Bioenerg. Res.
DOI 10.1007/s12155-015-9674-2

Genome-Scale Identification of Cell-Wall-Related Genes in Switchgrass through Comparative Genomics and Computational Analyses of Transcriptomic Data

Xin Chen[1,2,3] · Qin Ma[2,3,6] · Xiaolan Rao[3,4] · Yuhong Tang[3,5] · Yan Wang[1] · Gaoyang Li[1,2] · Chi Zhang[2] · Xizeng Mao[2,7] · Richard A. Dixon[3,4] · Ying Xu[1,2,3,8]

Abstract Large numbers of plant cell-wall (CW)-related genes have been identified or predicted in several plant genomes such as *Arabidopsis thaliana*, *Oryza sativa* (rice), and *Zea mays* (maize), as results of intensive studies of these organisms in the past 2 decades. However, no such gene list has been identified in switchgrass (*Panicum virgatum*), a key bioenergy crop. Here, we present a computational study for prediction of CW genes in switchgrass using a two-step procedure: (i) homology mapping of all annotated CW genes in the fore-mentioned species to switchgrass, giving rise to a total of 991 genes, and (ii) candidate prediction of CW genes based on switchgrass genes co-expressed with the 991 genes under a large number of experimental conditions. Specifically, our co-expression analyses using the 991 genes as seeds led to the identification of 104 large clusters of co-expressed genes, each referred to as a co-expression module (CEM), covering 830 of the 991 genes plus 823 additional genes that are strongly co-expressed with some of the 104 CEMs. These 1653 genes represent our prediction of CW genes in switchgrass, 112 of which are homologous to predicted CW genes in *Arabidopsis*. Functional inference of these genes is conducted to derive the possible functional relations among these predicted CW genes. Overall, these data may offer a highly useful information source for cell-wall biologists of switchgrass as well as plants in general.

Keywords Switchgrass · Plant cell wall · Homology mapping · Co-expression analysis

Introduction

Substantial efforts have been invested in the past 2 decades into explorations of future fuels in the post-fossil fuel era [1–3]. Lignocellulosic biomass is considered one of the most promising next-generation energy sources [4], which can be converted into biofuels such as bioethanol through degradation of the cellulose fibers and hemicellulosic polymers into

Xin Chen and Qin Ma contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s12155-015-9674-2) contains supplementary material, which is available to authorized users.

✉ Ying Xu
xyn@bmb.uga.edu

1 College of Computer Science and Technology, and School of Public Health, Jilin University, Changchun, China

2 Computational Systems Biology Lab, Department of Biochemistry and Molecular Biology and Institute of Bioinformatics, University of Georgia, Athens, GA, USA

3 US Department of Energy, BioEnergy Science Center (BESC), Oak Ridge, TN 37831, USA

4 Department of Biological Sciences, University of North Texas, Denton, TX 76203, USA

5 Plant Biology Division, The Samuel Roberts Noble Foundation, Ardmore, OK 73401, USA

6 Present address: Department of Plant Science, South Dakota State University, Brookings, SD 57006, USA

7 Present address: Institute of Applied Cancer Center, MD Anderson Cancer Center, Houston, TX 77054, USA

8 A110 Life Science building, University of Georgia, Athens, GA 30602, USA

Published online: 14 September 2015

monosaccharides followed by fermentation. One major challenge in utilizing this form of biofuel lies in the high costs of releasing the component sugars of the (hemi)cellulosic fraction, primarily from secondary cell walls, an issue that is widely referred to as biomass recalcitrance [5]. Overcoming this issue represents a major scientific focus in the biofuel research domain.

One proposed solution to this challenging problem is through first understanding how plant (secondary) cell walls are synthesized at the molecular level and then possibly rewiring the synthesis processes to reduce biomass recalcitrance to a minimal level without affecting the major properties of a plant [6]. A first step toward accomplishing this very ambitious goal is to identify all the genes involved in or related to CW construction, remodeling, and degradation.

It has been estimated that ~10 % of the protein-encoding genes in a plant genome have functions related to CW [7], which fall into five large functional classes: (i) substrate generation, (ii) polysaccharide synthesis and glycosyl transfer, (iii) assembly, architecture, and growth, (iv) differentiation and secondary wall formation, and (v) signaling and response mechanisms (https://cellwall.genomics.purdue.edu). A number of plant species are considered as good sources for harvesting lignocellulose for biofuel production, such as poplar (*Populus trichocarpa*) and switchgrass (*Panicum virgatum*) [8–11]. Switchgrass can be found in most areas of the USA, Canada, and Mexico and, with its high adaptability and biomass production, has been selected as one of the major biofuel crops [12]. Genomic resources are being developed for this species, including EST and Affymetrix microarray data [13], a draft genome sequence from the Joint Genome Institute (http://www.phytozome.net/panicumvirgatum.php), and efficient genetic transformation systems for multiple genotypes [14]. However, as of today, no genome-scale lists of cell-wall genes have been published, in comparison with model plant species such as *Arabidopsis* [15, 16], *Oryza sativa* (rice) [17, 15], and *Zea mays* (maize) [16]. This is largely due to the reality that the genome of switchgrass has not been fully completed, and only limited genome-scale transcriptomic data are publicly available [13].

