Genome-wide analysis of the cellulose synthase-like (Csl) gene family in bread wheat (Triticum aestivum L.)

Simerjeet Kaur, Kanwarpal S. Dhugga, Robin Beech and Jaswinder Singh

Abstract

Background: Hemicelluloses are a diverse group of complex, non-cellulosic polysaccharides, which constitute approximately one-third of the plant cell wall and find use as dietary fibres, food additives and raw materials for biofuels. Genes involved in hemicellulose synthesis have not been extensively studied in small grain cereals.

Results: In efforts to isolate the sequences for the cellulose synthase-like (Csl) gene family from wheat, we identified 108 genes (hereafter referred to as TaCsl). Each gene was represented by two to three homeoalleles, which are named as TaCslXY_ZA, TaCslXY_ZB, or TaCslXY_ZD, where X denotes the Csl subfamily, Y the gene number and Z the wheat chromosome where it is located. A quarter of these genes were predicted to have 2 to 3 splice variants, resulting in a total of 137 putative translated products. Approximately 45% of TaCsl genes were located on chromosomes 2 and 3. Sequences from the subfamilies C and D were interspersed between the dicots and grasses but those from subfamily A clustered within each group of plants. Proximity of the dicot-specific subfamilies B and G, to the grass-specific subfamilies H and J, respectively, points to their common origin. In silico expression analysis in different tissues revealed that most of the genes were expressed ubiquitously and some were tissue-specific. More than half of the genes had introns in phase 0, one-third in phase 2, and a few in phase 1.

Conclusion: Detailed characterization of the wheat Csl genes has enhanced the understanding of their structural, functional, and evolutionary features. This information will be helpful in designing experiments for genetic manipulation of hemicellulose synthesis with the goal of developing improved cultivars for biofuel production and increased tolerance against various stresses.

Keywords: Arabinoxylan, Bioenergy, Biofuels, Cell wall, Cellulose, CesA, Csl, Glucuronoarabinoxylan, Mixed-linked glucan, Wheat

Background

Plant cell wall consists of three main polysaccharide fractions: cellulose, hemicellulose, and pectin, with lignin and proteins being the other two constituents. Grass walls contain mainly two of the three polysaccharide fractions with pectin being a rather minor constituent. Hemicelluloses are plant cell wall matrix polysaccharides that possess diverse linear or branched structures [1, 2]. These mainly encompass 1–4-β-glucan, 1,3;1,4-β-glucan, galactan, and glucomannan in grasses [3]. In addition, glucuronoarabinoxylan is a major grass wall constituent. Because of the presence of heterogeneous substituents or other linkages on their polymer backbones, hemicelluloses are non-crystalline and can be readily hydrolysed in comparison to cellulose. These polysaccharides can interact with cellulose microfibrils through hydrogen bonds [4].

Hemicellulosic polysaccharide backbones in plants are made by the cellulose synthase-like (Csl) enzymes, which are members of a much larger superfamily of genes referred to as glycosyltransferase 2 (GT2) [5]. Several other GTs, i.e., xyloglucan α-1,6-xylosyltransferases (GT34), xyloglucan fucosyltransferases (GT37), and xyloglucan galactosyltransferases (GT47) have been reported to be involved in the biosynthesis of xyloglucans [6]. Genes...
encoding Csl enzymes share sequence similarity with the cellulose synthase A (CesA) gene family known to form cellulose throughout the plant kingdom [7]. A variable number of Csl genes ranging from 30 to 50 have been reported from different plant species and are classified into nine subfamilies (CslA–CslH and CslI) [8, 9]. Cereals generally lack CslB and CslG families. Among the remaining families, CslA, CslC, and CslD are conserved in all land plants, whereas CslF, CslH are restricted to grasses [10, 11]. A poorly understood subfamily, CslJ, has been reported in grasses as well as dicots, which contrasts with the previous claims of its occurrence only in grasses [12, 13]. Similarly, the subfamilies CslB and CslG were previously reported to be specific to dicots [14]. However, a recent report established the presence of the CslB subfamily in monocots as well [12]. Several of the Csl subfamilies have been reported to be involved in the biosynthesis of different cell wall polysaccharides. For example, subfamily CslA was shown to form β-1,4-mannan backbone of galactomannan and glucomannan [15, 16]. Similarly, CslF and CslH subfamilies were shown to make 1–3,1–4-β-glucan in grasses [17, 18], whereas CslC genes were associated with the formation of the 1–4-β-glucan backbone of a xyloglucan and some other polysaccharides [19]. Wheat is a major cereal crop grown on the largest area of arable land in the world, is second only to maize in grain production, and feeds approximately 40% of the world population [20]. It has a large genome size (~17 Gb), of which ~80–90% is repetitive [21]. Even after the complete genome sequence became available [22], Csl genes remain unidentified and uncharacterized in bread wheat. In general, homologous copies of most of the genes are located on each of the three chromosomes belonging to each of the subgenomes (A, B, and D), suggesting that the number of Csl genes is expected to approximately three-times that of a diploid species like rice. We used publicly available resources to retrieve wheat genome sequence. Large-scale data mining was performed using the Pfam domain models for the identification of Csl gene family members, which are reported in this study.

Methods

Data sources and sequence retrieval

Wheat genome data were downloaded from the Ensembl Plants FTP server (ftp://ftp.ensemblgenomes.org/pub/current/plants/fasta/triticum_aestivum/), generated by the International Wheat Genome Sequencing Consortium (IWGSC) and converted into a local BLAST database using the UNIX pipeline. BLAST analyses (BLASTN as well as BLASTP) were performed using the standalone command line version of NCBI (National Center for Biotechnology Information) blast 2.2.28+ (ftp://ftp.ncbi.nih.gov/blast/executables/LATEST/), released March 19, 2013. A query file was generated from Pfam domain models; PF00535 (GT2) domain and PF03552 (Cellulose_synt) downloaded from Pfam 30.0 June 2016 release [23]. The sequences of splice variants were also retrieved from Ensembl Plants browser (http://plants.ensembl.org/Triticum_aestivum/Info/Index). Analysis of splice variants was conducted as described by Kim et al. (2007) [24]. Previously known Csl sequences from Arabidopsis thaliana, Oryza sativa, and Zea mays were downloaded from the Cell Wall Navigator database [25]. For Brachypodium, sequences were retrieved from phytomine (https://phytosome.jgi.doe.gov). Amino acid sequences of the aforementioned CSL proteins are given in Additional file 1: Figure S1.

Blast searches for wheat homologs

All query files containing the two Pfam domain models (PF00535 and PF03552) were used to perform the BLASTn searches against the local blast database of bread wheat. All blast hits with E-value >1.0 were removed. Using cut-off E- value <1.0, all previously known CesA genes were retrieved. After the compilation of all the sequences below the cut-off value, CD-hit program was used to obtain non-redundant sequences. Higher cut-off E- value was used to ascertain the identification of all the genes that possessed the Pfam domains PF00535 and PF03552. These genes were further filtered through phylogenetic analysis along with previously known CSL proteins from Arabidopsis, Brachypodium, maize, and rice, which reflected some non-targeted genes that were removed from further analysis [26]. Phylogenetic analysis was also implemented to categorize different Csl sub-families. CesA genes were distinguished from the Csl genes with the CXC motif, which is diagnostic of the CesA but absent from the Csl proteins [7, 27]. Presence of the conserved domains Cellulose_synt/GT2 was confirmed using a batch blast search at the CDD (conserved domain database) of NCBI. Homologous genes from each of the three genomes were named TaCslXY_ZA, TaCslXY_ZB, or TaCslXY_ZD, where X denotes the Csl subfamily, Y the gene number and Z the wheat chromosome where it is located. Alignment of the sequences of all newly identified wheat Csl genes is given in Additional file 2: Figure S2.

