A genome-wide analysis of the *cellulose synthase-like* (*Csl*) gene family in maize (*Zea mays*)

Yongkai Li [1], **Xiaojie Cheng** [2], **Yaqin Fu** [1], **Qinqin Wu** [1], **Yuli Guo** [1], **Jiayu Peng** [1], **Wei Zhang** Corresp., [3], **Bin He** Corresp. [1]

[1] Jiangxi Science & Technology Normal University, Nanchang, Jiangxi, China
[2] Sichuan University, Chengdu, China
[3] Honghe University, Mengzi, China

Corresponding Authors: Wei Zhang, Bin He
Email address: zw_biology2@126.com, hebin.li@foxmail.com

Cell walls play an important role in the structure and morphology of plants as well as stress response, including various biotic and abiotic stresses. Although the comprehensive analysis of genes involved in cellulose synthase have been performed in model plants, such as *Arabidopsis thaliana* and rice, information regarding *cellulose synthase-like* (*Csl*) genes in maize is extremely limited. In this study, a total of 56 members of *Csl* gene family were identified in maize genome, which were classified into six subfamilies. Analysis of gene structure and conserved motif indicated functional similarities among the ZmCsl proteins within the same subfamily. Additionally, the 56 *ZmCsl* genes were dispersed on 10 chromosomes. The expression patterns of *ZmCsl* genes in different tissues using the transcriptome data revealed that most of *ZmCsl* genes had a relatively high expression in root and tassel tissues. Moreover, the expression profiles of *ZmCsl* genes under drought and re-watering indicated that the expression of *ZmCsl* genes were mainly responsive to early stage of drought stress. The protein-protein interaction network of *ZmCsl* genes proposed some potential interacted proteins. The data presented a comprehensive survey of *Csl* gene family in maize. The detailed description of maize *Csl* genes will be beneficial to understand their structural, functional, and evolutionary features. Importantly, we have described the differential expression profiles of these members across different tissues and under drought. This information will provide an important foundation for studying the roles of these *ZmCsl* genes in response to biotic and abiotic stresses.

A genome-wide analysis of the *cellulose synthase-like* (*Csl*) gene family in maize (*Zea mays*)

Yongkai Li[1#], Xiaojie Cheng[2#], Yaqin Fu[1], Qinqin Wu[1], Yuli Guo[1], Jiayu Peng[1], Wei Zhang[3]* and Bin He[1]*

[1] Jiangxi Key Laboratory of Bioprocess Engineering and Co-Innovation Center for In-vitro Diagnostic Reagents and Devices of Jiangxi Province, College of Life Sciences, Jiangxi Science & Technology Normal University, Nanchang 330013, China

[2] Key Laboratory of Bio-resources and Eco-environment, Ministry of Education, Sichuan Key Laboratory of Molecular Biology and Biotechnology, College of Life Sciences, Sichuan University, Chengdu 610064, China

[3] Honghe University，Mengzi 661100，China

[#]These authors contributed equally to this work

* Corresponding Author

Wei Zhang[3]*

Xuefu Road, Mengzi, Yunnan, 661100, China

Email address: zw_biology2@126.com

Bin He[1]*

Fenglin Road, Nanchang, Jiangxi, 330013, China

Email address: hebin.li@foxmail.com

Abstract

Cell walls play an important role in the structure and morphology of plants as well as stress response, including various biotic and abiotic stresses. Although the comprehensive analysis of genes involved in cellulose synthase have been performed in model plants, such as *Arabidopsis thaliana* and rice, information regarding *cellulose synthase-like* (*Csl*) genes in maize is extremely limited. In this study, a total of 56 members of *Csl* gene family were identified in maize genome, which were classified into six subfamilies. Analysis of gene structure and conserved motif indicated functional similarities among the ZmCsl proteins within the same subfamily. Additionally, the 56 *ZmCsl* genes were dispersed on 10 chromosomes. The expression patterns of *ZmCsl* genes in different tissues using the transcriptome data revealed that most of *ZmCsl* genes had a relatively high expression in root and tassel tissues. Moreover, the expression profiles of *ZmCsl* genes under drought and re-watering indicated that the expression of *ZmCsl* genes were mainly responsive to early stage of drought stress. The protein-protein interaction network of *ZmCsl* genes proposed some potential interacted proteins. The data presented a comprehensive survey of *Csl* gene family in maize. The detailed description of maize *Csl* genes will be beneficial to understand their structural, functional, and evolutionary features. Importantly, we have described the differential expression profiles of these members across different tissues and under drought. This information will provide an important foundation for studying the roles of these *ZmCsl* genes in response to biotic and abiotic stresses.

Keywords: cellulose synthase-like; *Zea mays*; genome-wide analysis; drought stress;

Introduction

The plant cell wall is essential for defining the shape of plant cells as well as providing structural integrity to plant tissues and organs. As the primary barrier, plant cell wall serves essential functions in resisting high cellular turgor pressures and protecting plants against environmental impacts (Kaur et al. 2017). Plant cell wall mainly consists of multiple polysaccharide fractions (Kaur et al. 2016). Among the various classes of polysaccharides, the major load-bearing component is cellulose. Cellulose is synthesized by cellulose synthase like (Csl) proteins, which belong to the superfamily of genes referred to as glycosyl transferase 2 (GT2). Due to the importance of the plant cell wall, the synthesis of cellulose and *Csl* genes have become the focus of significant research interest.

Csl genes are found throughout the plant kingdom. However, a variable number of *Csl* genes ranging from 30 to 50 have been reported from different plant species. The members of *Csl* gene family are classified into nine subfamilies: CslA, CslB, CslC, CslD, CslE, CslF, CslG, CslH and CslJ. Recently, a number of plant Csl genes have been identified, including rice (Hazen et al. 2002), poplar (Djerbi et al. 2005; Suzuki et al. 2006) and the moss *Physcomitrella patens* (Roberts & Bushoven 2007). CslJ is a new Csl family discovered in cereals (Yin et al. 2009), whereas CslB and CslG families are lacked in cereals generally (Yin et al. 2014). Among the nine subfamilies, CslA, CslC, and CslD are conserved in all land plants, whereas CslF and CslH families are restricted to grasses (Schreiber et al. 2014).

Several of the Csl subfamilies have been reported to be involved in the biosynthesis of different cell wall polysaccharides. For example, *CslA* and *CslC* genes have been identified as enzymes that encode the activity of 1,4-β-glucoside synthetase, and they are involved in the biosynthesis of mannan and xylan respectively (Cocuron et al. 2007; Liepman et al. 2005). Vega-Sánchez et al. have confirmed that rice *CslF6* gene mediates MLG (Mixed-linkage glucan) biosynthesis through over expression in tobacco, while Schreiber et al. reported that the CslF gene families in barley may be involved in (1,3;1,4)-β-glucan synthesis (Schreiber et al. 2014; Vega-Sánchez et al. 2012). In all Csl families, *CslD* has the oldest intron-exon structure and provide 1,4-β-linked glucan synthase activity, and is found in all sequenced terrestrial plant genomes, including *Physcomitrella* and *Selaginella*, indicating that the entire plant community has a highly conserved function (Douchkov et al. 2016; Verhertbruggen et al. 2011). However, CslD proteins have also been suggested to synthesize xylan, homogalacturonan, or mannan polysaccharides. Therefore, the true nature of Csl biochemical activity remains an open question.

Maize is one of the three major food crops with an ancient cultivation history, and has great potential for the

development of high nutrition, high biological functional food and biofuel (Wright et al. 2005). The synthesis of cellulose is of vital importance to protect maize against environmental impacts and provide raw materials for industrial applications. However, there are few reports on the related genes and mechanisms of cellulose synthesis in maize. Therefore, this study identify the Csl family members in the whole maize genome through large-scale data mining, providing a basis for the cloning of *Csl* gene family in maize, and analyzing its evolution and structure. Furthermore, the results lay a good foundation of further study on the function and the mechanism of cellulose synthesis.

Methods

Determine the Csl family members in maize genome

Because the naming of the Rice *Csl* genes is well documented, it is used as a reference sequence. All the sequences of rice *Csl* genes (*OsCsl*) were downloaded from the Rice Genome Annotation Project. The whole genome sequence of maize was downloaded from NCBI. To identify the *Csl* gene family members in maize, the genome sequence of maize was aligned with all rice *Csl* genes using NCBI-BLAST-2.7.1+ software. The E-value was set to 1E-5. Then the identity and cover region (more than 50%) were used as a filter criteria to eliminate improper *Csl* genes. These genes were further filtered through domain analysis to ensure that the selected sequences were non-redundant sequences to ultimately determine the candidate *Csl* gene family members. All candidate members should contain one of the two Pfam domain models (PF00535 and PF03552).