We present here a computational prediction of CW-related genes in switchgrass, along with a computational pipeline that was developed for making this prediction. Figure 1 outlines the functionalities of the pipeline, which consists of the following components: (1) homology mapping of known CW genes in *Arabidopsis*, maize, and rice [15] to switchgrass, (2) assignment of the mapped genes to the five aforementioned CW-associated functional categories, (3) biclustering analyses of the mapped genes based on similarities among their expression profiles for inference of functionally associated gene groups, under the assumption that co-expressed genes under multiple conditions tend to be functionally related [18] and that mapped genes not co-expressed with other mapped genes can therefore be removed from further consideration as candidate CW genes, (4) expansion of the predicted co-expression clusters to include additional genes that are co-expressed with some of the predicted CW gene clusters, giving rise to a collection of co-expressed gene modules (CEMs), (5) identification of transcription factors (TFs) within each CEM as candidate regulators for the transcription of the genes in the CEM, and (6) visualization of each co-expression gene module using Cytoscape [19]. The predicted CEMs are compared with annotated CW genes in *Arabidopsis* as a way to validate the prediction.

Overall, 991 switchgrass genes were predicted to be CW-related based on sequence homology, 830 of which also have strong support from gene-expression data and 741 of which have functional relationships with each other based on published data. Furthermore, 823 additional CW genes were predicted based on the similarities of their co-expression patterns with those of the 830 predicted CW genes. One hundred and twelve of these are homologous to predicted CW genes of *Arabidopsis*, hence providing additional supporting evidence for their prediction. We believe that these computational results offer a reliable list of CW genes in switchgrass, which can be used as an information source for experimental studies of cell wall synthesis and modification in this organism.

Materials and Methods

Data

The switchgrass transcriptomic data in our study was retrieved from the Switchgrass Gene Expression Atlas at the Samuel Roberts Noble Foundation [13], which contains genome-scale expression values of 122,973 probes, collected under 94 conditions. Of the 122,973 probes, 120,871 matched 92,686 Unique Transcript Sequences (UTSs) in the transcriptomic dataset of the switchgrass Atlas server (http://switchgrassgenomics.noble.org).

Eight hundred and forty-six, 982, and 716 CW-related proteins from *Arabidopsis*, rice and maize, respectively, were downloaded from the Purdue Cell-Wall-Genomics Database and used for homology mapping in this study. In addition, 56 HMM models for transcription factors (TFs) were downloaded from the PlantTFDB database V3.0 (release of 24/07/2013) [20] and used to predict TFs in switchgrass through scanning each HMM model against the whole genome of switchgrass. Throughout this paper, genes and probes are used interchangeably, and the mapping list between genes and probes is given in Supplementary File 1.

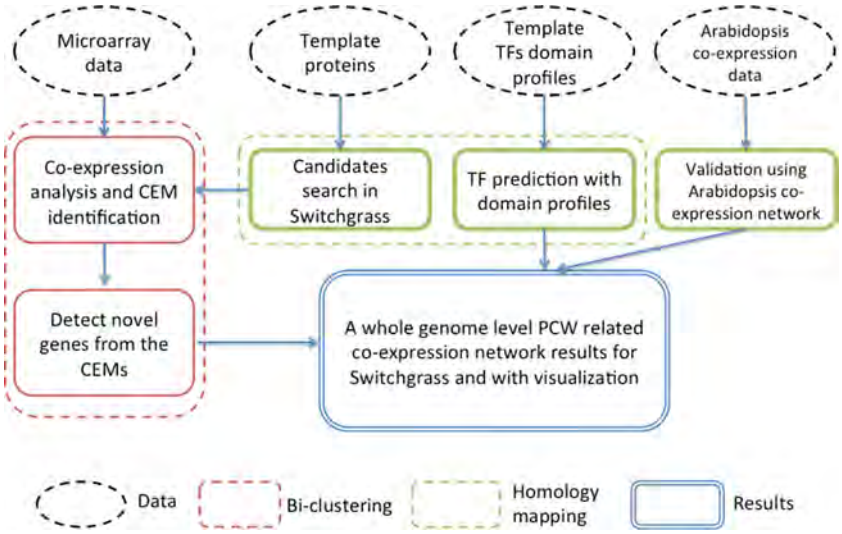

Fig. 1 A schematic of our prediction pipeline

Prediction of CW Genes and TFs Through Homology Mapping

Tblastn was used to homology-map each of the CW genes in the three template organisms to switchgrass, and only the best hits with Blast *e*-value<1e-30 and b-score>90 % were considered in our prediction of CW genes in switchgrass. A genome-scale prediction of TFs in switchgrass was conducted through scanning the 56 HMMs against the whole genome of switchgrass using HMMER3 with default parameters, which can be found in Supplementary File 1.