Protein structure and motif/domain identification

Protein sequences were downloaded from the Ensembl Plants FTP server (ftp://ftp.ensemblgenomes.org/pub/current/plants/fasta/triticum_aestivum/), developed by the International Wheat Genome Sequencing Consortium (IWGSC) [22]. Multiple protein sequence alignments were performed using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) [28]. The resulting
Table 1 Homeologous copies of the bread wheat Csl genes

| No. | Ensembl ID                  | Gene name         | Corresponding gene in rice |
|-----|-----------------------------|-------------------|-----------------------------|
| 1   | TRIAE_CS42_6BS_TGACv1_513375_AA1639370.1 | TaCslA1_6BS      | CslA1                        |
| 2   | TRIAE_CS42_6AS_TGACv1_485966_AA1554960.1 | TaCslA1_6AS      | CslA1                        |
| 3   | TRIAE_CS42_2AL_TGACv1_093375_AA0278800.1 | TaCslA2_2AL      | CslOS09G39920                |
| 4   | TRIAE_CS42_2BL_TGACv1_129747_AA0394630.1 | TaCslA2_2BL      | CslOS09G39920                |
| 5   | TRIAE_CS42_2DL_TGACv1_160461_AA0550770.1 | TaCslA2_2DL      | CslOS09G39920                |
| 6   | TRIAE_CS42_1AS_TGACv1_019142_AA0061550.1 | TaCslA2_1AS      | CslOS09G39920                |
| 7   | TRIAE_CS42_7BS_TGACv1_502860_AA1945380.1 | TaCslA3_7BS      | CslA3                        |
| 8   | TRIAE_CS42_7DS_TGACv1_623146_AA2050070.1 | TaCslA3_7DS      | CslA3                        |
| 9   | TRIAE_CS42_7AS_TGACv1_569190_AA1809650.1 | TaCslA3_7AS      | CslA3                        |
| 10  | TRIAE_CS42_6DS_TGACv1_543811_AA1744360.1 | TaCslA4_6DS      | CslA10/4/2                   |
| 11  | TRIAE_CS42_6AS_TGACv1_487286_AA1569690.1 | TaCslA4_6AS      | CslA10/4/2                   |
| 12  | TRIAE_CS42_2BS_TGACv1_146583_AA0468630.1 | TaCslA4_2BS      | CslA5/7                      |
| 13  | TRIAE_CS42_2AS_TGACv1_113418_AA0355820.1 | TaCslA5_2AS      | CslA5/7                      |
| 14  | TRIAE_CS42_2DS_TGACv1_177473_AA0578070.1 | TaCslA5_2DS      | CslA5/7                      |
| 15  | TRIAE_CS42_3DL_TGACv1_249033_AA0835410.1 | TaCslA6_3DL      | CslA11                       |
| 16  | TRIAE_CS42_3B_TGACv1_221079_AA0729630.1 | TaCslA6_3B       | CslA11                       |
| 17  | TRIAE_CS42_3AL_TGACv1_197519_AA0666560.1 | TaCslA6_3AL      | CslA11                       |
| 18  | TRIAE_CS42_3AS_TGACv1_113300_AA0354190.1 | TaCslA7_3AS      | CslA5/7                      |
| 19  | TRIAE_CS42_3DS_TGACv1_177798_AA0584795.1 | TaCslA7_3DS      | CslA5/7                      |
| 20  | TRIAE_CS42_3B_TGACv1_220828_AA0720500.1 | TaCslA8_3B       | CslA11                       |
| 21  | TRIAE_CS42_3DS_TGACv1_273022_AA0927600.1 | TaCslA8_3DS      | CslA11                       |
| 22  | TRIAE_CS42_3AS_TGACv1_558725_AA1795700.1 | TaCslA8_3AS      | CslA11                       |
| 23  | TRIAE_CS42_3DL_TGACv1_22332 AA0780350.1 | TaCslA9_3DL      | CslA9                        |
| 24  | TRIAE_CS42_3B_TGACv1_221079_AA0729630.1 | TaCslA9_3B       | CslA9                        |
| 25  | TRIAE_CS42_3AS_TGACv1_558725_AA1795700.1 | TaCslA9_3AS      | CslA9                        |
| 26  | TRIAE_CS42_3DS_TGACv1_558725_AA1795700.1 | TaCslA9_3DS      | CslA9                        |
| 27  | TRIAE_CS42_3B_TGACv1_221079_AA0729630.1 | TaCslA9_3AS      | CslA9                        |
| 28  | TRIAE_CS42_3AS_TGACv1_558725_AA1795700.1 | TaCslA9_3DS      | CslA9                        |
| 29  | TRIAE_CS42_3DS_TGACv1_558725_AA1795700.1 | TaCslA9_3AS      | CslA9                        |
| 30  | TRIAE_CS42_3AS_TGACv1_558725_AA1795700.1 | TaCslA9_3DS      | CslA9                        |
| 31  | TRIAE_CS42_3AS_TGACv1_558725_AA1795700.1 | TaCslA9_3AS      | CslA9                        |
| 32  | TRIAE_CS42_3AS_TGACv1_558725_AA1795700.1 | TaCslA9_3DS      | CslA9                        |
| 33  | TRIAE_CS42_3AS_TGACv1_558725_AA1795700.1 | TaCslA9_3AS      | CslA9                        |
| 34  | TRIAE_CS42_3AS_TGACv1_558725_AA1795700.1 | TaCslA9_3DS      | CslA9                        |
| 35  | TRIAE_CS42_3AS_TGACv1_558725_AA1795700.1 | TaCslA9_3AS      | CslA9                        |
| 36  | TRIAE_CS42_3AS_TGACv1_558725_AA1795700.1 | TaCslA9_3DS      | CslA9                        |
| 37  | TRIAE_CS42_3AS_TGACv1_558725_AA1795700.1 | TaCslA9_3AS      | CslA9                        |
| 38  | TRIAE_CS42_3AS_TGACv1_558725_AA1795700.1 | TaCslA9_3DS      | CslA9                        |
| 39  | TRIAE_CS42_3AS_TGACv1_558725_AA1795700.1 | TaCslA9_3AS      | CslA9                        |
| 40  | TRIAE_CS42_3AS_TGACv1_558725_AA1795700.1 | TaCslA9_3DS      | CslA9                        |
| 41  | TRIAE_CS42_3AS_TGACv1_558725_AA1795700.1 | TaCslA9_3AS      | CslA9                        |
| 42  | TRIAE_CS42_3AS_TGACv1_558725_AA1795700.1 | TaCslA9_3DS      | CslA9                        |
| 43  | TRIAE_CS42_3AS_TGACv1_558725_AA1795700.1 | TaCslA9_3AS      | CslA9                        |
Table 1  Homeologous copies of the bread wheat Csl genes (Continued)

