GENOME-WIDE IDENTIFICATION AND ANALYSIS OF THE CslF GENE FAMILY IN BARLEY (Hordeum vulgare L.)

Muttanathirige Don Lalith Chandana Nishantha*1,2, Diddugodage Chamila Jeewant1,2, Guangwei Xing1, Xiaojun Nie1 and Song Weining2

Address(es):
1 State Key Laboratory of Crop Stress Biology in Arid Areas, College of Agronomy, Northwest A&F University, Yangling, Shaanxi, 712100, China.
2 Directorate of Agriculture and Livestock, Army Cantonment, Panagoda, Homagama, 10200, Sri Lanka.

*Corresponding author: mdlnnishantha@gmail.com

doi: 10.1541/jmbfs.2020.10.1.122-126

ARTICLE INFO
Received 27. 6. 2019
Revised 11. 5. 2020
Accepted 22. 5. 2020
Published 1. 8. 2020

ABSTRACT

The Cellulose synthase-like (Csl) F family has been considered as one of the most crucial genes regulating β-glucan synthesis. It is a cereal cell wall component, holding advantages for human nutrition, but disadvantages in animal nutrition, malting and brewing industries. Based on a genome-wide search method, present study identified barley CslF gene family members by considering the importance of (1,3;1,4)-β-D-glucan and newly completed barley genome. A sum of Eighteen CslF genes were recognized in the barley genome. Then, phylogenetic analyses classified them into 3 groups, that shared conserved motif compositions. A new motif, D.D.WQxQxW was also found, which was responsible for the cellulose synthase. Furthermore, using RNA-seq data, the HvCslFs expression profiles were systematically examined in different tissues and tissue-specific candidates were found. Lastly, interaction network analysis identified 11 CslF genes involved in the interaction network. All together, present task provides valuable evidence about the genomic organization and evolutionary relationship of the CslF gene family in barley, and facilitate the functional surveys of CslF genes in barley and beyond.

Keywords: barley; β-glucan; CslF; Phylogenetic analysis; Expression profile

INTRODUCTION

Barley (Hordeum vulgare), the fourth main cereal cultivated worldwide is an ancient crop, which is used as feed, food, malt and brewing industries (Arrigzen et al., 2011). It is also a highly adapted crop species, which can be grown in both desert and fertile lands (Newman et al., 2006). Barley grain endosperm cell wall contains (1, 3; 1, 4)-β-D-glucan (hereafter mentioned as β-glucan) (Fincher et al., 2004) and considering the rich β-glucan availability, researches on barley are increased during last decades (Bhatty, 1999; Bilgi et al., 2004). β-glucan is a partially water-soluble linear polysaccharides molecule, which contains glucose (Johansson et al., 2004) linked by both β-(1→3) and β-(1→4)-linkages (Rimsten et al., 2003). Barley β-glucan is beneficial for human health, which may reduce the risks of cardiovascular diseases mainly coronary heart disease, high serum cholesterol, colorectal cancer, non-insulin-dependent diabetes, obesity and hypertension (Li et al., 2003; Koghi et al., 2003; Brenman et al., 2005). Also, it has immune modulating properties, increasing vitamin and mineral bio-availability, and important in gut physiology while influencing spatial memory performance of children (Klopfenstein, 1988; Thorsburn et al., 1993; Murphy et al., 2004; http://www.nutraingredients.com). At the same time, β-glucan has adverse impact on processing applications of cereals mainly on brewing and malting while anti-nutritive for mono-gastric animals feed formulations. In feed formulations, it affects growth and feed conversion efficiency, nutritional intake of animals and stickiness of droppings. Further, when used for brewing, it reduces haze formation and the rate of wort filtration in beer, and negatively affect the malt extraction recovery (Hesselman et al., 1982; Wang et al., 1992; Brennan et al., 2005). Glycosyltransferases are in charge of the production of major wall polysaccharides i.e.; mannans, xyloglucans and β-glucans (Scheller et al., 2010). The backbones of wall polysaccharides synthases are encrypted by an enormous multigene family which is named as the Cellulose synthase/Cellulose synthase-like (CesA/Csl) superfamily (Richmond et al., 2000). It includes several sub families such as cellulose synthase subfamily (CesA) and cellulose synthase-like (Csl) sub-families, CesA to Csl, and all of them consist of multiple genes (Schwerdt et al., 2015). β-glucan synthesis is done by the CslF gene family (Richmond et al., 2000; Burton et al., 2006; Schreiber et al., 2014).