Biclustering Analyses of Switchgrass Genes Based on Similarities Among Their Gene-Expression Patterns

QUBIC is a biclustering analysis tool designed for co-expression analyses of genes based on their gene-expression patterns under multiple conditions. The software can generally identify all statistically significant groups, or biclusters, of genes with similar expression patterns under at least a specific number of experimental conditions, which tend to be more sensitive and more specific than other biclustering tools [21]. Using this program and default parameter values (i.e., k=5 for the minimum number of genes per cluster; f=30 % for the maximum overlap between two clusters; r=1 or −1 represents upregulation or downregulation), we carried out biclustering analyses needed in this study. In addition, the program has an option (−s) for identification of additional genes co-expressed with a specific bicluster under the defining conditions of the bicluster. We used this option to expand the initial biclusters to include additional genes that may be CW related. Specifically, for any given bicluster, "QUBIC –s" identifies outstanding expression patterns across a substantial fraction of all the genes in the bicluster under the defining conditions of the bicluster and then search for additional genes with expression profiles highly similar to the bicluster-based expression patterns under the same set of conditions.

Prediction of Cellular Functions of Predicted CW Genes

Switchgrass genes predicted to be CW related were mapped to *Arabidopsis* via Blastn with *e*-value<e-30. The mapped genes were then subjected to function prediction using a DAVID-based pathway enrichment analysis [22] against the GO and KEGG databases. All genes enriching the same functional category with an enrichment score>1 were considered to form a co-functional module (CFM), a threshold widely used in DAVID-based functional enrichment analyses.

Assessment of Significant Associations Between CEMs

We calculated the level of co-expression between two CEMs through a permutation-based test by using the correlation coefficients between co-expression patterns of each pair of genes in the two CEMs, averaged across the defining conditions of the CEMs. Specifically, this is calculated as follows:

Considering two CEMs $A=\{a_1,a_2,\ldots a_m\}$ and $B=\{b_1,b_2,\ldots b_n\}$ with m and n genes, respectively, the co-expression level (CS) between A and B is measured using

$$\mathrm{CS(A,\ B)} = \frac{\sum_{i=1..mj=1..n} cor(a_i, b_j)}{m*n},$$

where $cor(a_i,b_j)$ is the Pearson correlation coefficient between genes a_i and b_j. The calculation of CS(A, B) is done using the following procedure:

1. Randomly, sample n and m genes from the whole gene set of switchgrass without replacement to construct nominal CEMs A' and B'.

2. Calculate CS(A', B') using the above equation.
3. Repeat steps 1 and 2 for 1,000,000 rounds and then plot an empirical distribution for CS(A', B')
4. Estimate the *P* value for the actual CS(A, B) based on the above empirical distribution.

The estimated *P* values among all the CEM pairs are then corrected using the "Bonferroni" method [ref??] to control the false discovery rate, and the corrected *P* value <0.05 is used as the threshold for statistical significance for each pair of CEMs to be co-expressed. A histogram of a simulated empirical null distribution of the CS versus the identified CS among the CEMs is shown in Supplementary File 5.

Results

Homology Mapping of CW-Related Genes from *Arabidopsis*, Maize, and Rice to Switchgrass

Eight hundred and forty-six, 982, and 716 CW genes in *Arabidopsis*, rice, and maize were retrieved from the Purdue Cell-Wall-Genomics Database (https://cellwall.genomics.purdue.edu) and homology-mapped to the genome of switchgrass. These genes fell into five major functional groups: (i) substrate generation, (ii) polysaccharide synthesis and glycosyl transfer, (iii) assembly, architecture, and growth, (iv) differentiation and secondary wall formation, and (v) signaling and response mechanisms. Of these functional groups, (iii) has the largest number of CW genes. The details of these genes and their functional classes are found in Table 1.

Using Tblastn and these template CW genes (Fig. 2, step A), 629, 1015, and 637 distinct switchgrass genes were homology-mapped from *Arabidopsis*, maize, and rice, respectively. Of these, 222 genes were mapped from all three organisms, and 16, 66, and 323 switchgrass genes were mapped from only one of the three organisms. The detailed information is shown in Supplementary File 2. Overall, 991 distinct switchgrass genes were mapped from at least one CW gene in the three organisms. The number of genes mapped from maize is consistently higher than from the other two organisms across all five functional classes (Fig. 3), which is expected since maize is evolutionarily the closest to switchgrass among the three.

Table 1 Functional classes of CW-related proteins in *Arabidopsis*, maize, and rice

Group	Function class	*Arabidopsis*	Maize	Rice
i	Substrate generation	117	140	120
ii	Polysaccharide synthases and glycosyl transferase	140	247	225
iii	Assembly, architecture and growth	409	325	327
iv	Differentiation and secondary wall formation	91	154	27
v	Signaling and response mechanisms	89	116	17
vi[a]	Secretion and targeting pathways	0	0	0

[a] This group was not included in our study due to the lack of CW-related genes in the Purdue Cell-Wall-Genomics database

Identification of CEMs Through Co-expression Analyses

We expect that a true CW gene should be functionally related to other CW genes [18]. Based on this premise, we have used co-expressions with other predicted CW genes under multiple conditions as a criterion for predicting a homology-mapped gene, derived from the previous section, as a CW gene.