| No. | Ensembl ID                | Gene name | Corresponding gene in rice |
|-----|---------------------------|-----------|----------------------------|
| 44  | TRIAE_CS42_5DL_TGACv1_435778_AA1454840.1 | TaCslC10_5DL | CslC10/9                   |
| 45  | TRIAE_CS42_5AL_TGACv1_374268_AA1195590.1 | TaCslC10_5AL | CslC10/9                   |
| 46  | TRIAE_CS42_1BL_TGACv1_030586_AA0094860.1 | TaCslD1_1BL | CslD1                      |
| 47  | TRIAE_CS42_1AL_TGACv1_001700_AA0034150.1 | TaCslD1_1AL | CslD1                      |
| 48  | TRIAE_CS42_1DL_TGACv1_063091_AA0223780.1 | TaCslD1_1DL | CslD1                      |
| 49  | TRIAE_CS42_2BS_TGACv1_148683_AA0494520.1 | TaCslD2_2BS | CslD3                      |
| 50  | TRIAE_CS42_2DS_TGACv1_177279_AA0572180.1 | TaCslD3_2DS | CslD3                      |
| 51  | TRIAE_CS42_2AS_TGACv1_114244_AA0365360.1 | TaCslD3_2AS | CslD3                      |
| 52  | TRIAE_CS42_1BS_TGACv1_049706_AA0160220.1 | TaCslD4_1BS | CslD4                      |
| 53  | TRIAE_CS42_5BS_TGACv1_433536_AA1415840.1 | TaCslD4_5BS | CslD4                      |
| 54  | TRIAE_CS42_6DL_TGACv1_526558_AA1687090.1 | TaCslE2_6DL | CslE2                      |
| 55  | TRIAE_CS42_6AL_TGACv1_471004_AA1500600.1 | TaCslE2_6AL | CslE2                      |
| 56  | TRIAE_CS42_6BL_TGACv1_499967_AA1596110.1 | TaCslE2_6BL | CslE2                      |
| 57  | TRIAE_CS42_5DL_TGACv1_603510_AA1985050.1 | TaCslE3     | CslE3                      |
| 58  | TRIAE_CS42_5AL_TGACv1_559436_AA1799630.1 | TaCslE3     | CslE3                      |
| 59  | TRIAE_CS42_5BL_TGACv1_406235_AA1342600.1 | TaCslE3     | CslE3                      |
| 60  | TRIAE_CS42_5DS_TGACv1_526558_AA1687090.1 | TaCslE3     | CslE3                      |
| 61  | TRIAE_CS42_5BL_TGACv1_406235_AA1342600.1 | TaCslE3     | CslE3                      |
| 62  | TRIAE_CS42_5AL_TGACv1_376126_AA1232370.1 | TaCslE3     | CslE3                      |
| 63  | TRIAE_CS42_5BS_TGACv1_425241_AA1392650.1 | TaCslE3     | CslE3                      |
| 64  | TRIAE_CS42_5DS_TGACv1_435778_AA1454840.1 | TaCslE3     | CslE3                      |
| 65  | TRIAE_CS42_5BL_TGACv1_406235_AA1342600.1 | TaCslE3     | CslE3                      |
| 66  | TRIAE_CS42_5AL_TGACv1_376126_AA1232370.1 | TaCslE3     | CslE3                      |
| 67  | TRIAE_CS42_5BS_TGACv1_425241_AA1392650.1 | TaCslE3     | CslE3                      |
| 68  | TRIAE_CS42_5BL_TGACv1_406235_AA1342600.1 | TaCslE3     | CslE3                      |
| 69  | TRIAE_CS42_5AL_TGACv1_376126_AA1232370.1 | TaCslE3     | CslE3                      |
| 70  | TRIAE_CS42_5BS_TGACv1_425241_AA1392650.1 | TaCslE3     | CslE3                      |
| 71  | TRIAE_CS42_5BL_TGACv1_406235_AA1342600.1 | TaCslE3     | CslE3                      |
| 72  | TRIAE_CS42_5AL_TGACv1_376126_AA1232370.1 | TaCslE3     | CslE3                      |
| 73  | TRIAE_CS42_5BS_TGACv1_425241_AA1392650.1 | TaCslE3     | CslE3                      |
| 74  | TRIAE_CS42_5BL_TGACv1_406235_AA1342600.1 | TaCslE3     | CslE3                      |
| 75  | TRIAE_CS42_5AL_TGACv1_376126_AA1232370.1 | TaCslE3     | CslE3                      |
| 76  | TRIAE_CS42_5BS_TGACv1_425241_AA1392650.1 | TaCslE3     | CslE3                      |
| 77  | TRIAE_CS42_5BL_TGACv1_406235_AA1342600.1 | TaCslE3     | CslE3                      |
| 78  | TRIAE_CS42_5AL_TGACv1_376126_AA1232370.1 | TaCslE3     | CslE3                      |
| 79  | TRIAE_CS42_5BS_TGACv1_425241_AA1392650.1 | TaCslE3     | CslE3                      |
| 80  | TRIAE_CS42_5BL_TGACv1_406235_AA1342600.1 | TaCslE3     | CslE3                      |
| 81  | TRIAE_CS42_5AL_TGACv1_376126_AA1232370.1 | TaCslE3     | CslE3                      |
| 82  | TRIAE_CS42_5BS_TGACv1_425241_AA1392650.1 | TaCslE3     | CslE3                      |
| 83  | TRIAE_CS42_5BL_TGACv1_406235_AA1342600.1 | TaCslE3     | CslE3                      |
| 84  | TRIAE_CS42_5AL_TGACv1_376126_AA1232370.1 | TaCslE3     | CslE3                      |
| 85  | TRIAE_CS42_5BS_TGACv1_425241_AA1392650.1 | TaCslE3     | CslE3                      |
| 86  | TRIAE_CS42_5BL_TGACv1_406235_AA1342600.1 | TaCslE3     | CslE3                      |
alignments were analysed for the presence of conserved motifs (D, D, DXD, QXXRW) of the GT2 superfamily. Conserved patterns of aligned sequences were highlighted using the sequence manipulation suite: Color align conservation (http://www.bioinformatics.org/sms2/color_align_cons.html) [29]. The conserved domains were predicted using CCD database (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) [22, 30, 31]. Wheat Csl genes were named based on their sequence identity, coverage, presence of conserved domains and motifs similar to those of the previously identified rice Csl genes. The number of genes in a subfamily exceeded that of rice, the additional genes were given new names. Because of the resemblance of CslD genes with CesA genes and their probable role in cellulose synthesis, we specifically focused on the TaCslD subfamily. Gene structures and intron evolution of TaCslD members were predicted using the gene structure display server 2.0 (http://gsds.cbi.pku.edu.cn/) using the genomic and cDNA sequences.

Evolutionary relationships of Csl genes
A total of 215 CSL proteins from Arabidopsis, maize, rice and wheat were aligned using MAAFT (v1.3.6) [32]. Sequences that did not extend over the conserved core region were removed. Positions where more than 40% of the sequences contained a gap were also removed. The phylogeny and 1000 bootstrap replications of these sequences was inferred using Seqboot (v3.696) [33] and FastTree (v2.1.10) implemented on the Guillimin cluster [34].

The phylogeny of the CslD subfamily was also determined separately from Arabidopsis, Brachypodium, maize, rice and wheat. For phylogenetic analysis, the amino acid sequences of CSL proteins were aligned using MUSCLE and their evolutionary history was inferred using Neighbor-Joining methods [35]. The tree was drawn to scale, with branch lengths being equivalent to the evolutionary distances used to infer the phylogenetic tree. Evolutionary distances were computed with a Poisson correction and are given as the number of amino acid substitutions per site. The rate of variation among sites was modeled with a gamma distribution (shape parameter = 1) and all positions containing gaps and missing data were removed. Evolutionary analyses were conducted in MEGA6 [36].

Table 1 Homeologous copies of the bread wheat Csl genes (Continued)