Multiple sequence alignment and the construction of phylogenetic tree

Multiple sequence alignment of rice *Csl* genes and *ZmCsl* gene family members was performed by the Muscle program with the default parameters. MEGA (v5.0.1.102) was applied to construct a maximum likelihood (ML) tree (Tamura et al. 2011). The phylogenetic tree was evaluated by bootstrap method, and the repeat value was set to 1000. The Figtree program (v1.4.3) was used to visualize it.

Analysis of conserved motif and gene structure

To identify the conserved motifs of each *ZmCsl* gene, deduced ZmCsl protein sequences were subjected to MEME version 4.12.0. The parameters were set as zero or one occurrence (of a contributing motif site) per sequence, the numbers of motif were chosen five motifs; the motif width was set 6 to 50 (Wu et al. 2016). The results are visualized by TBtools software. To further study the characteristics of *ZmCsl* genes, we used the GSDS2.0 software to analyze the structure and distribution of the intron and exons for each gene. The corresponding CDS sequences and genomic sequences were found in the NCBI database according to the accession number.

The location of genes on chromosomes

To determine the chromosomal locations of all *Csl* genes in maize, the information of locus coordinates were obtained from NCBI. The location and relative distance of *Csl* gene were displayed by MapChart (v3.2) software (Wen et al. 2016).

Expression analysis of *ZmCsl* during drought stress and different tissues

The genome-wide transcriptome data of maize in treatment with drought were obtained from NCBI SRA databases from accession SRR6665368 to SRR6665379. For drought treatment，when seedlings grow to 3-fully expanded leaf period, 20% polyethylene glycol (PEG) 6000 was added to the nutrient solution. In the 60h, 96h and recovery 3d after the stress treatment (denoted as T60, T96 and TR3d respectively), leaves were harvested and then immediately frozen in liquid nitrogen. Control seedlings were grown under the same conditions but without PEG treatment (denoted as CK60, CK96h and CKR3d respectively).

The transcriptome data of maize during different tissues have been submitted to NCBI SRA databases under accession SRP067440. The gene expression levels were estimated by the TopHat/Cufflinks pipeline described in the previous reports with FPKM (fragments per kilobase of exon per million fragments mapped) values. The heatmaps for expression profiles were generated with the OmicShare Tools (http://www.omicshare.com/tools/Home/Index/index.html).

Construction of protein–protein interaction network

Protein-protein interaction (PPI) data was obtained from the online database of STRING, which is an open source software for predicting and visualizing complex networks. These interactions were derived from literature of experimental validation, including physical interactions and enzymatic reactions found in signal transduction pathways. The PPI data were preprocessed, including removing redundancy and self-loops (Wenqing et al. 2016). Targets with a high confidence score >0.7 were selected to construct the PPI networks (Fu et al. 2016). PPI networks are visualized in Cytoscape with the nodes representing the proteins/genes and the edges representing interactions between any two proteins/genes.

Results

Identification of the *ZmCsl* gene family members

A total of 59 *Csl* genes were identified from the genome of maize using the Blast programs with the query sequences of rice *Csl* genes. Subsequently, Database searches for maize using conserved pfam PF00535 and PF03552, resulted in the identification of 56 *cellulose synthase-like* (*ZmCsl*) genes. The identified genes were

named following the nomenclature of rice, which shares synteny with maize (Table 1). The length of protein is ranged from 136 to 1514. All *ZmCsl* genes analyzed in this study possess the structures typical of the Csl family, including a glycosyl transferase domain (PF00535) or cellulose synthase domain (PF03552). Except the structures mentioned above, ZmCslD1, ZmCslD2-7, ZmCslD3 and ZmCslD4-3 have a RING/Ubox like zinc-binding domain (PF14570), while ZmCslD2-1, ZmCslD2-3, ZmCslD2-4 and ZmCslD2-6 have a Zinc-binding RING-finger domain (PF14569).

Phylogenetic analysis of the *Csl* gene family between rice and maize

To investigate the phylogenetic relationship of the *Csl* gene family in maize and rice (*Oryza sativa*), the amino acid sequences of 56 Csl members from maize and 34 from *Oryza sativa* were used to construct a phylogenetic tree. Based on the phylogenetic tree, it clearly shows that all the maize Csl were classified into the same corresponding categories in *O. sativa*, which include CslA, CslC, CslD, CslE, CslF and CslH subfamilies (Figure 1). There were 19 members from maize and 9 from *O. sativa* in the CslA subfamilies, which was the largest subfamily in the six subfamilies. In addition, there were 11 and 13 members from maize classified into CslC and CslD subfamily, only 2 members in CslH and CslE subfamily, while 10 ZmCsl grouped in the CslF subfamily. In the six clades, *Csl* genes have their clear orthologues in the genome of *O. sativa*, which suggests that these genes might be conserved for some specific functions in the two species. Furthermore, many *Csl* genes were grouped together, suggesting that these homologous genes may have derived from multiple duplications after the speciation of maize during the evolution.

Analysis of gene structure and conserved motif

Analysis of exon-intron structure in the members of gene family is beneficial to obtain further insight into their evolutionary trajectory. We used the GSDS software to compare the full-length cDNA with its corresponding genome sequence to detect the exon-intron structure of *ZmCsl* genes. The results indicated that the exon-intron structure of *ZmCs*l genes was conserved to a certain degree within subfamilies, which supported the evolutionary relationships among members of each clade. However, as shown in Figure 2, the number of introns in *ZmCsl* genes was highly variable among different clades, ranging from 1 to 13 introns. For example, *ZmCslF* genes only contained 1 or 2 introns, while the majority of *ZmCsl* genes in other subfamilies contained more introns. The large variation in structures of *ZmCsl* genes suggests that the maize genome has changed significantly during its long evolutionary history.

To obtain insights into the diversity of motif compositions in ZmCsl proteins, putative motifs were predicted using the MEME program. A total of 10 conserved motifs were identified. The relative location of these motifs within the protein is represented in Figure 3. The identified consensus sequence for the motifs is shown in Table 2. As expected, members who had similar motif compositions could be clustered into one class,

suggesting functional similarities among the ZmCsl proteins within the same subfamily. ZmCslA and ZmCslC members, which have a closer phylogenetic relationship, contained motif 1, motif 2, motif 3, motif4, motif 5, motif 9 and motif 10, while motif 6, motif 7, and motif 8 are concentrated in other ZmCsl members. The distribution of motifs in different subfamilies implied sources of functional differentiation in *ZmCsl* genes in the evolutionary processes. Additionally, the motif distribution further confirmed the accuracy of the phylogenetic relationship of *ZmCsl* genes.

The location of genes on chromosomes

The chromosomal distribution of the *ZmCsl* genes was visualized by MapChart program. As shown in Figure 4, the 56 *ZmCsl* genes were dispersed on 10 chromosomes. Chromosome 1 and 7 harboured the most (10 of 56) ZmCsl genes, whereas 3 *ZmCsl* genes were found on chromosome 8. In addition, there were 6 *ZmCsl* genes on chromosome 5 and 9.

The tissue-specific expression of *ZmCsl* genes

To investigate the expression patterns of *ZmCsl* genes in different tissues, we utilized publicly available transcriptome data to survey the transcription levels of the *ZmCsl* genes in the ear, embryo, endosperm, pollen, root and tassel. From the results, we observed that 20 *ZmCsl* genes had a relatively high expression in root tissues, mainly including *ZmCslD* and *ZmCslF* genes (Figure 5). Moreover, there were 17 *ZmCsl* genes that had very high transcript abundance in tassel, mainly including *ZmCslD*, *ZmCslA* and *ZmCslC* genes. Besides, 6 genes (*ZmCslA3-1*, *ZmCslA3-2*, *ZmCslA3-3*, *ZmCslC1-1*, *ZmCslD4-1* and *ZmCslD4-3*) and 2 genes (*ZmCslD5* and *ZmCslD9*) were highly expressed in the ear and pollen, respectively.

The analysis of *ZmCsl* genes in response to drought stress

In order to elucidate the roles of *ZmCsl* genes in response to drought stress, we analyzed their expression profiles under drought and re-watering. Significant more *ZmCsl* genes are up-regulated after drought treatment for 60 hours (Figure 6). For example, *ZmCslF6-1*, *ZmCslD1*, *ZmCslC2-1* and *ZmCslC2-2* have displayed maximum expression after devoid of water supply for 60 hours. Some genes were lightly up-regulated, such as *ZmCslD2-2* and *ZmCslD4-2*. However, when recovering the water supply, only *ZmCslD6, ZmCslC2-1* and *ZmCslC2-2* exhibited high transcript accumulation. The results revealed that the expression of *ZmCsl* genes were responsive to early drought stress.