A biclustering analysis was applied to the 991 mapped genes, using our in-house program QUBIC [21], as shown in Fig. 2, to detect co-expressed gene clusters under a substantial subset of 94 conditions that were previously collected. The gene-expression data used here consist of samples collected from different tissue types, including root, stem internode, and shoot, under different growth conditions [13]. Specifically, each bicluster consists of at least 10 genes and five conditions with a *P* value at most 0.001.

One hundred and four biclusters were found to satisfy the above conditions, some of which may overlap with each other but none with more than 30 % overlap (Table 2). Overall, 830 of the 991 genes (84 %) were found in at least one bicluster, which are predicted to be CW-related genes in switchgrass. If we relax the above condition to *P* values <0.01, the biclusters will cover 920 out of the 991 genes (92.8 %). While we try to be conservative in our prediction, this information suggests the overall high reliability of our initial prediction of CW genes.

A functional analysis was conducted on the 830 genes, aiming to gain a pathway-level understanding of the functions of these genes. Specifically, DAVID [23] was used to carry out this analysis against the GO and KEGG pathway databases using the default parameters [24]. A total of 45 co-functional modules were found. Seven hundred and eighty-three of the 830 predicted CW genes (94.3 %) were covered by these CFMs. A detailed list of the covered genes is given in Supplementary File 3.

To further ensure those genes in the 104 identified CEMs are indeed CW related, we have examined the conditions under which some CEMs are detected. Specifically, eight suspension cell samples were treated with hormone combinations of brassinolide and 2,4-D, which all led to the production of extracellular and cell wall-associated lignin, suggesting that these conditions are related to lignin synthesis. Interestingly, all the conditions are included in the conditions associated with group (i) CEMs [25].

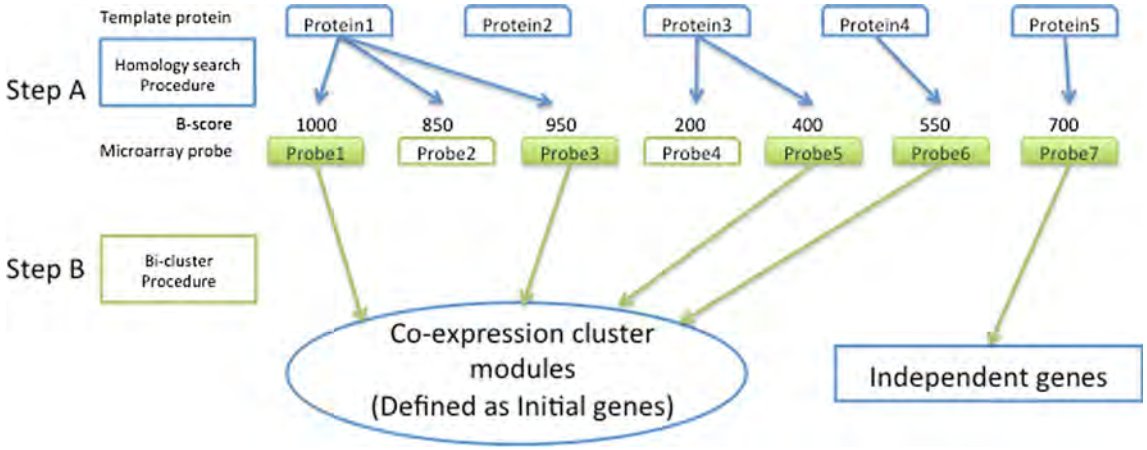

Fig. 2 A schematic diagram of homology mapping and co-expression analysis. **a** Homology mapping procedure. Tblastn was used to map template genes to the genome of switchgrass, using the following cutoffs: Tblastn *e*-value<1e-30, sequence identity>60 %, and *b*-score> 90 % of the best hits. The numbers above genes represent mapping *b*-scores between a template gene and a mapped gene. The *solid blocks* represent genes above the aforementioned cutoff values, and the hollow blocks are genes below the cutoff values and hence discarded in the next step. **b** QUBIC is used for biclustering analysis of the mapped genes to detect the co-expressed gene clusters

Prediction of Additional CW Genes Through Expanded Co-expression Analyses

We then conducted an expanded co-expression analysis to include all the switchgrass genes to identify additional genes that are strongly co-expressed with genes in the 104 predicted CEMs under the defining conditions of the 104 CEMs. The rationale is that genes having strong co-expressions with the 104 CEMs should be functionally related to genes in the 104 CEMs. For each CEM, we identified non-CEM genes having similar expression patterns to those of the genes in the CEM under the defining conditions of the CEM and then applied the same criteria as used for identifying the 104 CEMs to the new genes to each CEM. The method developed for accomplishing this is given in "Materials and Methods." Overall, 823 additional genes were predicted as CW genes in switchgrass, which are detailed in Table 3 and Supplementary File 3.