| No. | Ensembl ID | Gene name | Corresponding gene in rice |
|-----|------------|-----------|----------------------------|
| 87  | TRIAE_CS42_3DL_TGACv1_433902_AA1424880.1 | TaCslF7_3DL | CslF7 |
| 88  | TRIAE_CS42_3AL_TGACv1_374191_AA1193100.1 | TaCslF7_3AL | CslF7 |
| 89  | TRIAE_CS42_2BS_TGACv1_148916_AA0495580.1 | TaCslF8_2BS | CslF8 |
| 90  | TRIAE_CS42_2DS_TGACv1_178471_AA0596060.1 | TaCslF8_2DS | CslF8 |
| 91  | TRIAE_CS42_2AS_TGACv1_112322_AA0335280.1 | TaCslF8_2AS | CslF8 |
| 92  | TRIAE_CS42_2AS_TGACv1_112322_AA0335290.1 | TaCslF9_2AS | CslF9 |
| 93  | TRIAE_CS42_2BS_TGACv1_147667_AA0486240.1 | TaCslF9_2BS | CslF9 |
| 94  | TRIAE_CS42_2DS_TGACv1_177329_AA0573830.1 | TaCslF9_2DS | CslF9 |
| 95  | TRIAE_CS42_1BS_TGACv1_641498_AA2096480.1 | TaCslF10 | CslF9 |
| 96  | TRIAE_CS42_2BS_TGACv1_049866_AA0163180.1 | TaCslF10_1BS | CslF9 |
| 97  | TRIAE_CS42_2AL_TGACv1_094351_AA0296300.1 | TaCslH1_2AL | CslH1/2 |
| 98  | TRIAE_CS42_2DL_TGACv1_158387_AA0517170.1 | TaCslH1_2DL | CslH1/2 |
| 99  | TRIAE_CS42_2BL_TGACv1_129372_AA0380770.1 | TaCslH1_2BL | CslH1/2 |
| 100 | TRIAE_CS42_2B_TGACv1_221049_AA0728260.1 | TaCslH2_2B | Csl |
| 101 | TRIAE_CS42_2DS_TGACv1_273502_AA0931770.1 | TaCslH2_2DS | Csl |
| 102 | TRIAE_CS42_3BS_TGACv1_271739_AA0907200.1 | TaCslH3_3BS | Csl |
| 103 | TRIAE_CS42_3AS_TGACv1_212952_AA0704280.1 | TaCslH3_3AS | CslH3 |
| 104 | TRIAE_CS42_3BL_TGACv1_222234_AA0760340.1 | TaCslH3_3B | Csl |
| 105 | TRIAE_CS42_3DS_TGACv1_272297_AA0918580.1 | TaCslJ1_3DS | Csl |
| 106 | TRIAE_CS42_3AS_TGACv1_210908_AA0681280.1 | TaCslJ1_3AS | Csl |
| 107 | TRIAE_CS42_3B_TGACv1_221705_AA0747940.1 | TaCslJ2_3B | Csl |
| 108 | TRIAE_CS42_3DS_TGACv1_272756_AA0924850.1 | TaCslJ2_3DS | Csl |
RNA-seq expression analysis

Publicly available RNA-seq data generated from bread wheat (var. Chinese Spring) was used to study the expression of newly identified wheat Csl genes. The data were compiled from five different wheat tissues (spike, leaf, stem, root, and grain) collected at seedling, vegetative and reproductive stages of development [37]. The relative expression of each TaCsl subfamily was presented as a heat map generated from the relative abundance of transcripts (per 10 million reads) for each gene.
Table 2 Splice variants of the bread wheat Csl genes

| Ensembl gene ID | Gene name      | Predicted amino acids | Spliced exon/intron       | Status         |
|-----------------|----------------|------------------------|---------------------------|----------------|
| TRIA_C542_6BS_TGACv1_513375_AA1639370.1 | TaCslA1_6BS       | 581                    | –                         | Wild type      |
| TRIA_C542_6BS_TGACv1_513375_AA1639370.2 | TaCslA1_6BS       | 390                    | Exon 1 and 2              | Exon skipping  |
| TRIA_C542_6BS_TGACv1_513376_AA1639390.1 | TaCslA4_6BS       | 528                    | –                         | Wild type      |
| TRIA_C542_7AS_TGACv1_569190_AA1809650.1 | TaCslA3_7AS       | 551                    | –                         | Wild type      |
| TRIA_C542_7AS_TGACv1_569190_AA1809650.2 | TaCslA3_7AS       | 380                    | Exon 7, 8 and 9           | Exon skipping  |
| TRIA_C542_7AS_TGACv1_569190_AA1809650.3 | TaCslA3_7AS       | 503                    | Exon 9                    | Exon skipping  |
| TRIA_C542_7DL_TGACv1_602617_AA1962870.1 | TaCslA10_7DL      | 515                    | –                         | Wild type      |
| TRIA_C542_7DL_TGACv1_602617_AA1962870.2 | TaCslA10_7DL      | 555                    | Intron 8                  | Intron retention |
| TRIA_C542_3DL_TGACv1_249033_AA0835410.1 | TaCslA6_3DL       | 524                    | –                         | Wild type      |
| TRIA_C542_3DL_TGACv1_249033_AA0835410.2 | TaCslA6_3DL       | 572                    | Intron 1                  | Intron retention |
| TRIA_C542_3B_TGACv1_221079_AA0729630.1 | TaCslA6_3B        | 571                    | –                         | Wild type      |
| TRIA_C542_3B_TGACv1_221079_AA0729630.2 | TaCslA6_3B        | 538                    | Exon 2                    | Exon skipping  |
| TRIA_C542_5BL_TGACv1_404820_AA1311790.1 | TaCslC10_5BL      | 712                    | –                         | Wild type      |
| TRIA_C542_5BL_TGACv1_404820_AA1311790.2 | TaCslC10_5BL      | 468                    | Exon 5                    | Alternative 5' site |
| TRIA_C542_5BL_TGACv1_404820_AA1311790.3 | TaCslC10_5BL      | 504                    | Exon 1                    | Exon skipping  |
| TRIA_C542_5DL_TGACv1_435778_AA1454840.1 | TaCslC10_5DL      | 708                    | –                         | Wild type      |
| TRIA_C542_5DL_TGACv1_435778_AA1454840.2 | TaCslC10_5DL      | 502                    | Exon 1                    | Exon skipping  |
| TRIA_C542_5AL_TGACv1_374268_AA1195590.3 | TaCslC10_5AL      | 703                    | –                         | Wild type      |
| TRIA_C542_5AL_TGACv1_374268_AA1195590.4 | TaCslC10_5AL      | 496                    | Exon 5                    | Alternative 5' site |
| TRIA_C542_5AL_TGACv1_374268_AA1195590.1 | TaCslC10_5AL      | 501                    | Exon 5                    | Exon skipping  |
| TRIA_C542_3DL_TGACv1_251593_AA0882850.1 | TaCslC1_3DL       | 704                    | –                         | Wild type      |
| TRIA_C542_3DL_TGACv1_251593_AA0882850.2 | TaCslC1_3DL       | 493                    | Exon 5                    | Exon skipping  |
| TRIA_C542_3DL_TGACv1_251593_AA0882850.3 | TaCslC1_3DL       | 679                    | Exon 1                    | Alternative 3' site |
| TRIA_C542_3AL_TGACv1_197197_AA0665370.1 | TaCslC1_3AL       | 704                    | –                         | Wild type      |
| TRIA_C542_3AL_TGACv1_197197_AA0665370.2 | TaCslC1_3AL       | 560                    | Exon 5                    | Alternative 3' site |
| TRIA_C542_3AL_TGACv1_197197_AA0665370.3 | TaCslC1_3AL       | 679                    | Exon 5                    | Alternative 5' site |
| TRIA_C542_6AL_TGACv1_471004_AA1500600.2 | TaCslE2_6AL       | 677                    | –                         | Wild type      |
| TRIA_C542_6AL_TGACv1_471004_AA1500600.3 | TaCslE2_6AL       | 737                    | Intron 8                  | Intron retention |
| TRIA_C542_6AL_TGACv1_471004_AA1500600.4 | TaCslE2_6AL       | 635                    | Exon 4                    | Alternative 5' site |
| TRIA_C542_6AL_TGACv1_471004_AA1500600.1 | TaCslE2_6AL       | 626                    | Exon 4                    | Alternative 5' site |
| TRIA_C542_6AL_TGACv1_471004_AA1500600.2 | TaCslE2_6AL       | 737                    | Intron 8                  | Intron retention |
| TRIA_C542_6AL_TGACv1_471004_AA1500600.3 | TaCslE2_6AL       | 635                    | Exon 4                    | Alternative 5' site |
| TRIA_C542_6AL_TGACv1_471004_AA1500600.4 | TaCslE2_6AL       | 626                    | Exon 4                    | Alternative 5' site |
| TRIA_C542_2DS_TGACv1_177641_AA0581710.1 | TaCslE2_2DS       | 847                    | –                         | Wild type      |
| TRIA_C542_2DS_TGACv1_177641_AA0581710.2 | TaCslE2_2DS       | 735                    | Exon 2                    | Alternative 3' site |
| TRIA_C542_2DS_TGACv1_179076_AA0604160.1 | TaCslF4_2DS       | 783                    | –                         | Wild type      |
| TRIA_C542_2DS_TGACv1_179076_AA0604160.2 | TaCslF4_2DS       | 970                    | Exon 1                    | Exon skipping  |
| TRIA_C542_2BS_TGACv1_147667_AA0486240.1 | TaCslF9_2BS       | 827                    | –                         | Wild type      |
| TRIA_C542_2BS_TGACv1_147667_AA0486240.2 | TaCslF9_2BS       | 796                    | Exon 1                    | Exon skipping  |
| TRIA_C542_5BL_TGACv1_409916_AA1366600.1 | TaCslF7_5BL       | 745                    | –                         | Wild type      |
| TRIA_C542_5BL_TGACv1_409916_AA1366600.2 | TaCslF7_5BL       | 815                    | Intron 2                  | Intron retention |
| TRIA_C542_5AL_TGACv1_374191_AA1193100.1 | TaCslF7_5AL       | 792                    | –                         | Wild type      |
| TRIA_C542_5AL_TGACv1_374191_AA1193100.2 | TaCslF7_5AL       | 807                    | Intron 1                  | Intron retention |
using wheat expression browser powered by expVIP (http://www.wheat-expression.com).