Considering the significance of β-glucan for human health, brewing and malting industries, it’s important to aware that CslF gene family which is direct for β-glucan synthesis. In this study we examined the CslF gene family in barley based on a bioinformatics search using latest genomic information. The phylogenetic tree, interaction network, conserved motifs and gene expression pattern of CslF were further systematically analyzed. Present study provides the basic genomic organization information of CslF genes in barley, and it will help for further functional studies.

MATERIAL AND METHODS

CslF genes in barley

Using the method given by Wang et al., 2016 all possible members of barley CslF gene family were recognized. To create a local protein database, all existing protein sequences for Hordeum vulgare L. were retrieved from the Ensemble database (http://plants.ensembl.org/index.html) (Bolser et al., 2016). The available CslF genes of Arundo donax, Avena sativa, Brachypodium distachyon, Zea mays, Sorghum bicolor, Triticum aestivum, Oryza sativa and Triticum urartu depositing in National Center for Biotechnology Information (NCBI) database and hmm-build tool embedded in HMMER 3.0 were utilized to build a hidden Markov model (HMM) profile. Then the HMM profile and the Hmmsearch tool embedded in HMMER 3.0 were applied to search barley proteins (Wheeler et al., 2013). Conserved domains of barley CslF members were further confirmed by InterProScan database (Zdobnov et al., 2001) and PAM (Finn et al., 2016). Lastly, sequence verification was done by a BLASTN (Nucleotide BLAST) similarity search compared to barley expressed sequence tags (ESTs) deposited in the NCBI database. The Mw (molecular weight) of candidate and theoretical pl (isoelectric point) value of genes were calculated by online compute pl/Mw tool (Gasteiger et al., 2003). Using the CELLO v2.5 web server, subcellular localization of those genes were predicted (Yu et al., 2006).

Phylogenetic analysis and multiple alignments

ClustalW tool was used to perform multiple sequence alignments (Larkin et al., 2007). Then the phylogenetic tree was created by combining neighbor-joining
method and bootstrap test method with thousand replications in MEGA 6.0 software (Tamura et al., 2013). The conserved motifs of CslF were predicted by the MEME program (Bailey et al., 2009).

CslF RNA-seq datasets expression profiles

RNA-seq datasets obtained from NCBI Sequence Read Archive (SRA) was utilized to learn the expression profile of HvCslF genes in various tissues. Sample information and the data used were depicted in Table S1. TopHat and Cufflinks software were used to analyze gene expressions (Trapnell et al., 2012). For each gene, the FPKM value was taken. To generate the heat map, log10-transform (FPKM +1) values of HvCslF genes were used.

Analysis of Co-expression network

To investigate the gene function and regulatory pathway, the most commonly used method is study of co-expression networks. Barley CslF genes co-expression network was created with the help of WGCNAR 1.49 package by analyzing RNA-seq data using weighted correlation network analysis (Langfelder et al., 2008).

RESULTS AND DISCUSSION

Genome-wide identification of the CslF gene family in barley

To find out the members of CslF family in barley, we performed a HMM search using the latest updated genome resource and totally 18 non-redundant CslF genes have been recognized in barley genome (Tab 1). Based on the chromosome location, the predicted barley CslF genes were then designated as HvCslF1 to HvCslF18. CslF genes count in barley (18) was greater than maize (7) and rice (8) (Schwerdt et al., 2015; Penning et al., 2009). HvCslF cascade genes locations were not random along barley chromosomes. Four HvCslF genes per chromosome (total 12) were located on 2, 5 and 7 chromosomes, while 3 genes were located in chromosome 1. One gene per each chromosome was on chromosome 3 and 6. There were no HvCslF genes found on chromosome 4.

According to the literature, previous studies have discovered only 10 CslF genes in the CslF subfamily in barley (Schreiber et al., 2014; Burton et al., 2011). Present experiment, identified 18 CslF genes in barley using newly completed barley genome through a genome-wide search. According to Burton et al., (2008), the 7 CslF genes found from barley were divided into two groups and 4 HvCslF genes were mapped to 2H chromosome. Rest of the HvCslF genes were mapped to chromosomes 7H, 5H and 1H. In 2014, Schreiber et al., (2014) mapped 10 HvCslF genes on barley chromosomes, where 5 HvCslF genes in chromosome 2H and the rest of HvCslF genes have been mapped to chromosomes 7H, 5H and 1H.