We have checked whether the 823 new predictions are homologous to any predicted CW genes in *Arabidopsis* [26], to provide additional supporting evidence for the predictions. Specifically, we compared these genes against all the predicted CW genes, totaling 3105 in *Arabidopsis* [26]. One hundred and twelve of the 823 genes have homologs among the 3105 genes. Hence, these 112 genes are considered as reliable predictions among the new additions. A statistical significance assessment, using a hypergeometric test, revealed that the *p* value for having 112 out of 823 genes with *Arabidopsis* CW homologs is <e-22, indicating that this result is highly significant. Table 3 shows a summary of these genes, and the detailed list of the 112 genes is given in Supplementary File 3.

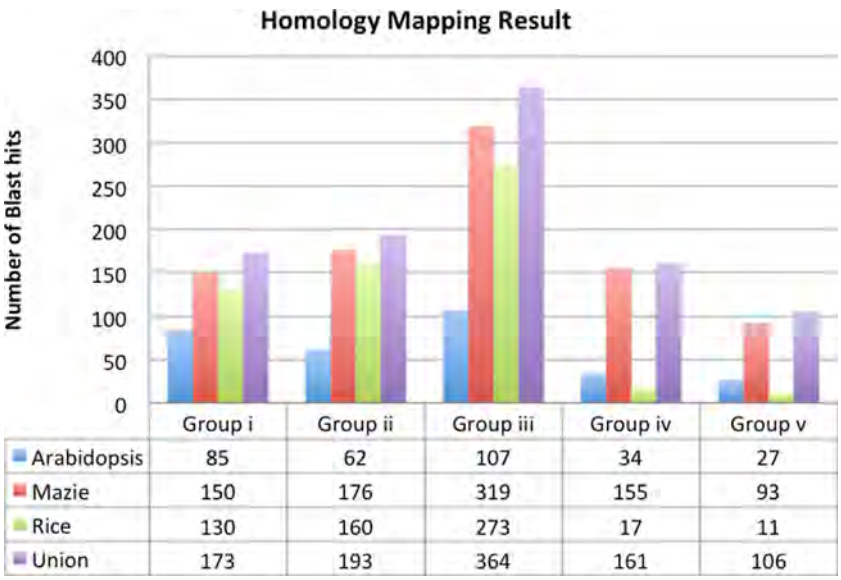

Fig. 3 Mapping results from CW genes of *Arabidopsis*, maize, and rice to switchgrass. Union represents the collection of all the unique switchgrass genes mapped from at least one of the three organisms in each functional group. Note: Group (iii) shared a total of six genes with groups (i) and (iv)

Table 2 Biclustering results and associated enriched CW pathways

Group	i	ii	iii	iv	v
#CEMs	21	38	26	13	6
#Genes	152	180	315	109	74
#CFMs	12	5	14	6	8
% of CEM genes coinciding with CFMs	98 %	86.1 %	95.6 %	100 %	93.2 %

#CEMs the number of co-expression modules in a specific functional group, *#Genes* the number of genes in at least one CEM, *#CFM* the number of co-functional modules, *% of CEM genes coinciding with CFMs* the percentage of genes in a CEM that also share the same CFM in a specific functional group

Among the five major functional groups, group (i) contains the highest number of the new CW genes, largely due to the presence of important cell wall-related transcription factors, such as MYB family genes [27, 28], in this group while group (iv) has the smallest number of the new CW genes.

Network-Level Functional Analyses and Predication of Transcriptional Regulators Within CEMs

Here, we aim to predict the main transcriptional regulators (TFs) for genes in each predicted CEM and the relationships among all the identified CEMs. To accomplish this, we have done a new round of biclustering analysis among all the 1653 predicted CW genes (830 initial and 823 additional), which consist of 22 TFs (see "Materials and Methods"). A total of 89 biclusters were identified across the 1653 genes with some of the original 104 CEMs being merged into larger biclusters bridged by some of the 823 new predictions.

The following observations were made regarding these CEMs: (i) 33 of the 89 (37 %) CEMs consist predominantly of the initial 830 genes mapped from the other plants, (ii) 40 of the 89 (45 %) CEMs consist of a substantial number of genes from both the initial 830 and the 823 new genes, suggesting that the newly predicted genes work closely with the initial genes, and (iii) the remaining 16 (18 %) CEMs tend to have fewer initial genes but more new ones, suggesting that the new genes in each of these CEMs may form a distinct pathway that can be extended from the initial genes. Figure 4 shows examples of each of these types, each represented as a graph with nodes representing genes and edges representing pairs of co-expressed genes.

We noted that 27 of the 89 (30 %) CEMs contain at least one TF per CEM. In addition, intra-modular connections (i.e., co-expressions) were assessed to identify the possible core genes using the WGCNA R package [29]. The result is summarized as follows: A total of 30 cores, i.e., subnetworks in CEMs with highly dense intra-connections (with density above a specific threshold), were detected in the 89 CEMs. Fifteen of these cores predominantly consist of the initial CW genes, and the remaining 15 consist of largely new CW genes as defined above. The detailed information of intra-connection within each CEM is given in Supplementary File 5.