Results

Identification and classification of Csl gene family members in bread wheat

Database searches for bread wheat using conserved pfam motifs PF00535 and PF03552, which are specific to the GT2 superfamily, resulted in the identification of 108 cellulose synthase-like (TaCsl) genes (Table 1). Two to three homologous copies of each gene from the A, B and D genomes were common. The identified genes were named following the nomenclature of rice, which shares synteny with wheat. To avoid the complexity of the nomenclature, a suffix corresponding to the chromosome number and the specific wheat genome identifier (A, B, or D) has been used for each gene name [7]. For example, the first gene of subfamily CslA; CslA1 on the long arm of chromosome 1 of genomes A, B, and D is named as TaCslA1_1AL, TaCslA1_1BL, and TaCslA1_1DL, respectively.

An unrooted neighbor-joining (NJ) tree for the 215 derived Csl proteins from Arabidopsis, maize, rice and wheat is shown in Fig. 1. TaCsl proteins grouped into seven subfamilies: TaCslA (32 proteins), TaCslC (13 proteins), TaCslD (12 proteins), TaCslE (10 proteins), TaCslF (29 proteins), TaCslH (8 proteins), and TaCslI (4 proteins) (Fig. 2). The TaCslA and TaCslC subfamilies were closely related as shown by their taxonomic distribution and phylogenies. As expected, these subfamilies were conserved across the plant species. Although TaCslD is present in all the plant species whereas TaCslI is specific to grasses, their proximity to each other suggests a common origin [12]. Among the sequences common to both dicots and grasses, subfamily CslA appeared to be the most divergent between these two groups of plants. Whereas the sequences within the subfamilies CslC and CslD were interspersed between Arabidopsis and grasses, all the subfamily CslA sequences of Arabidopsis clustered together, separately from the grass CslA sequences. Proximity of the CslB and CslH subfamilies points to their common origin before the separation of grasses from dicots. Similarly, CslG and CslJ apparently had a common origin.

Splice variants of Csl genes

Twenty two of the 108 genes appeared to encode two or more proteins because of the presence of alternative splicing sites, as predicted by Ensembl database, which would result in 137 probable Csl protein products (Table 2). Splice variants were predicted in all the subfamilies of the TaCsl genes except TaCslD (Table 2). In the subfamily TaCslA, 6 genes alternatively spliced to form 13 putative proteins whereas in the subfamily TaCslC, 5 genes were alternatively spliced resulting in 14 putative proteins. Similarly, for the subfamilies TaCslE and TaCslF, alternative splicing resulted in 7 and 10 splice variants, respectively. Alternative splicing of 1 and 2 genes respectively generated 3 and 4 putative proteins in the CslH and CslI subfamilies (Fig. 2). More than half (51%) of the splice variants stemmed from exon skipping, ~24% from alternative 5′ and 3′ splice sites, and the rest, ~24%, from intron retention (Table 2).

Conserved motifs and domains

All predicted TaCsl proteins contain either the pfam glycosyltransferase family 2_3 (GT) domain (PF13641) or the cellulose_synt domain (PF03552), considered to be the signature domains of the GT2 superfamily [12, 26]. Subfamilies TaCslA and TaCslC contained GT 2_3, and CslD, CslE, CslF, CslH, and CslJ contained the cellulose_ - synt domain (Fig. 2). All the TaCsl transplanted products contained the motifs D, DXD, D and QXXRW except eight truncated genes that lacked some of these motifs apparently because of the missing sequence (TaCslA7_2DS, TaCslD4_1BS, TaCslD4_5BS, TaCslF2_7BL, TaCslF6_7AL, TaCslF6_7DL, TaCslH3_3AS, TaCslH2_3B). Rice CesA10, 11 and CslH3 also contained only the DXD but lacked the D and QXXRW motifs [38]. The variable amino acids in the conserved motifs DXD and QXXRW were diverse in different subfamilies of Csl genes, for example, for TaCslA (DMD, QQH/FRW), TaCslC (DMD, QQHRW), TaCslD (DCD, QVLRW), TaCslE (DCD, QHKRW), TaCslF (DC/ GD, QV/L/VRW), TaCslH (DCD QF/YKRW), TaCslI

Table 2 Splice variants of the bread wheat Csl genes (Continued)

| Ensembl gene ID | Gene name | Predicted amino acids | Spliced exon/introns | Status          |
|-----------------|-----------|-----------------------|----------------------|-----------------|
| TRIAE_CS42_2AL_TGACv1_094351_AA0296300.1 | TaCslH1_2AL | 737                   | Exon 9               | Wild type       |
| TRIAE_CS42_2AL_TGACv1_094351_AA0296300.2 | TaCslH1_2AL | 660                   | Exon skipping        | Wild type       |
| TRIAE_CS42_2AL_TGACv1_094351_AA0296300.3 | TaCslH1_2AL | 480                   | Exon 6, 7, 8 and 9   | Exon skipping   |
| TRIAE_CS42_3AS_TGACv1_210908_AA0681280.1 | TaCslJ3_3AS | 738                   | Exon 4               | Intron retention|
| TRIAE_CS42_3AS_TGACv1_210908_AA0681280.2 | TaCslJ3_3AS | 766                   | Intron 1             | Intron retention|
| TRIAE_CS42_3DS_TGACv1_272756_AA0924850.1 | TaCslJ3_3DS | 609                   | Wild type            |                 |
| TRIAE_CS42_3DS_TGACv1_272756_AA0924850.2 | TaCslJ3_3DS | 734                   | Wild type            |                 |
Phylogenetic analysis of the CslD subfamily

The evolutionary history of the CslD subfamily from Arabidopsis, Brachypodium, rice, maize and wheat was inferred using the Neighbor-Joining method, in MEGA6 [36], after grouping the orthologs from various species into different clades (Fig. 3). Rice Csl genes were used as reference because their complete nomenclature is well documented. All the genes grouped into three clades. The first clade contained CslD2 and CslD1 genes from rice and their orthologs from the remaining species. The three homeologous genes of wheat branched together with OsCslD1; wheat genes under this clade were named TaCslD1_1AL, TaCslD1_1BL, and TaCslD1_1DL. The second clade contained two subgroups with the orthologs of rice genes CslD3 and CslD5 from different species. The genes in the first subgroup were named TaCslD3_2AS, TaCslD3_2BS, and TaCslD3_2DS, and those of the second subgroup TaCslD5_7AL, TaCslD5_7BL, and TaCslD5_7DL. The last clade was composed of the orthologs of the rice CslD4 and wheat genes TaCslD4_5BS, TaCslD4_1BS and TaCslD4_5DS. Here we found only two homeologs of TaCslD4, but a gene from the 1BS genome (TaCslD4_1BS) of wheat grouped together with TaCslD4 genes (bootstrap = 1000), pointing to a translocation from its original A genome (Table 1). This gene shared sequence identity of 85% with TaCslD4_5BS at the amino acid level. OsCslD genes shared 73–86% sequence identity with the corresponding wheat orthologs.