The putative HvCslF proteins’ length was starting from 606 to 1207 amino acids, with theoretical pl extending from 6.04 to 8.83 and putative molecular weights (Mw) starting from 68500.08 up to 132914.06 Da. According to the Subcellular localization analysis, most of the HvCslF genes were localized in Inner Membrane (Table 1). BLASTN search against UniGene database and barley EST using the HvCslF genes as queries was done to find out the actual presence of above putative genes. Results showed that all HvCslF genes had EST support, suggesting they are truly found in barley genome.

| Gene  | Ensemble Barley Gene ID | Chromosome Number | Amino Acid Length | pI | Mw (Da)               | EST Count | Subcellular Location |
|-------|-------------------------|-------------------|-------------------|----|----------------------|-----------|----------------------|
| HvCslF1 | HORVU0Hr1G038120 | Unknown | 1144 | 8.26 | 128803.08 | 19 | Cytoplasmic |
| HvCslF2 | HORVU1Hr1G022900 | 1 | 802 | 8.18 | 90216.61 | 9 | InnerMembrane |
| HvCslF3 | HORVU1Hr1G026320 | 1 | 920 | 6.5 | 100340.87 | 18 | Cytoplasmic |
| HvCslF4 | HORVU1Hr1G039250 | 1 | 966 | 8.39 | 109510.29 | 21 | InnerMembrane |
| HvCslF5 | HORVU2Hr1G042240 | 2 | 821 | 6.18 | 92464.71 | 9 | InnerMembrane |
| HvCslF6 | HORVU2Hr1G042250 | 2 | 897 | 7.89 | 99823.28 | 9 | InnerMembrane |
| HvCslF7 | HORVU2Hr1G042350 | 2 | 869 | 7.92 | 97507.51 | 9 | InnerMembrane |
| HvCslF8 | HORVU2Hr1G042370 | 2 | 900 | 7.01 | 101803.96 | 7 | InnerMembrane |
| HvCslF9 | HORVU3Hr1G071770 | 3 | 1041 | 6.04 | 116685.80 | 20 | Cytoplasmic |
| HvCslF10 | HORVU5Hr1G023640 | 5 | 1207 | 8.32 | 132914.06 | 5 | InnerMembrane |
| HvCslF11 | HORVU5Hr1G064230 | 5 | 686 | 8.72 | 77120.31 | 23 | InnerMembrane |
| HvCslF12 | HORVU5Hr1G10000 | 5 | 1023 | 8.22 | 114840.73 | 17 | InnerMembrane |
| HvCslF13 | HORVU5Hr1G118270 | 5 | 1141 | 8.83 | 129402.91 | 16 | InnerMembrane |
| HvCslF14 | HORVU6Hr1G050750 | 6 | 838 | 8.72 | 94826.85 | 21 | InnerMembrane |
| HvCslF15 | HORVU7Hr1G005270 | 7 | 1188 | 7.07 | 131851.52 | 26 | InnerMembrane |
| HvCslF16 | HORVU7Hr1G070010 | 7 | 947 | 8.54 | 103629.28 | 20 | InnerMembrane |
| HvCslF17 | HORVU7Hr1G081850 | 7 | 606 | 8.51 | 68590.08 | 5 | InnerMembrane |
| HvCslF18 | HORVU7Hr1G121040 | 7 | 834 | 6.59 | 93563.78 | 9 | InnerMembrane |

Legend: Molecular weight (Mw), Expressed sequence tags (EST) and Isoelectric point (pI),

Analysis of multiple alignments, phylogeny and conserved motif of HvCslF

Using ClustalW software, the full-length protein sequences of the 18 HvCslF were aligned to evaluate the phylogenetic relationships of the HvCslF genes (Larkin et al., 2007). Further, the MEGA 6.0 with neighbor joining (NJ) method has been used to phylogenetic tree construction (Tamura et al., 2013). We observed highly conserved motifs regions in the sequences of the HvCslF genes (Figure 1A). Considering the phylogenetic analysis HvCslFs were divided into three groups (Figure 1B), which included 6 (Group I), 4 (Group II), and 8 (Group III) members, respectively. The results showed some evolutionary changes found in barley CslF genes, which was consistent with those in Oryza sativa (Burton et al., 2008). Furthermore, each CslF cluster recognized by phylogenetic analysis, has same composition of conserved motifs. Totally 15 motifs have been recognized in HvCslF proteins. All of the CslF genes products contain a D,D,D,QxrxRW motif, which has been called as the nucleotide sugar-binding domain as well as the catalytic site of enzymes (Richmond et al., 2000; Schwerdt et al., 2015; Doblin et al., 2009). Motif 3 (IPR00150) is an important motif in the HvCslF family, being responsible for the cellulose synthase, which was found by InterProScan analysis. Other most important motifs in the HvCslF family are Motifs 1 and 2, which containing D,D,WQxxD. Conserved domains analysis and identification may assist to identify gene’s functional units as well as elaborate their tasks in growth and development of a plant. The Motif D, D,D,QxrxRW is available in other cereals like rice, maize, and Brachypodium distachyon (Schwerdt et al., 2015).
HvCslF genes expression pattern