Analysis of protein-protein interaction (PPI) network

To illustrate the molecular mechanisms of ZmCsl proteins, PPI network was constructed using the data from STRING database. As shown in Figure 7, four ZmCsl proteins, including ZmCslD2-4, ZmCslA4-1, ZmCslF6-1 and ZmCslF6-2, were included in the network. Furthermore, ZmCslD2-4, GRMZM2G467943_P01 and

GRMZM2G025016_P01 were located in the more important positions of network, suggesting those proteins play critical roles in the maintaining the whole protein interactions in the network. However, the function of GRMZM2G467943_P01 and GRMZM2G025016_P01 were unknown. The results in this study were beneficial to identify more important proteins and biological modules that interacted with ZmCsl proteins. Additionally, some proteins, such as Zm.21995 (eukaryotic translation initiation factor) and GRMZM5G821988_P01 (eukaryotic translation initiation factor), were interacted with ZmCslD2-4 as well. The detailed information of the proteins in the PPI network was listed in Table S1.

Disscussion

In recent years, genome-wide analysis of gene family has become an efficient approach to understand gene structure, function, and evolution. The *Csl* genes are reported to involve in cellular component and various biological processes in many plant species. However, the comprehensive analysis of *Csl* gene family has been limited in *Arabidopsis*, rice and wheat. In this study, we conducted a comprehensive analysis of the *ZmCsl* gene family, including identification of members, phylogenetic relationships, chromosomal organization, expression profiles in different tissues and under drought stress conditions. The results will assist in understanding the roles of these *ZmCsl* genes and their potential molecular mechanisms in response to drought stress.

A total of 56 putative *Csl* genes were identified in the maize genome, the number of which was distinct with *Arabidopsis thaliana* (36) and rice (34), the dicotyledonous and monocotyledonous plants, respectively. Comparison of the maize genome (2500Mb) with A. thaliana (125Mb) and rice (466Mb), maize had about 20 times and five times larger genome size than A. thaliana and rice (Inititiative 2000; Sasaki et al. 2002; Shukla et al. 2009). However, the number of *Csl* genes in these species is not proportional with the genome size. The reason may be that the maize genome has undergone significant gene loss since the duplication event. The results from Lai et al further confirmed the speculation, showing that, only a small proportion of genes were conserved as duplicate factors in orthologous intervals of maize, when comparing the two homoeologous regions of the maize genome and the single homoeologous regions of the sorghum and rice genomes (Lai et al. 2004; Messing et al. 2004). Besides, phylogenetic analysis divides the members of the *ZmCsl* gene family into six distinct groups, including CslA, CslC, CslD, CslE, CslF and CslH, which was largely similar to that found in rice, while the Csl genes in A. thaliana were divided into other six groups, including CslA, CslB, CslC, CslD, CslE, and CslG. The results suggested that *Csl* genes may have a potential functional diversity between dicot and monocot plants.

The role of *Csl* genes have been reported mainly in model plant, *A. thaliana* and rice. For example, by

destroying *CsLD5* expression in *Arabidopsis*, Bernal et al. found that the growth of stem and root was significantly reduced in *CslD5* knockout plants, and the activity of xylan and homologous glucuronate synthetase was also reduced, indicating the possible role of this gene and other *CslD* in cell wall biosynthesis (Bernal et al. 2010). Florence et al. observed glucomannan deficiency in *CslA* mutant *Arabidopsis*, indicating that CslA family codes for glucomannan synthetase (Goubet et al. 2010). In rice, there were continuous researches on the *Csl* gene family since the *Csl* gene family in rice has been clearly identified (Hazen et al. 2002). For example, Wang et al. found that *Csl* genes in rice play important roles on the potentially functional complement for cell wall synthesis using transcriptional profiling and co-expression analyses (Wang et al. 2010). However, limited researches of *Csl* genes focused on maize with the exception of *ZmCslD1*, which is integral to normal cell division, expansion, leaf development and controls organ size in maize (Hunter 2010; Li et al. 2018).

In this study, a systematic analysis of *ZmCsl* genes was performed, including the identification of members, phylogenetic analysis and gene structure analysis. Furthermore, we explored the expression evidence for all putative *ZmCsl* genes in different tissues (ear, embryo, endosperm, pollen, root and tassel) using the publicly available transcriptome data. *ZmCslD* and *ZmCslF* genes exhibited a relatively high expression in root tissues, consistent with the previous research, indicating that they could play a role in the development of the plant root (Hansey et al. 2010; Wang et al. 2001). Moreover, *Csl* genes have been shown to play crucial roles in plants in responding to various biotic and abiotic stresses. In *Arabidopsis*, *AtCslD5* is not essential for the normal growth and development but play a critical role in osmotic stress tolerance which may be likely involved in the regulation of ROS under stress (Zhu et al. 2010). In another study of *Arabidopsis*, a cellulose synthase gene was reported to enhance drought and osmotic stress tolerance through regulating cellulose synthesis (Chen et al. 2010). In barley (*Hordeum vulgare*), *HvCslD2* represent an important defence reaction both for nonhost and for quantitative host resistance against nonadapted wheat and host-adapted barley powdery mildew pathogens, respectively, through mediating cell wall changes in the epidermal layer (Douchkov et al. 2016). In the present study, significant more *ZmCsl* genes are up-regulated after drought treatment for 60 hours, suggesting that these *Csl* genes may play positive roles in response to early drought stress. However, the specific mechanism of action needs to be further demonstrated and analyzed experimentally.

Regardless, in this study, we provided a comprehensive analysis of *ZmCsl* genes in the maize genome. The 56 *ZmCsl* genes were classified into six subfamilies, and the structural and functional properties of each *ZmCsl* genes were characterised. Most of *ZmCsl* genes had a relatively high expression in root tissues and up-regulated after drought treatment for 60 hours. The results will provide an important foundation for studying the roles of these *ZmCsl* genes in response to biotic and abiotic stresses, and be beneficial to understand their

potential interactions with defence-related genes in the regulation network.

Conclusion

In conclusion, this study provides a systematic and comprehensive analysis of the *cellulose synthase-like* (*Csl*) gene family in maize. In total, 56 members of *Csl* gene family were identified in maize genome, which were classified into six subfamilies, including include CslA, CslC, CslD, CslE, CslF and CslH subfamilies. The expression profile analyses of *ZmCsl* genes in different tissues were performed to reveal that most of *ZmCsl* genes had a relatively high expression in root and tassel tissues. Moreover, the expression patterns of *ZmCsl* genes under drought and re-watering indicated that the expression of *ZmCsl* genes were mainly responsive to early stage of drought stress. The detailed description of maize *Csl* genes will be beneficial to understand their structural, functional, and evolutionary features. This information will provide an important foundation for studying the roles of these *ZmCsl* genes in response to biotic and abiotic stresses.

Acknowledgements

We thank Yayi Tu and Min Sun for their suggestion in paper writting.

References

Bernal AJ, Jensen JK, Harholt J, Sørensen S, Moller I, Blaukopf C, Bo J, Lotto RD, Pauly M, and Scheller HV. 2010. Disruption of ATCSLD5 results in reduced growth, reduced xylan and homogalacturonan synthase activity and altered xylan occurrence in Arabidopsis. *Plant Journal* 52:791-802.

Chen Z, Hong X, Zhang H, Wang Y, Li X, and Gong Z. 2010. Disruption of the cellulose synthase gene, AtCesA8/IRX1, enhances drought and osmotic stress tolerance in Arabidopsis. *Plant Journal* 43:273-283.

Cocuron JC, Lerouxel O, Drakakaki G, Alonso AP, Liepman AH, Keegstra K, Raikhel N, and Wilkerson CG. 2007. A gene from the cellulose synthase-like C family encodes a beta-1,4 glucan synthase. *Proceedings of the National Academy of Sciences of the United States of America* 104:8550-8555.

Djerbi S, Lindskog M, Arvestad L, Sterky F, and Teeri TT. 2005. The genome sequence of black cottonwood (Populus trichocarpa) reveals 18 conserved cellulose synthase (CesA) genes. *Planta* 221:739-746.

Douchkov D, Lueck S, Hensel G, Kumlehn J, Rajaraman J, Johrde A, Doblin MS, Beahan CT, Kopischke M, and Fuchs R. 2016. The barley (Hordeum vulgare) cellulose synthase-like D2 gene (HvCslD2) mediates penetration resistance to host-adapted and nonhost isolates of the powdery mildew fungus. *New Phytologist* 212:421-433.

Fu Q, Huang Y, Wang Z, Chen F, Huang D, Lu Y, Liang X, and Zhang M. 2016. Proteome Profile and Quantitative Proteomic Analysis of Buffalo (Bubalusbubalis) Follicular Fluid during Follicle Development. *International Journal of Molecular Sciences* 17:618.

Goubet F, Barton CJ, Mortimer JC, Yu X, Zhang Z, Miles GP, Richens J, Liepman AH, Seffen K, and Dupree P. 2010. Cell wall glucomannan in Arabidopsis is synthesised by CSLA glycosyltransferases, and influences the progression of embryogenesis. *Plant Journal* 60:527-538.