Gene structure and intron evolution of TaCslD subfamily

The 12 TaCslD genes identified from bread wheat ranged in size from 1519 to 5864 bp. The TaCslD4_1BS gene was the shortest and TaCslD1_1AL was the longest. Homeologous copies of all the genes shared sequence identity ranging from 87 to 94% at the nucleotide level. The variation in size among different genes was primarily because of the number and length of introns but also because of a lack of the complete sequences in the database (Fig. 4). The number of introns in all the genes varied from 2 to 4. Two homeologs: TaCslD1_1AL and TaCslD1_1BL each contained three introns whereas, a third homeolog (TaCslD1_1DL) had four. The genes TaCslD3, TaCslD4 and their homeologs contained three introns each, except TaCslD4_1BS with only two introns. TaCslD5 and its homeologs also had two introns each. For the phases of introns, the genes from the TaCslD subfamily exhibited variable patterns of distribution. Introns 1, 2 and 3 of TaCslD1_1AL, TaCslD1_1BL and TaCslD1_1DL were in 2, 0, and 0 phase whereas the 4th intron of TaCslD1_1DL was in 0 phase. Introns 1 and 2 of TaCslD3_2AS, TaCslD3_2BS and TaCslD3_2DS both were in 0
The third intron of these genes was in phase 2, 1 and 2 respectively. The genes TaCslD4_5BS, TaCslD4_5DS, TaCslD5_7AL, TaCslD5_7BL and TaCslD5_7DL had intron 1 and 2 in phases 2 and 0, respectively, and the third intron of TaCslD4_5BS and TaCslD4_5DS was in phase 0 and 2, respectively. TaCslD4_1BS had introns 1 and 2 in phases 1 and 0. The largest proportion of introns (60%) of all the genes was in phase 0, followed by phase 2 (34%) with a few in phase 1 (6%).

Expression analysis of TaCsl genes from bread wheat
Publicly available RNA-Seq datasets were used to analyse the expression of TaCsl genes over three developmental stages and different tissues of wheat including root, stem, leaf, spike, and grain. Expression data were available for 32 of the TaCslA genes. Two genes (TaCslA1_6AS and TaCslA1_6BS) were expressed in all the tissues except reproductive stem and leaves. Four genes (TaCslA5_2BS, TaCslA5_2DS, TaCslA6_3B, and TaCslA6_3AL) were expressed moderately. TaCslA9 gene was highly expressed in the leaf tissue from the reproductive stage while the transcript abundance of the remaining genes was low (Fig. 5). TaCslC subfamily genes, with the exception of TaCslC3, TaCslC9 and two homeologs of TaCslC10, were expressed highly in root and spike tissues. Two genes, TaCslC1 and TaCslC7 and their homeologs displayed moderate to high expression in all the tissues at seedling and vegetative stage. One gene (TaCslC10_5DL) exhibited moderate to high expression in all the tissues studied except reproductive stem and grain (Fig. 6). Expression of most of the genes of the TaCslD subfamily ranged from moderate to high in the spike and root tissues but was very low in all the other tissues (Fig. 7). Three of the 10 TaCslE subfamily genes (TaCslE2_6AL, TaCslE2_6BL and TaCslE3) were expressed from moderate to high levels in all the tissues. The remaining genes were expressed at a very low level in all the tissues (Fig. 8). A mixed pattern of expression was observed in the large TaCslF subfamily. Three genes (TaCslF6_7AL, TaCslF6_7BL and TaCslF6_7DL) were highly expressed in all the tissues except the leaves at the reproductive stage. Two genes (TaCslF4_2BS and TaCslF4_2DS) were highly expressed in the stem tissue, but only at a low or moderate level in all other tissues. All other genes expressed at low or moderate levels in one or more tissues (Fig. 9). In the TaCslH subfamily, one of the eight genes, TaCslH1_2BS, was expressed from moderate to high levels in the leaf, stem and spike tissues. The remaining genes were expressed from low to moderate levels in all the tissues (Fig. 10). Three out of four members of the subfamily TaCslI were expressed from low to moderate levels in the leaf and root tissues while one gene (TaCslI1_3DS) was poorly expressed in all the tissues studied (Fig. 10).

Discussion
Grass cell walls contain 20–40% non-cellulosic polysaccharides. The proportion and composition of these polysaccharides varies in different plant species [39]. After the first report demonstrating the β-glucan synthase activity in a Csl-encoded protein was published [15], several members of the Csl gene family have been reported to be involved in the formation of the backbone of the hemicellulosic polysaccharides [16, 18, 19, 26, 38, 40, 41]. As information on the identity of the Csl genes in wheat was lacking, we undertook this study to fill this gap.

We retrieved 108 TaCsl genes from wheat using two conserved domains, PF00535, and PF03552, which were previously shown to be present in the derived proteins of all the Csl genes [12]. These genes include homeologs from A, B and D genome of bread wheat. Similar patterns of homeologous genes were found for FLOWERING LOCUS T (FT), Pairing homeologous 1 (Ph1) and ADP-
glucose pyrophosphorylase (AGPase) gene families of hexaploid wheat. Approximately, a quarter of the identified Csl genes were predicted to be alternatively spliced, possibly contributing to the diversity of encoded enzymes. A recent study suggested that alternative splicing was common in plants and accounted for about 20% of the loci transcribed in the leaf and spike tissues of Aegilops tauschii. In the case of germinating barley embryos, 14–20% of intron-containing genes were alternatively spliced [42]. This phenomenon, apparently meant to increase the fitness of an organism, has not thus far been reported for the Csl genes from other species [43].

The TaCsl genes were distributed across all the wheat chromosomes except one, chromosome 4 (Fig. 11). A similar trend of Csl gene distribution was observed in barley [9, 44, 45]. More than half the TaCsl genes were located on only two chromosomes: 2 (32%) and 3 (22%). This suggests hyper-multiplication of the Csl genes on these chromosomes although the reasons for this phenomenon are unknown. It appears, though, that cis duplication of the Csl genes was favored over trans duplication in wheat. Five of the nine CslF genes in barley were located on chromosome 2H [40]. In fact, the barley CslF gene was assigned its role in mixed-
linked glucan (MLG) formation via syntenic orthology with rice long before the barely genome sequence became available [40]. A detailed analysis of the rice syntenic region corresponding to a known QTL for MLG from barley, which had been published previously, initially led to the breakthrough of the role of CslF in the formation of this polysaccharide [40]). A similar cluster of CslF genes was also detected in the conserved syntenic regions of Brachypodium and sorghum on chromosomes 1 and 2, respectively [9].