We have also assessed inter-CEM connections across the 89 CEMs as shown in Fig. 5 (and Supplementary File 5). We note the following: (1) CEMs in the same functional group ((i)–(v)) do not necessarily have stronger functional relations compared with CEMs in different functional groups, which seems to be true across all five functional groups, indicating that CEMs in the same group do not necessarily work more closely with each other compared to CEMs in other groups, (2) group (v) has the highest density per CEM in terms of their connections with CEMs in other groups while group (ii) seems to have the lowest connection density between its CEMs and CEMs in other groups, and (3) there seem to be multiple parallel "pathways," each of which connects largely distinct CEMs in each of the five groups, particularly groups (i)–(iii), suggesting the possibility that there are multiple, largely nonoverlapping larger functional molecules, each consisting CEMs in each of functional groups. Clearly, this requires more detailed analyses of the hierarchical organization of the CEMs, which may lead to new knowledge and understanding about the synthesis and remodeling of plant cell walls in switchgrass.

Table 3 Predicted CW genes with supporting evidence

Group	i	ii	iii	iv	v
# new CW genes	313	233	136	35	101
# new genes homologous to CW genes in *Arabidopsis*	43	25	11	6	27

Discussion

Two types of information were used to predict CW genes in switchgrass: (i) homology mapping of annotated CW genes in other plant genomes to switchgrass and (ii) switchgrass genes co-expressed with genes in (i). Out of the 1653 predicted CW genes, 942 are considered as reliable predictions since they each have at least two pieces of independent evidence supporting the prediction: homology to known or predicted CW genes in other plants and co-expressed, hence functionally related, with other predicted CW genes in switchgrass. In addition, all the remaining 711 of the 1653 genes have not been reported to be CW related. These genes, along with their predicted co-expression and functional relationships, can serve as a basis for further experimental studies of CW-related genes in switchgrass.

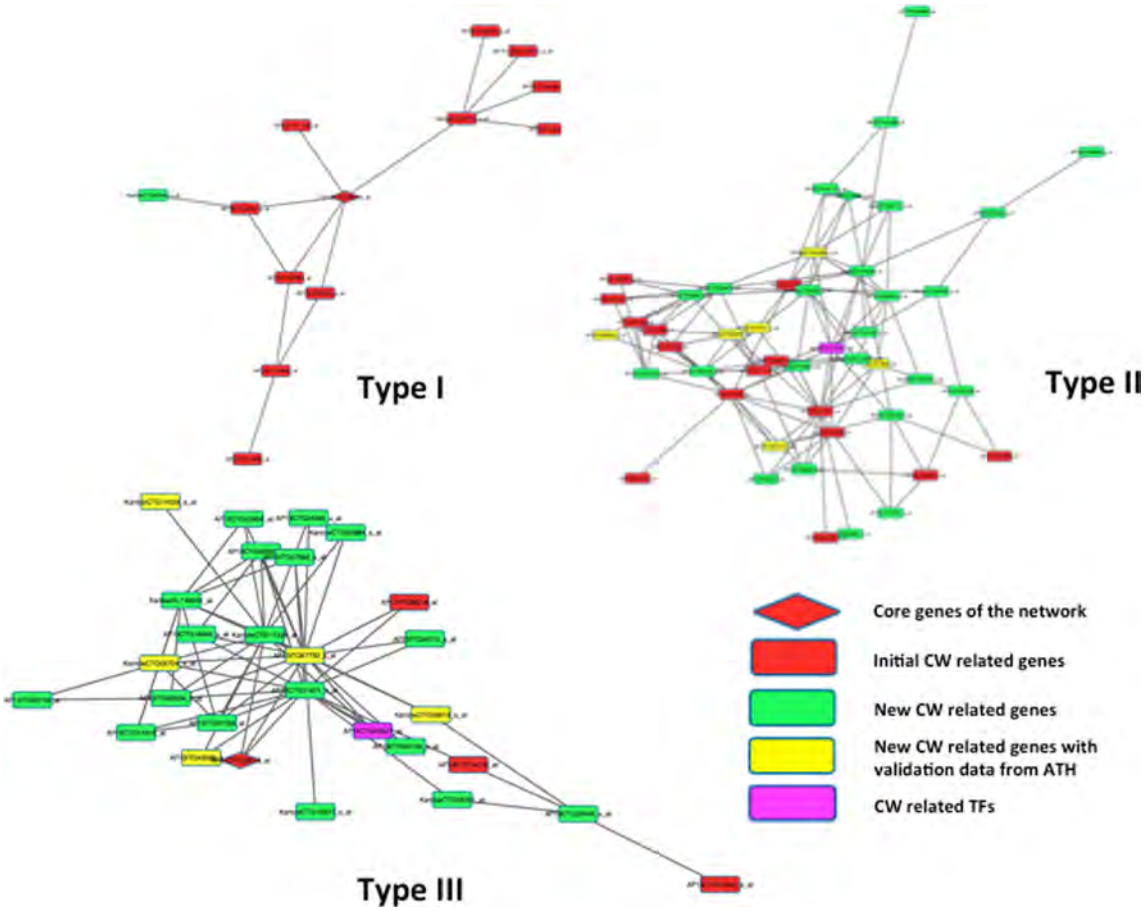

Fig. 4 Examples of three CEM types with purple nodes for TFs. (I) CEMs consisting of predominantly initial predictions of CW genes, (II) CEMs consisting of both types of genes (initial and new), and (III) CEMs consisting of predominantly new predictions of CW genes. The networks are drawn using Cytoscape (Supplementary File 5). Note that we consider only co-expressions in a network with Spearman correlation coefficient >0.6. The network modules are generated using the Cytoscape software [19]

