sheath boundary. *Genes Dev* 12:208–218
40. Finkelstein R, Gampala SS, Lynch TJ, Thomas TL, Rock CD (2005) Redundant and distinct functions of the ABA response loci ABA-INSENSITIVE (ABI) 5 and ABRE-BINDING FACTOR (ABF) 3. *Plant Mol Biol* 59:253–267
41. Ehlert A, Weltmeier F, Wang X, Mayer CS, Smeekens S et al (2006) Two-hybrid protein-protein interaction analysis in *Arabidopsis* protoplasts: establishment of a heterodimerization map of group C and group S bZIP transcription factors. *Plant J* 46:890–900
42. Vagujfalvi A, Aprile A, Miller A, Dubcovsky J, Delugu G et al (2005) The expression of several Cbf genes at the Fr-A2 locus is linked to frost resistance in wheat. *Mol Gen Genomics* 274:506–514
43. Mu J, Tan H, Zheng Q, Fu F, Liang Y et al (2008) LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in *Arabidopsis*. *Plant Physiol* 148:1042–1054
44. Sarnowski TJ, Rios G, Jasik J, Swiezewski S, Kaczanowski S et al (2005) SWI3 subunits of putative SWI/SNF chromatin-remodeling complexes play distinct roles during *Arabidopsis* development. *Plant Cell* 17:2454–2472
45. Yamanouchi U, Yano M, Lin H, Ashikari M, Yamada K (2002) A rice spotted leaf gene, *Spl7*, encodes a heat stress transcription factor protein. *Proc Natl Acad Sci U S A* 99:7530–7535
46. Kumar R, Kushalappa K, Godt D, Pidkowich MS, Pastorelli S et al (2007) The *Arabidopsis* BEL1-LIKE HOMEODOMAIN proteins SAW1 and SAW2 act redundantly to regulate KNOX expression spatially in leaf margins. *Plant Cell* 19:2719–2735
47. Jeong JH, Song HR, Ko JH, Jeong YM, Kwon YE et al (2009) Repression of FLOWERING LOCUS T chromatin by functionally redundant histone H3 lysine 4 demethylases in *Arabidopsis*. *PLoS One* 4:e8033
48. Lu F, Cui X, Zhang S, Jenuwein T, Cao X (2011) *Arabidopsis* REF6 is a histone H3 lysine 27 demethylase. *Nat Genet* 43:715–719
49. Rotman N, Durberry A, Wardle A, Yang WC, Chaboud A et al (2005) A novel class of MYB factors controls sperm-cell formation in plants. *Curr Biol* 15:244–248
50. Reyes JL, Chua NH (2007) ABA induction of miR159 controls transcript levels of two MYB factors during *Arabidopsis* seed germination. *Plant J* 49:592–606
51. Yoshiyama K, Conklin PA, Huefner ND, Britt AB (2009) Suppressor of gamma response 1 (SOG1) encodes a putative transcription factor governing multiple responses to DNA damage. *Proc Natl Acad Sci U S A* 106:12843–12848
52. Castaings L, Camargo A, Pocholle D, Gaudon V, Texier Y et al (2009) The nodule inception-like protein 7 modulates nitrate sensing and metabolism in *Arabidopsis*. *Plant J* 57:426–435
53. Asano T, Masuda D, Yasuda M, Nakashita H, Kudo T et al (2008) AtNFXL1, an *Arabidopsis* homologue of the human transcription factor NF-X1, functions as a negative regulator of the trichothecene phytoxin-induced defense response. *Plant J* 53:450–464
54. Doyle MR, Bizzell CM, Keller MR, Michaels SD, Song J et al (2005) HUA2 is required for the expression of floral repressors in *Arabidopsis thaliana*. *Plant J* 41:376–385
55. Eulgem T, Rushton PJ, Robatzek S, Somssich IE (2000) The WRKY superfamily of plant transcription factors. *Trends Plant Sci* 5:199–206