The observation that only half of genes from the subfamily CslA were expressed at varying levels in the studied tissues suggests that the apparently silent genes may provide a backup under stressful conditions. Alternatively, they may express only transiently in specialized cells or cell parts at levels too low to be detected by the method used to study expression. The first biochemical evidence for the relationship of CslA genes with mannan synthase activity came from the expression of a guar CSLA cDNA in soybean somatic embryos [15].
Fig. 8 Heat map of the expression profiling of wheat TaCslE genes at seedling, vegetative and reproductive stages. RNA-seq data were obtained from root, leaf, stem, spike and grain of Chinese spring cultivar. The respective transcripts per 10 million values were used to construct heat map with scale bar showing expression of the genes.

| Gene Name     | Seedling | Vegetative | Reproductive |
|---------------|----------|------------|--------------|
|               | Root     | Leaf       | Root         | Leaf | Stem | Spike | Root | Leaf | Stem | Spike | Grain |
| TaCslE1_5BL   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslE1_5DL   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslE2_6AL   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslE2_6BL   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslE2_6DL   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslE3       |          |            |              |      |      |        |      |      |      |        |       |
| TaCslE4_6DS   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslE6_5AL   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslE6_5BL   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslE6_5DL   |          |            |              |      |      |        |      |      |      |        |       |

Fig. 9 Heat map of the expression profiling of wheat TaCslF genes at seedling, vegetative and reproductive stages. RNA-seq data were obtained from root, leaf, stem, spike and grain of Chinese spring cultivar. The respective transcripts per 10 million values were used to construct heat map with scale bar showing expression of the genes.

| Gene Name     | Seedling | Vegetative | Reproductive |
|---------------|----------|------------|--------------|
|               | Root     | Leaf       | Root         | Leaf | Stem | Spike | Root | Leaf | Stem | Spike | Grain |
| TaCslF1_2AL   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF1_2BL   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF1_2DL   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF2_7AL   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF2_7BL   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF2_7DL   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF3_2AS   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF3_2BS   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF3_2DS   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF4_2BS   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF4_2DS   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF5_2AS   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF5_2BS   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF5_2DS   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF6_7AL   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF6_7BL   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF6_7DL   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF7_5AL   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF7_5BL   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF7_5DL   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF8_2AS   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF8_2BS   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF8_2DS   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF9_2AS   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF9_2BS   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF9_2DS   |          |            |              |      |      |        |      |      |      |        |       |
| TaCslF10      |          |            |              |      |      |        |      |      |      |        |       |
Subsequent studies in insect cells demonstrated the role of CslA family members in the glucomannan synthases [16, 46]. Reverse genetic and biochemical approaches in Arabidopsis and Dendrobium officinale have also allowed association of certain CslA genes with glucomannan biosynthesis [41, 47]. A recent study in wheat suggested the involvement of a gene from the CslA subfamily in the development of tillers, cell wall composition and stem strength. This study further suggested the probable role of CslA gene transcript levels in carbon partitioning throughout the plant [48].

For the subfamilies TaCslC and TaCslD, most of the genes were relatively highly expressed in the root and spike tissues during the vegetative as well as reproductive phases. Heterologous expression in Pichia revealed that the CslC-encoded enzymes made β-1,4-glucan, the backbone of xyloglucan [19]. The CslD subfamily is conserved in all land plants and is most closely related to the CesA gene family with 40–50% sequence similarity at the amino acid level [49]. Similar to CesAs, the CslD subfamily is ubiquitous in all plant genomes examined to date, unlike other, taxa-specific Csl subfamilies [50]. Previous reports also showed the involvement of certain members of the CslD subfamily in tip growth, for example development of root hairs and pollen tube elongation [51, 52], normal plant growth [50, 53], and meristem morphology [53, 54]. More recently, their role in resistance against biotic stresses has been described [55]. Adding to this discussion, our in silico expression analysis suggests the involvement of certain TaCslD genes in spike development. This suggestion is supported by the observation that a mutant, slender leaf 1 (sle1), which encodes the CSLD4 protein in rice, reduces the number and width of spikelets in the panicle [56].

Two groups of Csl genes, CslF and CslH, have evolved independently in grasses [57]. A third group CslI, originally believed to be specific to grasses, was recently identified in some dicots [11, 13]. Although TaCslF6 gene showed higher expression in all the studied tissues except the leaf tissue from reproductive stage, it was the
only member of the TaCslF subfamily which expressed highly in the grain tissue. Several studies have demonstrated the functional role of CslF6 and CslH in the synthesis of MLG [18, 44, 58, 59]. Only one member of all the genes in these families, CslF6, was expressed in the grain, suggesting that it was responsible for MLG formation. MLG is a desirable polysaccharide as a dietary fiber but undesirable for the brewery industry because it causes haze in beer. It should be possible to select natural variants for the expression of the CslF6 gene to select for an increased or reduced MLG content depending upon the target market for the grain.

Differential expression patterns were observed among homeologous copies from three different genomes of bread wheat, which agree with the previous studies reporting unequal contributions of the three genomes toward gene expression. Interestingly, the homeologous copies of TaCslID genes also differed from each other in terms of intron phase evolution, indicating structural and functional divergence of homeologous gene copies (Fig. 4). Most introns were present in phase 0, which is in accordance with previous findings showing an intron bias in favor of phase 0 [7, 60, 61]. The three homeologs of each gene were not observed for all the genes reported in this study. This could be because of the incomplete sequencing information or because of the elimination of the genes during the allopolyploidization of wheat.

Conclusions
We have identified 108 TaCsl genes in bread wheat and classified them into seven subfamilies (CslA, CslC, CslD, CslE, CslF, CslH, and CslI). Two or three homeoealleles were identified for most of the Csl genes. Although located on all the wheat chromosomes except chromosome 4, the Csl genes were especially concentrated on chromosomes 2 and 3, suggesting selective, localized duplication in cis phase. Only one of the 29 CslF genes, CslF6, was expressed in the grain, suggesting its role in mixed-linked glucan formation. Neither CslI nor CslH was expressed in the grain. Information in this report will be helpful in designing experiments to alter wall composition in wheat for improving grain quality, culm strength, or culm composition for biofuels.

Abbreviations
CesA: Cellulose synthase; Csl: Cellulose synthase-like; GT: Glycosyltransferase; MLG: Mixed-linked glucan

Acknowledgements
This work was supported by the CGIAR’s Consortium Research Program WHEAT (KSD), Canada Foundation for Innovation (CFI), the ministère de l’Économie, de la science et de l’innovation du Québec (MESI) and the Fonds de recherche du Québec - Nature et technologies (FRQ-NT) (RB), and Natural Sciences and Engineering Research Council of Canada through discovery program (JS).

Computations were made on the supercomputer Guillimin from McGill University, managed by Calcul Québec and Compute Canada. The operation of this supercomputer is funded by the Canada Foundation for Innovation (CFI), the ministère de l’Économie, de la science et de l’innovation du Québec (MESI) and the Fonds de recherche du Québec - Nature et technologies (FRQ-NT).

Funding
Natural sciences and engineering research council of Canada.

Availability of data and materials
Yes, all the data are included in the supplement already.

Authors’ contributions
SK extracted the sequences, analyzed them, and wrote the paper; KSD conceived the project along with JS, analyzed the data, wrote parts of the paper, and edited the manuscript; RB carried out the phylogenetic analysis and constructed the phylogenetic tree; JS conceived and supervised the project, and helped write the paper. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
1Department of Plant Science, McGill University, Sainte Anne de Bellevue, QC, Canada. 2International Maize and Wheat Improvement Center (CIMMYT), El Batán, Texcoco, Estado de México, Mexico. 3Institute of Parasitology, McGill University, Sainte Anne de Bellevue, Montreal, QC, Canada.

Received: 20 April 2017 Accepted: 26 October 2017

References
1. Pauly M, Keegstra K. Cell-wall carbohydrates and their modification as a resource for biofuels. Plant J. 2008;54(4):559–68.
2. Sandhu APS, Randhawa GS, Dhugga KS. Plant cell wall matrix polysaccharide biosynthesis. Mol Plant. 2009;2(5):840–50.
3. Sorek N, Yeats TH, Sazemneyi H, Youngs H, Somerville CR. The implications of Lignocellulosic biomass chemical composition for the production of advanced biofuels. Bioscience. 2014;64(3):192–201.
4. Pauly M, Gilles S, Liu L, Mansoori N, de Souza A, Schultink A, Xiong G. Hemicellulose biosynthesis. Planta. 2013;238(4):527–42.
5. Richmond TA, Somerville CR. The cellulose synthase superfamily. Plant Physiol. 2000;124(2):495–8.
6. Rai KM, Thu SW, Balasubramanian VK, Cobos CJ, Disasa T, Mendu V. Identification, characterization, and expression analysis of Cell Wall related genes in Sorghum Bicolor (L) Moench, a food, fodder, and biofuel crop. Front Plant Sci. 2016;1287.