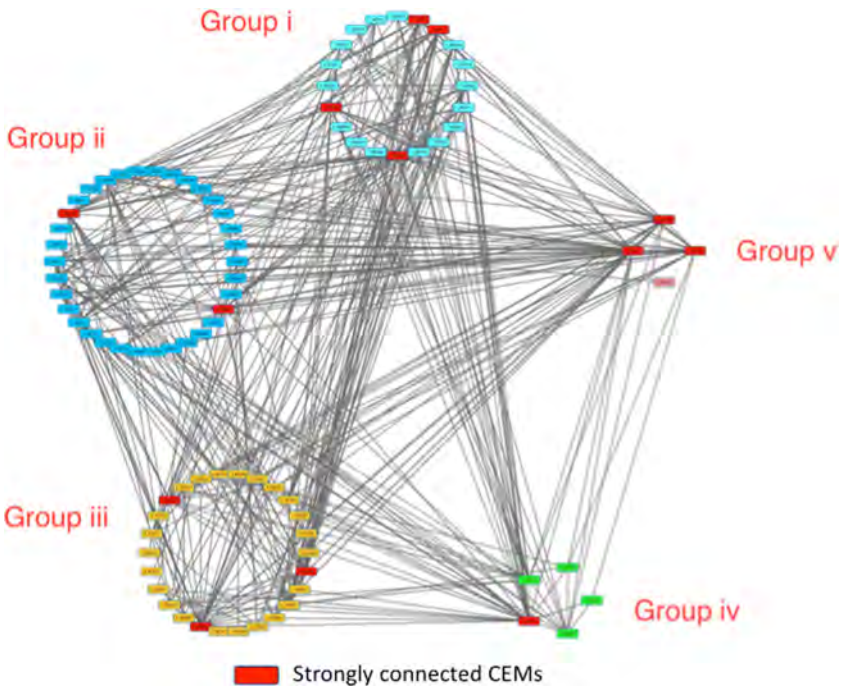

Fig. 5 Inter-CEM connections. Each node represents a CEM. Edges are defined based on the average co-expression correlation coefficients between two CEMs

For further work, we see a few ways to extend and refine our predictions. One is to increase the prediction reliability of the co-expressed gene modules by checking if the predicted genes tend to share common *cis* regulatory motifs that can potentially bind with the predicted TF(s), hence offering additional information in support of transcriptional co-regulation of the predicted co-expressed genes. A reliable approach for doing this is to generate genome-scale ChIP-seq data of the relevant TFs. Another area for improvement is to use a larger number of gene-expression datasets, particularly datasets collected under conditions specifically designed to study plant (secondary) cell walls, to refine our co-expression analyses for inference of CW genes. We expect that such data may become available once the switchgrass genome is better assembled. A third area is to conduct more detailed analyses of the large co-expression networks spanning all the 89 CW-related CEMs and five functional groups.

Acknowledgments This work was supported in part by the National Science Foundation (DEB-0830024 and DBI-0542119) and the DOE BioEnergy Science Center grant (DE-PS02-06ER64304), which is supported by the Office of Biological and Environmental Research in the Department of Energy Office of Science. This work was also supported in part by the Agriculture Experiment Station and the Biochemical Spatio-temporal Network Resource Center (3SP680) of South Dakota State University.

XC and QM participated in the coordination of the paper, carried out or participated all the analyses of transcriptomic data and the comparative genomics framework, and drafted the manuscript; XM participated in framework design. YT provided the transcriptomic data along with relevant data details, XR offered biology guidance in co-expression analysis, and YW and GL proved the TF prediction results. CZ designed the network analysis part. RAD reviewed and edited the paper and assisted in interpretation of data, and YX conceived the study, participated in its design and coordination, and revised the manuscript. All authors read and approved the final manuscript.