Hansey CN, Lorenz AJ, and Leon ND. 2010. Cell Wall Composition and Ruminant Digestibility of Various Maize Tissues Across Development. *Bioenergy Research* 3:28-37.

Hazen SP, Scottcraig JS, and Walton JD. 2002. Cellulose synthase-like genes of rice. *Plant Physiology* 128:336-340.

Hunter CT, Iii. 2010. The Mu transportation of Zea Mays and their use in determining gene function: Cellulose

synthase-like D genes in plant and cell development. *Dissertations & Theses - Gradworks*.
Inititiative TAG. 2000. Analysis of the genome sequence of Arabidopsis thaliana. *Nature* 408:796-815.
Kaur S, Dhugga KS, Beech R, and Singh J. 2017. Genome-wide analysis of the cellulose synthase-like (Csl) gene family in bread wheat (Triticum aestivum L.). *Bmc Plant Biology* 17:193.
Kaur S, Dhugga KS, Gill K, and Singh J. 2016. Novel structural and functional motifs in cellulose synthase (CesA) genes of bread wheat (Triticum aestivum, L.). *Plos One* 11:e0147046.
Lai J, Ma J, Swigonová Z, Ramakrishna W, Linton E, Llaca V, Tanyolac B, Park YJ, Jeong OY, and Bennetzen JL. 2004. Gene loss and movement in the maize genome. *Genome Research* 14:1924-1931.
Li W, Yang Z, Yao J, Li J, Song W, and Yang X. 2018. Cellulose synthase-like D1 controls organ size in maize. *Bmc Plant Biology* 18:239. 10.1186/s12870-018-1453-8
Liepman AH, Wilkerson CG, Keegstra K, and Kende HJ. 2005. Expression of Cellulose Synthase-Like (Csl) Genes in Insect Cells Reveals That CslA Family Members Encode Mannan Synthases. *Proceedings of the National Academy of Sciences of the United States of America* 102:2221-2226.
Messing J, Bharti AK, Karlowski WM, Gundlach H, Kim HR, Yu Y, Wei F, Fuks G, Soderlund CA, and Mayer KF. 2004. Sequence composition and genome organization of maize. *Proceedings of the National Academy of Sciences of the United States of America* 101:14349-14354.
Roberts AW, and Bushoven JT. 2007. The cellulose synthase (CESA) gene superfamily of the moss Physcomitrella patens. *Plant Molecular Biology* 63:207-219.
Sasaki T, Matsumoto T, Yamamoto K, Sakata K, Baba T, Katayose Y, Wu J, Niimura Y, Cheng Z, and Nagamura Y. 2002. The genome sequence and structure of rice chromosome 1. *Nature* 420:312-316.
Schreiber M, Wright F, Mackenzie K, Hedley PE, Schwerdt JG, Little A, Burton RA, Fincher GB, Marshall D, and Waugh R. 2014. The Barley Genome Sequence Assembly Reveals Three Additional Members of the CslF (1,3;1,4)-β-Glucan Synthase Gene Family. *Plos One* 9:e90888.
Shukla VK, Doyon Y, Miller JC, Dekelver RC, Moehle EA, Worden SE, Mitchell JC, Arnold NL, Gopalan S, and Meng X. 2009. Precise genome modification in the crop species Zea mays using zinc-finger nucleases. *Nature* 459:437-441.
Suzuki S, Li L, Sun YH, and Chiang VL. 2006. The cellulose synthase gene superfamily and biochemical functions of xylem-specific cellulose synthase-like genes in Populus trichocarpa. *Plant Physiology* 142:1233-1245.
Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology & Evolution* 28:2731.
Vega-Sánchez ME, Verhertbruggen Y, Christensen U, Chen X, Sharma V, Varanasi P, Jobling SA, Talbot M, White RG, and Joo M. 2012. Loss of Cellulose synthase-like F6 function affects mixed-linkage glucan deposition, cell wall mechanical properties, and defense responses in vegetative tissues of rice. *Plant Physiology* 159:56-69.
Verhertbruggen Y, Yin L, Oikawa A, and Scheller HV. 2011. Mannan synthase activity in the CSLD family. *Plant Signaling & Behavior* 6:1620-1623.
Wang L, Guo K, Li Y, Tu Y, Hu H, Wang B, Cui X, and Peng L. 2010. Expression profiling and integrative analysis of the CESA/CSL superfamily in rice. *Bmc Plant Biology* 10:282.
Wang X, Cnops G, Vanderhaeghen R, and Block SD. 2001. AtCSLD3, a cellulose synthase-like gene important for root hair growth in arabidopsis. *Plant Physiology* 126:575-586.
Wen C, Cheng Q, Zhao L, Mao A, Yang J, Yu S, Weng Y, and Xu Y. 2016. Identification and characterisation of Dof transcription factors in the cucumber genome. *Scientific Reports* 6:23072.
Wenqing N, Diane T, Lu J, Zhang K, Wu H, Fu Y, Wang Y, Ou Z, Shan L, and Ding Y. 2016. Identification of novel genes and pathways in carotid atheroma using integrated bioinformatic methods. *Scientific Reports* 6:18764.
Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, Mcmullen MD, and Gaut BS. 2005. The effects of artificial selection on the maize genome. *Science* 308:1310-1314.
Wu Z, Cheng J, Cui J, Xu X, Liang G, Luo X, Chen X, Tang X, Hu K, and Cheng Q. 2016. Genome-Wide Identification and Expression Profile of Dof Transcription Factor Gene Family in Pepper (Capsicum annuumL.). *Frontiers in Plant Science* 7.
Yin Y, Huang J, and Xu Y. 2009. The cellulose synthase superfamily in fully sequenced plants and algae. *Bmc Plant Biology* 9:99.
Yin Y, Johns MA, Cao H, and Rupani M. 2014. A survey of plant and algal genomes and transcriptomes reveals

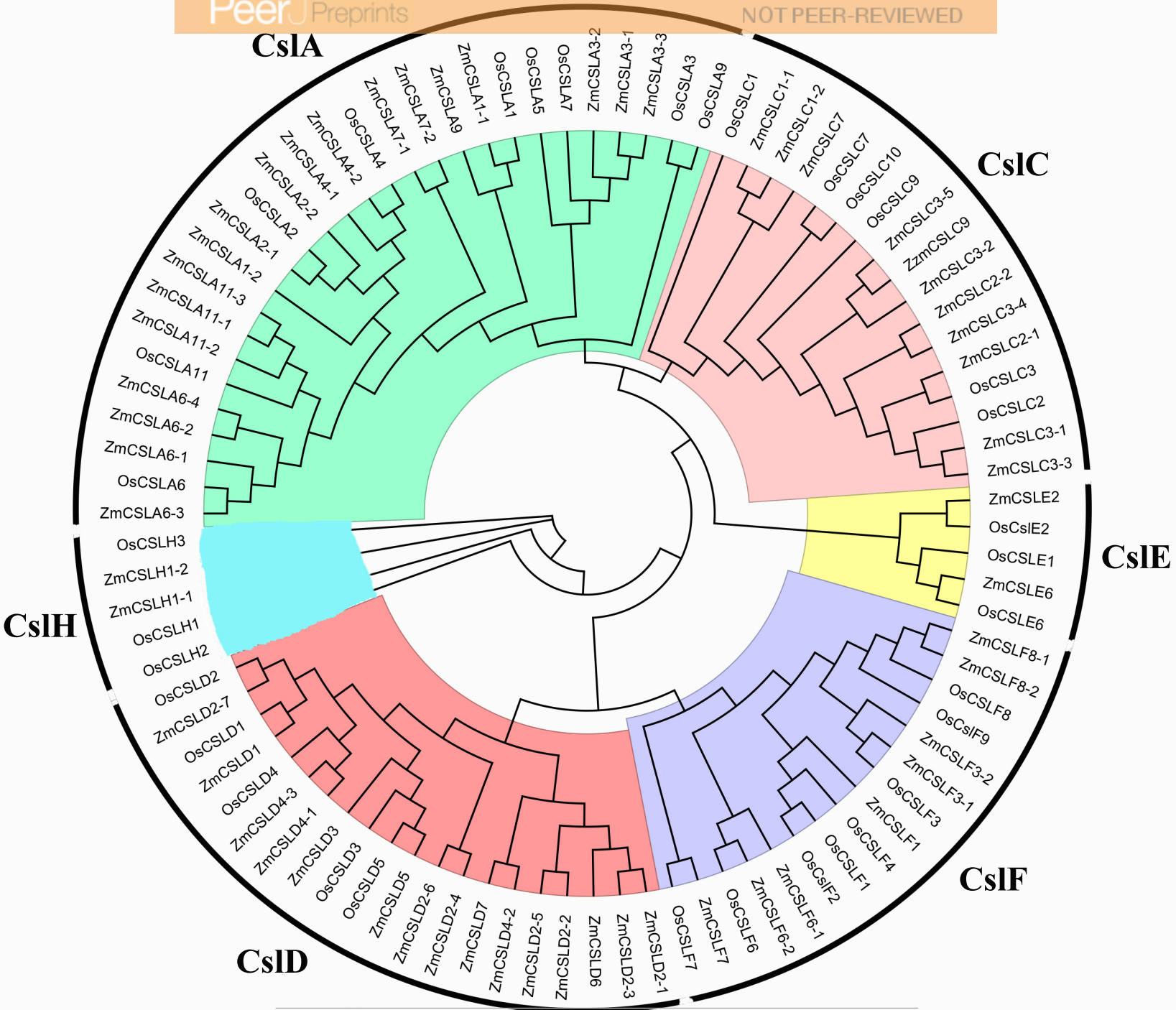
new insights into the evolution and function of the cellulose synthase superfamily. *BMC Genomics,15,1(2014-04-04)* 15:260.