Additional files
Additional file 1: Figure S1. FASTA sequences of CSL proteins used for the phylogenetic analysis. (PDF 453 kb)

Additional file 2: Figure S2. List of Csl subfamily genes, their protein sizes (number of amino acids), and multiple protein sequence alignments for different subfamilies. The conserved motifs (D, D, DXD, QXXRW) diagnostic of CSL proteins are highlighted with red boxes for each of the subfamilies. (PDF 465 kb)
7. Kaur S, Dhugga KS, Gill K, Singh J. Novel structural and functional motifs in cellulose synthase (CesA) genes of bread wheat (Triticum aestivum, L.). PLoS One. 2016;11(1):e0147046.

8. Hazen SP, Scott-Craig JS, Walton JD. Cellulose synthase-like genes of rice. Plant Physiol. 2002;128(3):36–40.

9. Schwerdt JG, Mackenzie K, Wright F, Dehme D, Wagner JM, Harvey AJ, Shirley NJ, Burton RA, Schreiber M, Halpin C. Evolutionary dynamics of the cellulose synthase gene superfamily in grasses. Plant Physiol. 2015;168(3):968–83.

10. Burton RA, Collins HM, Kibble NA, Smith JA, Shirley NJ, Jobling SA, Henderson M, Singh RR, Pettolino F, Wilson SM, et al. Over-expression of specific HvCesA cellulose synthase-like genes in transgenic barley increases the levels of cell wall (1,3;1,4)-beta-D-glucans and alters their fine structure. Plant Biotechnol J. 2011;9(2):117–35.

11. Farrokh N, Burton RA, Brownfield L, Hrmoova M, Wilson SM, Baic A, Fincher GB. Plant cell wall biosynthesis: genetic, biochemical and functional genomics approaches to the identification of key genes. Plant Biotechnol J. 2006;4(2):145–67.

12. Yin Y, Johns MA, Cao H, Rupani M. A survey of plant and algal genomes and transcriptomes reveals new insights into the evolution and function of the cellulose synthase superfamily. BMC Genomics. 2014;15(11).

13. Vogel JP, Gavin DF, Mockler TC, Schmutz J, Rokhsar D, Bevan MW, Barry K, et al. The genome of the perennial grass Brachypodium distachyon. Nature. 2010;463(7282):763–8.

14. Doblin MS, Pettolino FA, Wilson SM, Campbell R, Burton RA, Fincher GB. Plant cell wall biosynthesis: genetic, biochemical and functional genomics approaches to the identification of key genes. Plant Biotechnol J. 2006;4(2):145–67.

15. Dhugga KS, Barreiro R, Whitten S, Stecca K, Hazebroek J, Randhawa GS, Dhugga KS, Barreiro R, Whitten S, Stecca K, Hazebroek J, Randhawa GS. Darlingtonia californica encodes a beta-1,4 glucan synthase. Proc Natl Acad Sci U S A. 2007;104(20):8283–8.

16. Yin Y, Johns MA, Cao H, Rupani M. A survey of plant and algal genomes and transcriptomes reveals new insights into the evolution and function of the cellulose synthase superfamily. BMC Genomics. 2014;15(11).

17. Dhugga KS. Biosynthesis of non-cellulosic polysaccharides of plant cell walls. Phytochemistry. 2012;74:8–19.

18. Dhugga KS, Barreiro R, Whitten S, Stecca K, Hazebroek J, Randhawa GS, Dolan M, Kinney AJ, Tomes D, Nichols S. Sugar seed mannan synthase is a member of the cellulose synthase super gene family. Science. 2004;303(5656):363–6.

19. Liepman JC, Lerouxel O, Drakakaki G, Alonso AP, Liepman AH, Keegstra K, Cocuron JC, Lerouxel O, Drakakaki G, Alonso AP, Liepman AH, Keegstra K. The Arabidopsis thaliana CesA family encodes a beta-1,4 glucan synthase. Proc Natl Acad Sci U S A. 2007;104(20):8283–8.

20. Gupta PK, Mir RR, Mohan A, Kumar J. Wheat genomics: present status and future prospects. Int J Plant Genomics. 2008;2008:896451.

21. Burton RA, Wilson SM, Hrmoova M, Harvey AJ, Shirley NJ, Medhurst A, Stone BA, Newbigin EJ, Baic A, Fincher GB. Cellulose synthase-like CesA genes mediate the synthesis of cell wall (1,3;1,4)-beta-D-glucans. Science. 2006;311(5769):1940–2.

22. Consortium IWGS. A chromosome-based draft sequence of the hexaploid wheat (Triticum aestivum L.) genome. Nature. 2018;551(7679):56–60.

23. Covarrubias PR, Huelsenbeck JP, O’Loughlin EE, Kjer KD, Rambaut A, Larget B, knowlus P. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics. 2004;20(8):3552–6.

24. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30(12):2725–9.

25. dhugga Ks. Biosynthesis of non-cellulosic polysaccharides of plant cell walls. Phytochemistry. 2012;74:8–19.

26. Dhugga KS, Barreiro R, Whitten S, Stecca K, Hazebroek J, Randhawa GS, Dolan M, Kinney AJ, Tomes D, Nichols S. Sugar seed mannan synthase is a member of the cellulose synthase super gene family. Science. 2004;303(5656):363–6.

27. dhugga Ks. Biosynthesis of non-cellulosic polysaccharides of plant cell walls. Phytochemistry. 2012;74:8–19.

28. dhugga Ks. Biosynthesis of non-cellulosic polysaccharides of plant cell walls. Phytochemistry. 2012;74:8–19.

29. dhugga Ks. Biosynthesis of non-cellulosic polysaccharides of plant cell walls. Phytochemistry. 2012;74:8–19.
53. Li M, Xiong G, Li R, Cui J, Tang D, Zhang B, Pauly M, Cheng Z, Zhou Y. Rice cellulose synthase-like D4 is essential for normal cell-wall biosynthesis and plant growth. Plant J. 2009;60(6):1055–69.

54. Bernal AJ, Jensen JK, Harholt J, Sørensen S, Moller I, Blaukopf C, Johansen B, De Lotto R, Pauly M, Scheller HV. Disruption of ATCSLDS results in reduced growth, reduced xylan and homogalacturonan synthase activity and altered xylan occurrence in Arabidopsis. Plant J. 2007;52(5):791–802.

55. Douchkov D, Lueck S, Hensel G, Kumlehn J, Rajaraman J, Johrde A, Doblin MS, Beahan CT, Kopischke M, Fuchs R. 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 Phytol. 2016;212:421–33.

56. Yoshikawa T, Eguchi M, Hibara K-I, Ito J-I, Nagato Y. Rice SLENDER LEAF 1 gene encodes cellulose synthase-like D4 and is specifically expressed in M-phase cells to regulate cell proliferation. J Exp Bot. 2013;64(7):2049–61.

57. Burton RA, Collins HM, Kibble NA, Smith JA, Shirley NJ, Jobling SA, Henderson M, Singh RR, Pettolino F, Wilson SM. 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 Biotechnol J. 2011;9(2):1117–35.

58. Taketa S, Yuo T, Tomoka T, Tsumuraya Y, Inagaki Y, Haruyama N, Larroque O, Jobling SA. Functional characterization of barley betaglucanless mutants demonstrates a unique role for CslF6 in (1,3;1,4)-β-D-glucan biosynthesis. J Exp Bot. 2012;63(1):381–92.

59. Nemeth C, Freeman J, Jones HD, Sparks C, Pellny TK, Wilkinson MD, Dunwell J, Anderson AAM, Aman P, Guillon F, et al. Down-regulation of the CslF6 gene results in decreased (1,3;1,4)-β-D-glucan in endosperm of wheat. Plant Physiol. 2010;152(3):1209–18.

60. Lynch M. Intron evolution as a population-genetic process. Proc Natl Acad Sci U S A. 2002;99(9):6118–23.

61. Bhattachan P, Dong B. Origin and evolutionary implications of introns from analysis of cellulose synthase gene. J Syst Evol. 2017;55(2):142–8.