References

1. Pauly M, Keegstra K (2010) Plant cell wall polymers as precursors for biofuels. Curr Opin Plant Biol 13(3):305–312. doi:10.1016/j.pbi.2009.12.009
2. Ho DP, Ngo HH, Guo W (2014) A mini review on renewable sources for biofuel. Bioresour Technol. doi:10.1016/j.biortech.2014.07.022
3. Divakara BN, Upadhyaya HD, Wani SP, Gowda CLL (2010) Biology and genetic improvement of *Jatropha curcas* L.: a review. Appl Energy 87(3):732–742. doi:10.1016/j.apenergy.2009.07.013
4. Konda NM, Shi J, Singh S, Blanch HW, Simmons BA, Klein-Marcuschamer D (2014) Understanding cost drivers and economic potential of two variants of ionic liquid pretreatment for cellulosic biofuel production. Biotechnol Biofuels 7:86. doi:10.1186/1754-6834-7-86
5. Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW et al (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science 315(5813):804–807. doi:10.1126/science.1137016
6. Kalluri UC, Yin H, Yang X, Davison BH (2014) Systems and synthetic biology approaches to alter plant cell walls and reduce biomass recalcitrance. Plant Biotechnol J 12(9):1207–1216. doi:10.1111/pbi.12283
7. McCann M, Rose J (2010) Blueprints for building plant cell walls. Plant Physiol 153(2):365. doi:10.1104/pp. 110.900324
8. 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. doi:10.1073/pnas.0704767105
9. Wu M, Wu Y, Wang M (2006) Energy and emission benefits of alternative transportation liquid fuels derived from switchgrass: a fuel life cycle assessment. Biotechnol Prog 22(4):1012–1024. doi:10.1021/bp050371p
10. Karp A, Hanley SJ, Trybush SO, Macalpine W, Pei M, Shield I (2011) Genetic improvement of willow for bioenergy and biofuels. J Integr Plant Biol 53(2):151–165. doi:10.1111/j.1744-7909.2010.01015.x
11. Sannigrahi P, Ragauskas AJ, Tuskan GA (2010) Poplar as a feedstock for biofuels: a review of compositional characteristics. Biofuels Bioprod Biorefin 4(2):209–226
12. Parrish DJ, Fike JH (2009) Selecting, establishing, and managing switchgrass (Panicum virgatum) for biofuels. Methods Mol Biol 581:27–40. doi:10.1007/978-1-60761-214-8_2
13. Zhang JY, Lee YC, Torres-Jerez I, Wang M, Yin Y, Chou WC et al (2013) Development of an integrated transcript sequence database and a gene expression atlas for gene discovery and analysis in switchgrass (Panicum virgatum L.). Plant J 74(1):160–173. doi:10.1111/tpj.12104
14. Xi Y, Ge Y, Wang ZY (2009) Genetic transformation of switchgrass. Methods Mol Biol 581:53–59. doi:10.1007/978-1-60761-214-8_4
15. Carpita N, Tierney M, Campbell M (2001) Molecular biology of the plant cell wall: searching for the genes that define structure, architecture and dynamics. Plant Mol Biol 47(1–2):1–5
16. Yokoyama R, Nishitani K (2004) Genomic basis for cell-wall diversity in plants. A comparative approach to gene families in rice and *Arabidopsis*. Plant Cell Physiol 45(9):1111–1121. doi:10.1093/pcp/pch151
17. Penning BW, Hunter CT 3rd, Tayengwa R, Eveland AL, Dugard CK, Olek AT et al (2009) Genetic resources for maize cell wall biology. Plant Physiol 151(4):1703–1728. doi:10.1104/pp. 109.136804
18. Stuart JM, Segal E, Koller D, Kim SK (2003) A gene-coexpression network for global discovery of conserved genetic modules. Science 302(5643):249–255. doi:10.1126/science.1087447
19. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C et al (2007) Integration of biological networks and gene expression data using Cytoscape. Nat Protoc 2(10):2366–2382. doi:10.1038/nprot.2007.324
20. Jin J, Zhang H, Kong L, Gao G, Luo J (2014) PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. Nucleic Acids Res 42(Database issue):D1182–D1187. doi:10.1093/nar/gkt1016
21. Li G, Ma Q, Tang H, Paterson AH, Xu Y (2009) QUBIC: a qualitative biclustering algorithm for analyses of gene expression data. Nucleic Acids Res 37(15), e101. doi:10.1093/nar/gkp491
22. Jiao X, Sherman BT, da Huang W, Stephens R, Baseler MW, Lane HC et al (2012) DAVID-WS: a stateful web service to facilitate gene/protein list analysis. Bioinformatics 28(13):1805–1806. doi:10.1093/bioinformatics/bts251
23. da Huang W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4(1):44–57. doi:10.1038/nprot.2008.211
24. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M (2014) Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res 42(Database issue):D199–D205. doi:10.1093/nar/gkt1076
25. Shen H, Mazarei M, Hisano H, Escamilla-Trevino L, Fu C, Pu Y et al (2013) A genomics approach to deciphering lignin biosynthesis in

switchgrass. Plant Cell 25(11):4342–4361. doi:10.1105/tpc.113.118828

26. Wang S, Yin Y, Ma Q, Tang X, Hao D, Xu Y (2012) Genome-scale identification of cell-wall related genes in *Arabidopsis* based on co-expression network analysis. BMC Plant Biol 12:138. doi:10.1186/1471-2229-12-138
27. Zhong R, Ye ZH (2012) MYB46 and MYB83 bind to the SMRE sites and directly activate a suite of transcription factors and secondary wall biosynthetic genes. Plant Cell Physiol 53(2):368–380. doi:10.1093/pcp/pcr185
28. Law JA, Vashisht AA, Wohlschlegel JA, Jacobsen SE (2011) SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. PLoS Genet 7(7), e1002195. doi:10.1371/journal.pgen.1002195
29. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinforma 9:559. doi:10.1186/1471-2105-9-559