Zhu J, Lee BH, Dellinger M, Cui X, Zhang C, Wu S, Nothnagel EA, and Zhu JK. 2010. A cellulose synthase-like protein is required for osmotic stress tolerance in Arabidopsis. *Plant Journal* 63:128-140.

Figure 1(on next page)

Phylogenetic relationship of rice and maize *Csl* genes

Maximum likelihood tree was constructed by MEGA program with 1000 bootstrap sampling using full-length sequences of 56 maize and 34 rice Csl proteins.

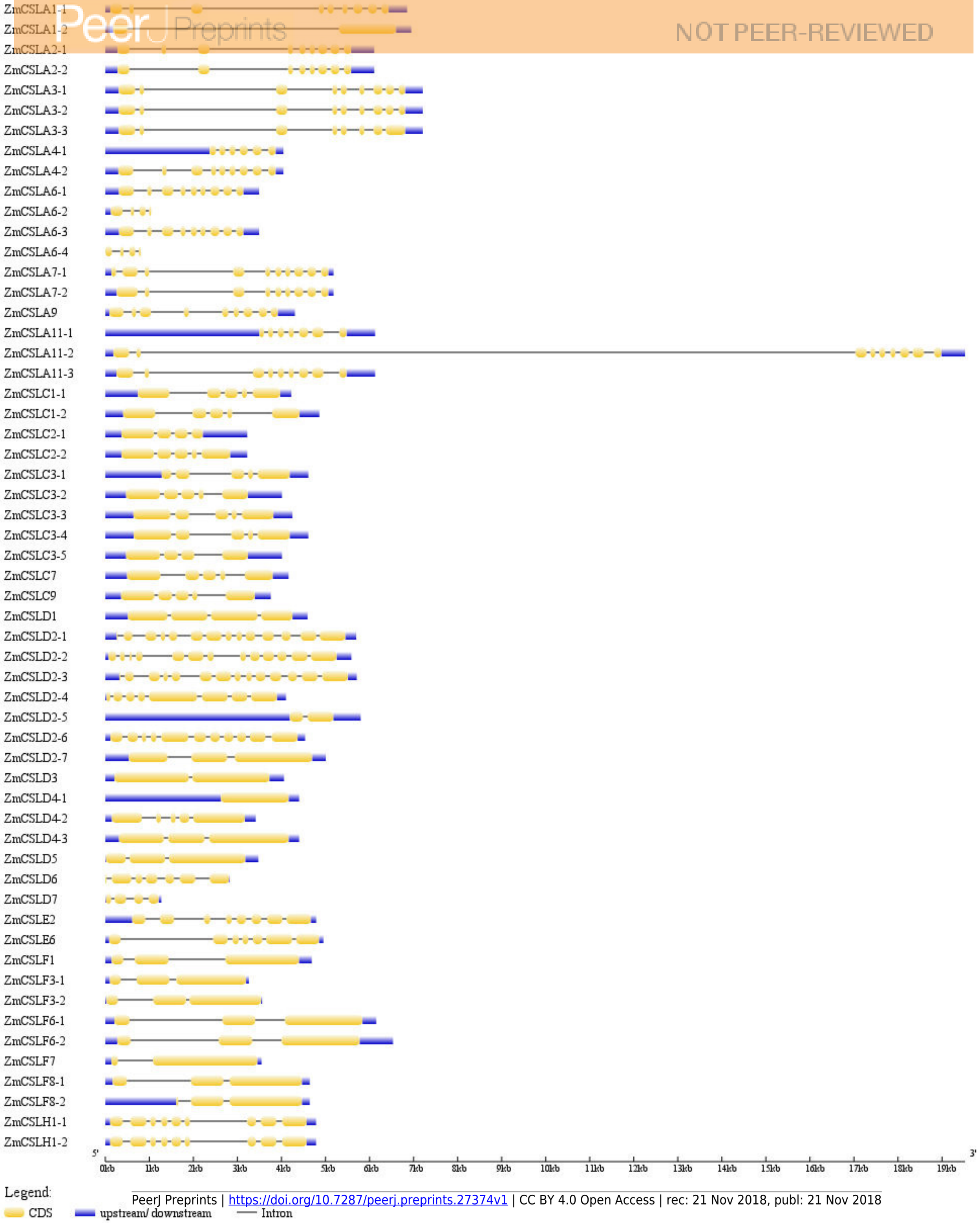


PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27374v1 | CC BY 4.0 Open Access | rec: 21 Nov 2018, publ: 21 Nov 2018

Figure 2(on next page)

The exon-intron structure of *Csl* genes in maize.

The yellow boxes indicate the exons while the single lines indicate introns. UTRs are displayed by thick purple lines at both ends. Gene models were drawn to scale as indicated at the bottom.

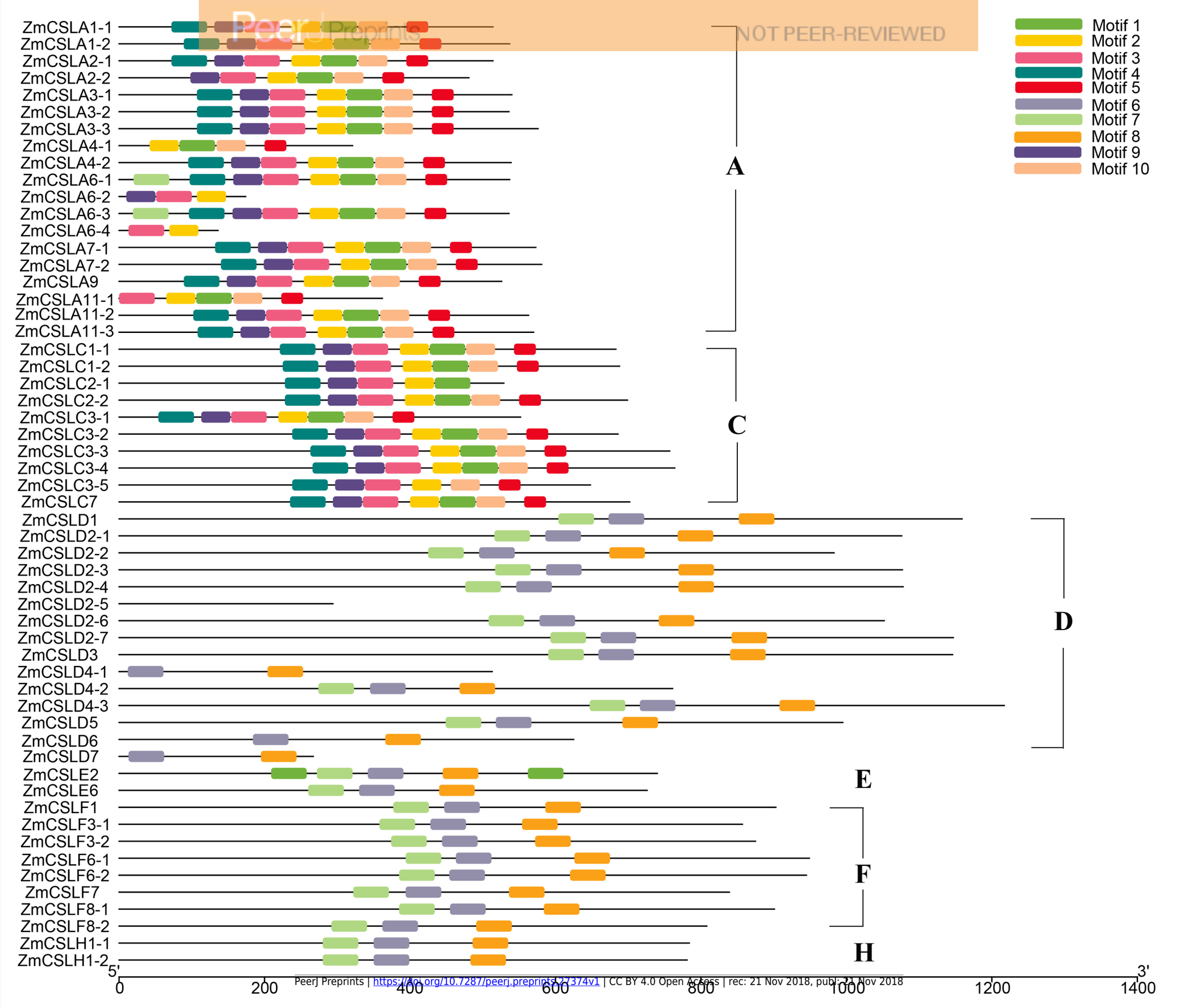


PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27374v1 | CC BY 4.0 Open Access | rec: 21 Nov 2018, publ: 21 Nov 2018

Figure 3(on next page)

Conserved motifs of *Csl* gene family in maize.