By using RNA-seq data from NCBI database (https://www.ncbi.nlm.nih.gov/), the expression patterns of HvCslFs in seven barley tissues was studied (Table S1). The heat map (Figure 2) indicated that, only 11 genes detected in several barley tissues and, their expression levels were highly variable. Most of the tissues showed higher expression of HvCslF1, 3, 4, 9 and 15 genes. Among these tissues, the highest number of genes (5) were expressed in the grain. In the grain, HvCslF9 shows the highest expression and HvCslF14 was only expressed in the grains. Hence, these two genes could be suggested as the main candidate genes for β-glucan synthesis in grain. Some HvCslFs were highly expressed in specific tissues such as, HvCslF1, 9 and 15 were highly expressed in palea, grain and lodicule, respectively, and HvCslF3 was also relatively highly expressed in palea and lodicule tissues, proposing that those genes may play crucial roles in these tissues. The other 5 genes, HvCslF2, 6, 8, 12, and 13 were not expressed in the studied tissues. It is worth mentioning that, the CslF genes specificity has been studied in rice and maize (Wang et al., 2010; Penning et al., 2009).

Several studies showed that, the CslF genes have a key role in β-glucan production (Burton et al., 2011). For example, CslF genes in Oryza sativa (Hazen et al., 2002), Triticum aestivum (Jobling, 2015) and Zea mays (Alexandrov et al., 2009) were reported to regulate β-glucan synthesis. Also, expression of barley CslF genes are highly variable under the abiotic stresses such as water stress. Quantity of barley β-glucan is affected by the quantity of water supply during the maturity. Moisture level in soil and β-glucan in the barley kernel has a negative correlation (Guler, 2003). Barley β-glucan content is highly affected by dry conditions prevailing at grain maturation period (Hang et al., 2007). β-glucan levels in barley grains change dramatically during its growth and development due to numerous expression patterns of HvCslF genes in various tissues at different times. Gilbeaut et al., (2005) observed, increased β-glucan level in barley coleoptiles walls at the elongation phase, followed by cessation of growth at about 5 days, β-glucan content rapidly decreases. The transient nature of β-glucan in maize coleoptiles is described by McCann et al., 2007.

Interactions between the HvCslF family members

Present experiment, created the interaction network of the HvCslF family considering the various tissues in barley. Using RNA-seq data, WGCNAR package (Langfelder et al., 2008) provided a wide-ranging functions to perform the analysis of weighted correlation network. For example in Figure 3, 11 out of 18 HvCslF genes were found in the co-relation network analysis. Nine HvCslF genes (HvCslF1, HvCslF2, HvCslF4, HvCslF9, HvCslF10, HvCslF11, HvCslF12, HvCslF13 and HvCslF15) were involved in a single cascade, illustrating that they have close relationship with each other, and 2 HvCslF genes (HvCslF5 and HvCslF7) located separately, indicating that they have no close relationship with the other 9 HvCslF genes. Furthermore, 3 of the 9 HvCslF genes in Group III (Figure 1), including HvCslF4, HvCslF9 and HvCslF11, showed a close association with each other than the other genes in the cascade.
In the network, the largest group was HvCslF10 followed by HvCslF2 and HvCslF12, respectively. This emphasizes that, these HvCslF genes have more relationship with other genes in the genome. For example, in the HvCslF10 group, five Cyclin B (CYCB) genes are interacting, which are responsible for the cell cycle regulation, that regulates the gap 2 (G2)/mitosis (M) transition (Ishida et al.; 2008Lin et al., 2017). The HvCslF2 group has interaction with GSL genes, suggested to have a crucial role in the synthesis of callose, which belongs to glucan synthase-like (GSL) family (Shu et al., 2014; Toller et al., 2008). The HvCslF22 group has interaction with UGP genes, which are responsible for Plant UDP-glucose (UDPG) pyrophosphorylase (UGPase) which are essential in the metabolism or production of UDPG, an essential metabolite for cell wall and sucrose synthesis (Meng et al., 2007; Meng et al., 2008). The interaction network constructed in this study between HvCslFs provides information for further studies on β-glucan synthesis and other genes which have interaction with HvCslF genes in barley and genomes of other species.