All conserved motifs of the ZmCsl proteins were identified by the MEME program. Protein sequences are indicated by thick gray lines, and the conserved motifs are represented by different colored boxes. The length (amino acids) of the protein and motif can be estimated using the scale bar at the bottom.



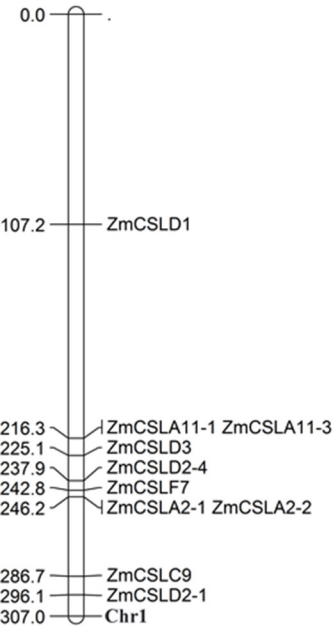
PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27374v1 | CC BY 4.0 Open Access | rec: 21 Nov 2018, publ: 21 Nov 2018

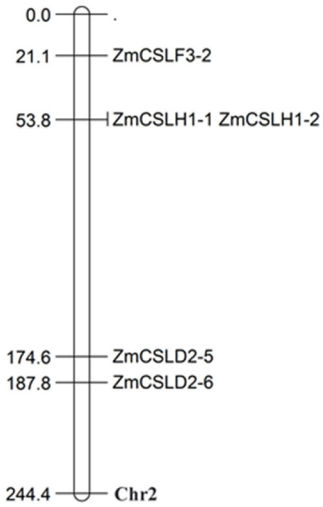
Figure 4(on next page)

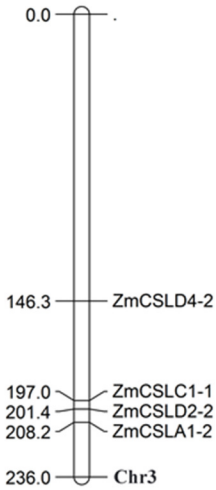
Chromosomal locations of *Csl* genes in maize genome.

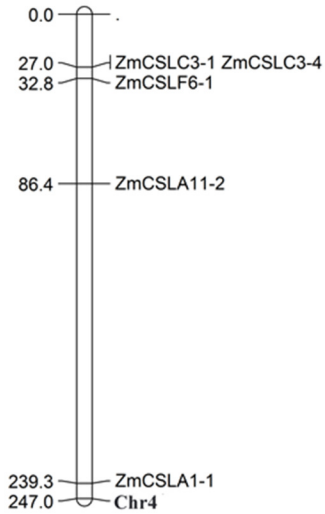
Chromosomes 1-10 are depicted as bars. *ZmCsl* genes are indicated by vertical black lines.

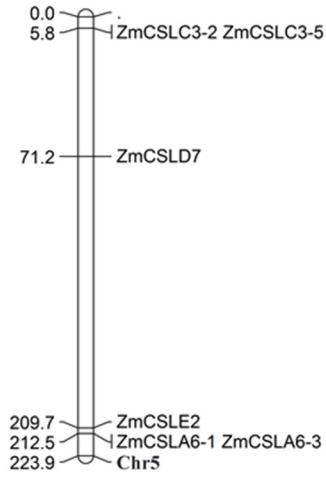
PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27374v1 | CC BY 4.0 Open Access | rec: 21 Nov 2018, publ: 21 Nov 2018

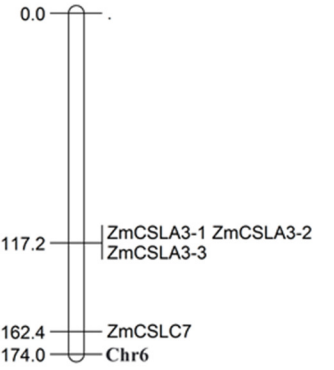


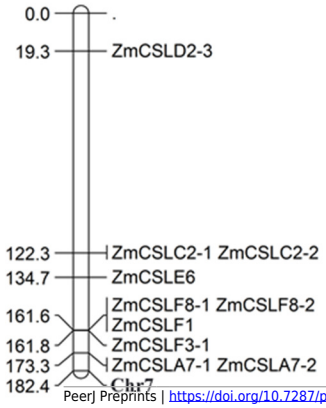


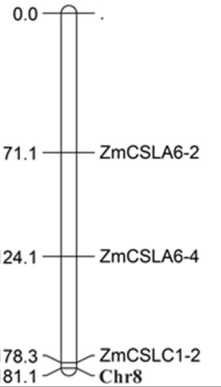


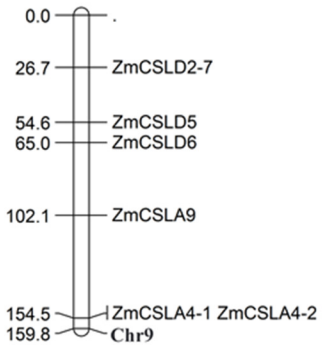












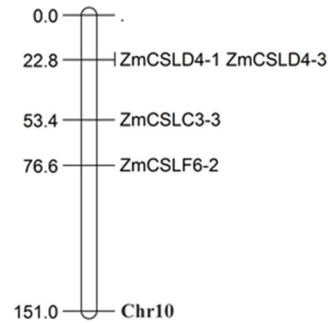
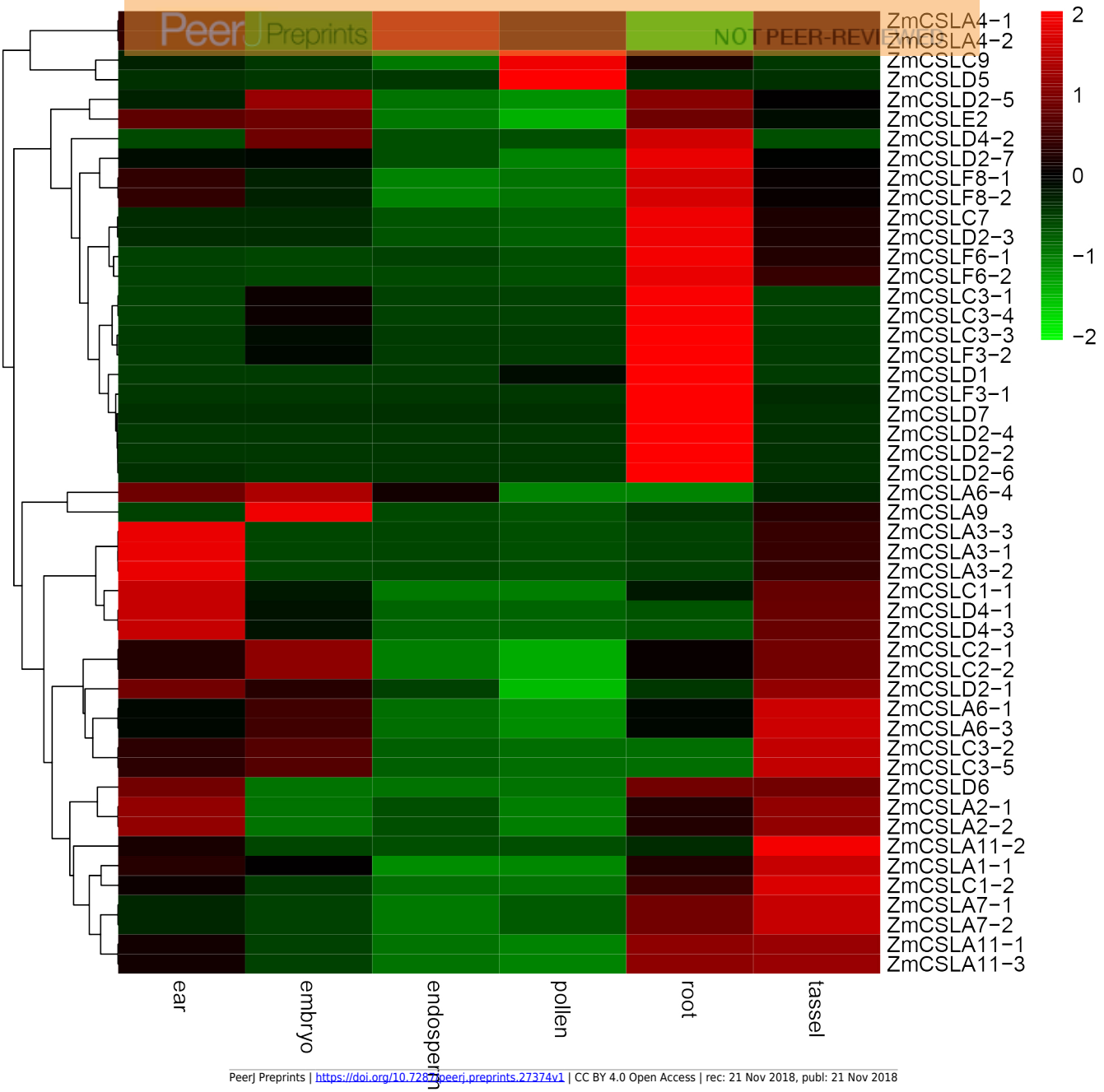


Figure 5(on next page)

Expression profiles of *ZmCsl* genes across different tissues.