CONCLUSION

The evolution relationship, expression profiles and genome organization of the CslF gene family in barley were investigated in the present study. Totally, 18 HvCslF genes were identified based on a bioinformatics search using latest genomic information. The gene expression pattern, phylogenetic tree, conserved motifs as well as interaction network of the CslF was further systematically analyzed. This is the first study to report the barley CslF family at the genome scale, which is providing the candidates for advance functional studies, and facilitates to expose the regulatory mechanism of the CslF family involving in development and growth in barley and beyond.

Acknowledgments: Authors are gratitude to China government scholarship.

REFERENCES

Alexandrov, N. N., Brover, V. V., Freidin, S., Troukhan, M. E., Tatarinova, T. V., Zhang, H., Feldmann, K. A. (2008). Insights into corn genes derived from large-scale cDNA sequencing. *Plant Molecular Biology*, 69(1-2), 179–194. https://doi.org/10.1007/s11103-008-9415-4.

Arngren, M., Hansen, P. W., Eriksen, B., Larsen, J., & Larsen, R. (2011). Analysis of Pregenerated Barley Using Hyperspectral Image Analysis. *Journal of Agricultural and Food Chemistry*, 59(21), 11385–11394. https://doi.org/10.1021/jf2017122y.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, 37(Web Server), W202–W208. https://doi.org/10.1093/nar/gkp335.

Bhat, R. S. (1999). β-Glucan and Flour Yield of Hull-less Barley. *Cereal Chemistry Journal*, 76(2), 314–315. https://doi.org/10.1094/chem.1999.76.2.314.

Bilgic, B., & Celik, S. (2004). Solubility and emulsifying properties of barley protein concentrates. *European Food Research and Technology*, 218(5), 437–441. https://doi.org/10.1007/s00217-004-0895-4.

Bolser, D., Staines, D. M., Pritchard, E., & Kersey, P. (2016). EnsemblPlants: Integrating Tools for Visualizing, Mining, and Analyzing Plant Genomics Data. *Methods in Molecular Biology*, 115–140. https://doi.org/10.1007/978-1-4939-3167-5.

Brennan, C. S., & Cleary, L. J. (2005). The potential use of cereal (1→3,1→4)-β-d-glucans as functional food ingredients. *Journal of Cereal Science*, 42(1), 1–13. https://doi.org/10.1016/j.jcs.2005.01.002.

Burton, R. A., Collins, H. M., Kibble, N. A. J., Smith, J. A., Shirley, N. J., Joblin g, S. A., … Fincher, G. B. (2011). Over-expression of specific HvCslF cellulose synthase-like genes in transgenic barley increases the levels of cell wall (1,3;1,4)-β-d-glucans and alters their fine structure. *Plant Biotechnology Journal*, 9(2), 117–135. https://doi.org/10.1111/j.1467-7652.2010.00532.x.

Burton, R. A., Jobling, S. A., Harvey, A. J., Shirley, N. J., Mather, D. E., Bacic, A., & Fincher, G. B. (2008). The Genetics and Transcriptional Profiles of the Cel lulose Synthase-Like HvCslF Gene Family in Barley. *Plant Physiology*, 146(4), 1 821–1833. https://doi.org/10.1104/pp.107.114694.

Burton, R. A. (2006). Cellulose Synthase-Like CslF Genes Mediate the Synthesis of Cell Wall (1,3;1,4)-β-D-Glucans. *Science*, 311(5769), 1940–1942. https://doi.org/10.1126/science.1122975.

Doblin, M. S., Pettolino, F. A., Wilson, S. M., Campbell, R., Burton, R. A., Finch er, G. B Bacic, A. (2009). A barley cellulose synthase-like CSLH gene mediates (1,3;1,4)-β-D-glycans synthesis in transgenic Arabidopsis. *Proceedings of the National Academy of Sciences*, 106(14), 5996–6001. https://doi.org/10.1073/pnas.090 2019106.