The *ZmCsl* genes and color scale were shown at the right of the heatmap. The genes with an RPKM equal to 0 were not used in this heatmap.

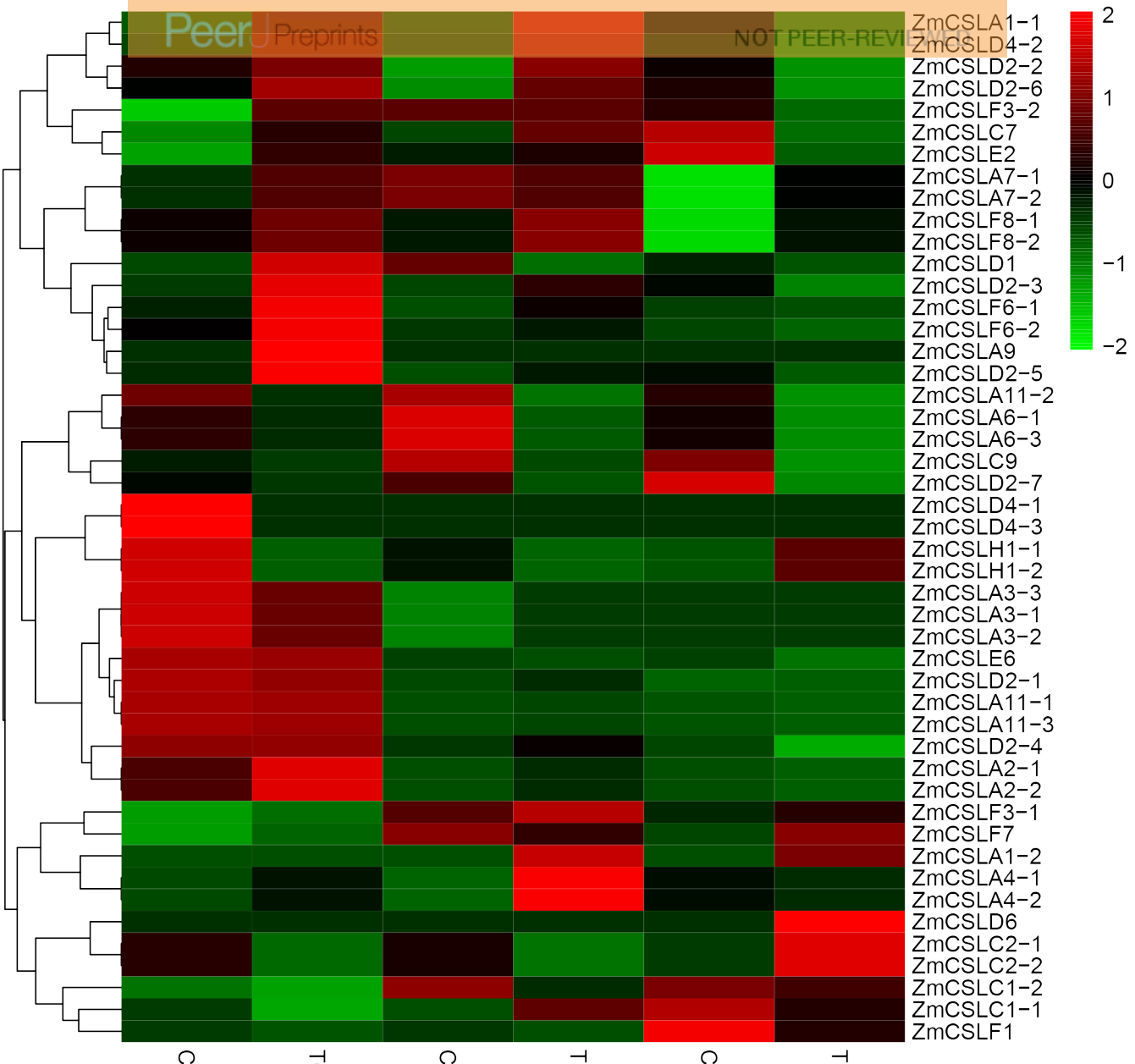


PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27374v1 | CC BY 4.0 Open Access | rec: 21 Nov 2018, publ: 21 Nov 2018

Figure 6(on next page)

Expression profiles of *ZmCsl* genes in response to drought stress treatment.

The expression color scale was shown at the right of the heatmap. Higher expression for each gene was presented in red; otherwise, green was used. T60, T96 and TR3d indicate samples were treated with PEG after 60h, 96h and recovery 3d. Control seedlings were grown under the same conditions but without PEG treatment (denoted as CK60, CK96h and CKR3d respectively).



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27374v1 | CC BY 4.0 Open Access | rec: 21 Nov 2018, publ: 21 Nov 2018

Figure 7 (on next page)

PPI interactions network of *ZmCsl* genes.

The nodes represent the proteins and the edges represent the corresponding PPI. The confidence score was required to greater than 0.7.

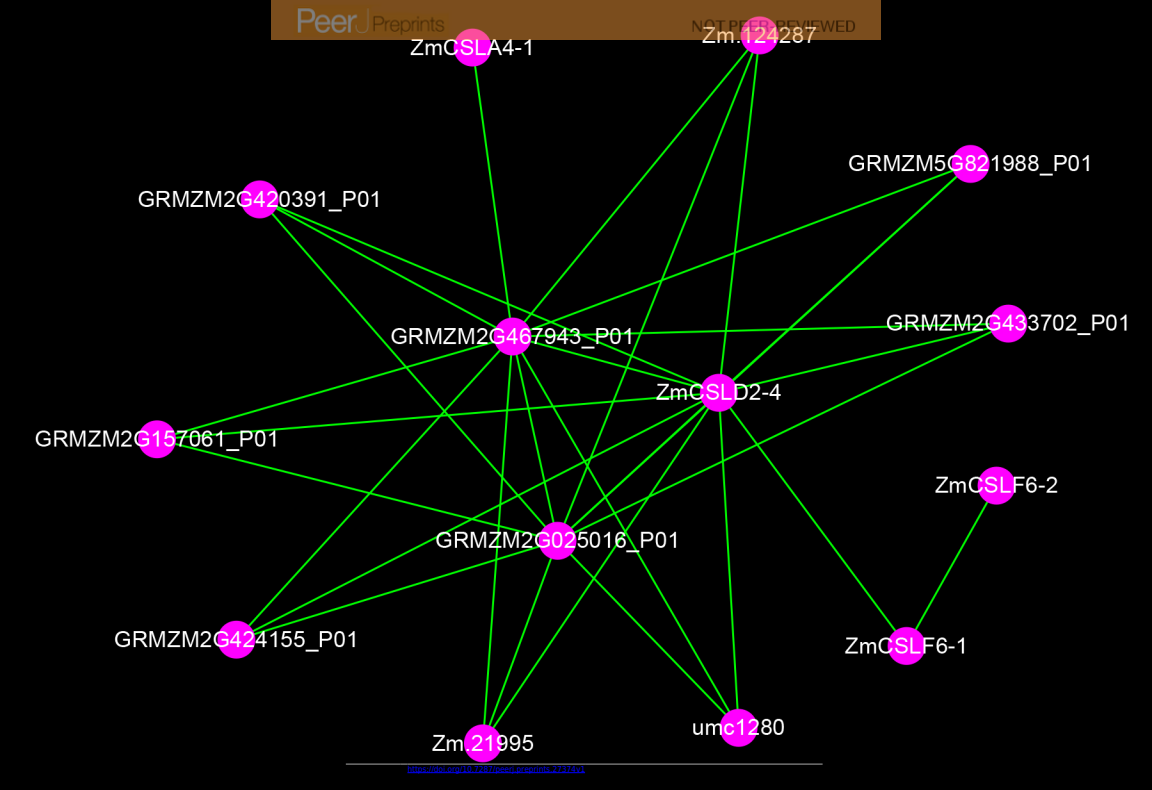


Table 1(on next page)

Csl gene family identified and characterized in maize

1

2

Nomenclature	Accession number in NCBI	Length of CDS （bp）	Chromosome	Protein length (aa)	Domain	Corresponding g rice
ZmCSLA1-1	NP_001131007.1	1542	4	504	Glycosyl transferase	CSLA1
ZmCSLA1-2	NP_001183100.1	1614	3	534	Glycosyl transferase	CSLA1
ZmCSLA2-1	XP_008664977.2	1545	1	1514	Glycosyl transferase	CSLA2
ZmCSLA2-2	XP_008664978.2	1446	1	481	Glycosyl transferase	CSLA2
ZmCSLA3-1	XP_008649262.1	1623	6	540	Glycosyl transferase	CSLA3
ZmCSLA3-2	XP_008649263.1	1611	6	536	Glycosyl transferase	CSLA3
ZmCSLA3-3	XP_020395052.1	1731	6	576	Glycosyl transferase	CSLA3
ZmCSLA4-1	NP_001130786.1	966	9	321	Glycosyl transferase	CSLA4
ZmCSLA4-2	NP_001347619.1	1620	9	539	Glycosyl transferase	CSLA4
ZmCSLA6-1	NP_001132315.1	1614	5	537	Glycosyl transferase	CSLA6
ZmCSLA6-2	XP_008657506.1	525	8	174	Glycosyl transferase	CSLA6
ZmCSLA6-3	XP_008681194.1	1611	5	536	Glycosyl transferase	CSLA6
ZmCSLA6-4	XP_020398689.1	411	8	136	Glycosyl transferase	CSLA6
ZmCSLA7-1	XP_008653381.1	1722	7	573	Glycosyl transferase	CSLA7
ZmCSLA7-2	XP_020396877.1	1746	7	581	Glycosyl transferase	CSLA7
ZmCSLA9	XP_008659807.1	1581	9	526	Glycosyl transferase	CSLA9
ZmCSLA11-1	NP_001136470.1	1089	1	362	Glycosyl transferase	CSLA11
ZmCSLA11-2	NP_001183513.2	1692	4	563	Glycosyl transferase	CSLA11
ZmCSLA11-3	NP_001346969.1	1713	1	570	Glycosyl transferase	CSLA11

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27374v1 | CC BY 4.0 Open Access | rec: 21 Nov 2018, publ: 21 Nov 2018

ZmCSLC1-1	NP_001169244.2	2052	3	683	Glycosyl transferase	CSLC1
ZmCSLC1-2	XP_008657194.1	2067	8	688	Glycosyl transferase	CSLC1
ZmCSLC2-1	NP_001335706.1	1509	7	529	Glycosyl transferase	CSLC2
ZmCSLC2-2	XP_008650958.1	2100	7	699	Glycosyl transferase	CSLC2
ZmCSLC3-1	NP_001141327.1	1659	4	552	Glycosyl transferase	CSLC3
ZmCSLC3-2	XP_008644464.1	2061	5	686	Glycosyl transferase	CSLC3
ZmCSLC3-3	XP_008662691.2	2274	10	757	Glycosyl transferase	CSLC3
ZmCSLC3-4	NP_001347177.1	2295	4	764	Glycosyl transferase	CSLC3
ZmCSLC3-5	XP_023155870.1	1947	5	648	Glycosyl transferase	CSLC3
ZmCSLC7	XP_008649860.1	2109	6	702	Glycosyl transferase	CSLC7
ZmCSLC9	NP_001334935.1	2130	1	709	Glycosyl transferase	CSLC9
ZmCSLD1	XP_008658791.1	3480	1	1159	RING/Ubox like zinc-binding Cellulose synthase	CSLD1
ZmCSLD2-1	NP_001104955.2	3231	1	1076	Zinc-binding RING-finger Cellulose synthase	CSLD2
ZmCSLD2-2	NP_001105236.2	2952	3	983	Cellulose synthase	CSLD2
ZmCSLD2-3	NP_001105621.2	3234	7	1077	Zinc-binding RING-finger Cellulose synthase	CSLD2
ZmCSLD2-4	NP_001105672.1	3237	1	1078	Zinc-binding RING-finger	CSLD2

ZmCSLD2-5	NP_001146633.1	885	2	294	Cellulose synthase Cellulose synthase	CSLD2
ZmCSLD2-6	NP_001306681.1	3159	2	1052	Zinc-binding RING-finger Cellulose synthase	CSLD2
ZmCSLD2-7	NP_001345017.1	3444	9	1147	RING/Ubox like zinc-binding Cellulose synthase	CSLD2
ZmCSLD3	XP_008664736.1	3441	1	1146	RING/Ubox like zinc-binding Cellulose synthase	CSLD3
ZmCSLD4-1	NP_001136615.1	1542	10	513	Cellulose synthase Cellulose synthase	CSLD4
ZmCSLD4-2	XP_020405933.1	2286	3	761	Cellulose synthase Cellulose synthase	CSLD4
ZmCSLD4-3	NP_001346418.1	3654	10	1217	RING/Ubox like zinc-binding Cellulose synthase	CSLD4
ZmCSLD5	XP_008659557.1	2988	9	995	Cellulose synthase Cellulose synthase	CSLD5
ZmCSLD6	XP_020400201.1	1878	9	625	Cellulose synthase	CSLF9
ZmCSLD7	XP_020393266.1	804	5	267	Cellulose synthase	CSLF3
ZmCSLE2	XP_020394267.1	2223	5	740	Cellulose synthase Cellulose synthase	CSLE2

ZmCSLE6	NP_001147894.1	2181	7	726	Cellulose synthase Cellulose synthase	CSLE6
ZmCSLF1	XP_008653206.1	2712	7	903	Cellulose synthase Cellulose synthase	CSLF1
ZmCSLF3-1	NP_001147926.1	2574	7	857	Cellulose synthase Cellulose synthase	CSLF3
ZmCSLF3-2	XP_008671789.1	2682	2	875	Cellulose synthase Cellulose synthase	CSLF3
ZmCSLF6-1	NP_001308343.1	2850	4	949	Cellulose synthase Cellulose synthase	CSLF6
ZmCSLF6-2	NP_001315306.1	2838	10	945	Cellulose synthase Cellulose synthase	CSLF6
ZmCSLF7	XP_008664938.1	2520	1	839	Cellulose synthase Cellulose synthase	CSLF7
ZmCSLF8-1	XP_008653203.1	2706	7	901	Cellulose synthase Cellulose synthase	CSLF8
ZmCSLF8-2	XP_008653204.1	2427	7	808	Cellulose synthase Cellulose synthase	CSLF8
ZmCSLH1-1	XP_008669410.1	2355	2	784	Cellulose synthase Cellulose synthase	CSLH1
ZmCSLH1-2	XP_008669411.1	2346	2	781	Cellulose synthase Cellulose synthase	CSLH1

3

Table 2(on next page)

The MEME motif sequences and lengths in ZmCsl proteins

1 **Table 2 The MEME motif sequences and lengths in ZmCsl proteins.**

Motif	Conserved amino acid sequences	E-values	Sites	Width
1	KGWKFIYVGDVKVKSELPSTYKAYRKQQHRWSCGPANL FRKMFPEIJKSK	5.1e-953	27	50
2	LNFFGFNGTAGVWRISAJEESGGWKDRTTVEDMDJAVRA HL	3.0e-965	30	41
3	IFDADFQPEPDFLKRTVPFLVHNPEJALVQARWSFVNKDE NLLTRJQEMN	3.8e-1044	29	50
4	SGYYPMVLVQIPMYNEREVYKLSIGAACGLDWPRDRFL VQVLDDSTDPVI	3.8e-868	25	50
5	SFHFVIFWILFENVMSVHRFKAAVSGLLZLG	4.2e-475	27	31
6	AYVQFPQRFDGIDPTDRYANHNRVFFDGNMRGLDGJQG PVYVGTGCVFRR	1.6e-861	26	50
7	PMLVYVSREKRPGYDHHKKAGAMNALVRVSAVLSNAP FILNLDCDHYVNN	7.2e-824	23	50
8	IYGSVTEDVVTGFRMHNRGWRSVYCSPKRDAFRGTAPIN LTDRLHQVLRW	1.8e-794	24	50
9	KWAQKGVNIKYEHRVNRKGYKAGNLKSGMECDYVKD CEFVA	3.7e-650	27	41
10	VSLWKKFNLJYLFFFVRKVVAPFYTFTLYCVIJPLSVFVPE	1.6e-573	27	41

2

3