Cellulose content variation and underlying gene families in bread wheat

Simerjeet Kaur

Department of Plant Science

McGill University, Montreal

Canada

April 2017

A thesis submitted to the McGill University in partial fulfilment of the requirements of the

degree of Doctor of Philosophy

©Simerjeet Kaur (2017)

1

Table of Contents

Cellulose content variation and underlying gene families in bread wheat	1
LIST OF TABLES	7
LIST OF FIGURES	8
LIST OF ABBREVIATIONS	13
ABSTRACT	16
ACKNOWLEDGEMENT	20
PREFACE AND CONTRIBUTION OF THE AUTHORS	23
a. Preface	23
b. Contribution of the authors	24
Chapter I: General introduction	26
1.1 General hypothesis	29
1.2 General objectives	29
Chapter II. Literature review	
2.1 Future energy requirements	30
2.2 Lignocellulosic materials as bioethanol	30
2.3 Structure and composition of lignocellulose	31
2.3.1 Cellulose	31
2.3.2 Hemicellulose	32
2.3.3 Lignin	33
2.4 Biofuels and plant cell walls	33
2.5 Functional significance and synthesis of key cell wall components	34
2.5.1 Genetics of cellulose synthesis	34
2.5.2 Cellulose Synthase-Like (Csl) genes and their importance	
2.6 Wheat straw and its potential as biofuel	
2.7 Importance of wheat and its genetics	
2.8 Molecular markers in wheat	
2.8.1 Random Markers (RDMs)	40
2.8.2 Gene Target Markers (GTMs)	40
2.8.3 Functional Markers (FMs)	41
2.9 Comparative genomics	41
2.10 Functional genomics in wheat	42

2.10.1 Gene silencing approach through RNA interference	43
2.10.2 Virus-induced gene silencing in wheat	43
2.11 Genomics-integrated breeding	44
2.11.1 Genome-wide association (GWA)	45
2.11.2 Genomic selection (GS)	46
CONNECTING STATEMENT FOR CHAPTER III	48
Chapter III. Novel structural and functional motifs in <i>Cellulose synthase A</i> (CesA) genes of bread (<i>Triticum aestivum</i> , L.)	l wheat 49
3.1 Abstract	49
3.2 Introduction	50
3.2.1 Hypothesis	53
3.2.2 Objective I	53
3.2.3 Objective II.	53
3.3 Methods and materials	53
3.3.1 Identification of CesAs in wheat and their true orthologs from different species	53
3.3.2 Gene structure analysis	54
3.3.3 Protein structure and motif identification	54
3.3.4 Phylogenetic analysis	55
3.3.5 RNA-seq expression profiling of <i>TaCesA</i> genes	56
3.4 Results	57
3.4.1 Identification and mapping of CesA gene family in wheat	57
3.4.2 DNA sequence comparison of primary and secondary cell wall TaCesA genes	58
3.4.3 Evolution of introns in <i>TaCesA</i> gene family	59
3.4.4 Amino acid variability of predicted TaCESA proteins	59
3.4.5 New motifs distinguishing PCW CESAs from SCW CESAs	60
3.4.6 Conservation of motifs in monocots and dicots	61
3.4.7 Unique motifs conserved among the CESA orthologs from different species	61
3.4.8 Motifs differentiating CESAs from monocots and dicots	62
3.4.9 Phylogenetic analysis	63
3.4.10 RNA-seq analysis of <i>TaCesA</i> genes	63
3.5 Discussion	64
3.6 Conclusion	67
CONNECTING STATEMENT FOR CHAPTER IV	77

Chapter IV. Functional characterization of secondary cell wall specific <i>CesA4</i> gene in br virus-induced gene silencing (VIGS)	read wheat using
4.1 Abstract	
4.2 Introduction	79
4.2.1 Hypothesis	81
4.2.2 Objective I	81
4.2.3 Objective II.	81
4.3 Materials and methods	81
4.3.1 <i>TaCesA4</i> gene structure analysis	81
4.3.2 In silico expression analysis of TaCesA homoeologs	81
4.3.3 Preparation of VIGS-construct	82
4.3.4 In vitro transcription of VIGS plasmids and rub inoculation	82
4.3.5 RNA Isolation and cDNA synthesis	83
4.3.6 Real-time PCR	83
4.3.7 Estimation of cellulose content	84
4.3.8 Microscopic analysis of stem sections	85
4.3.9 Statistical analysis	85
4.4 Results	85
4.4.1 TaCesA4 gene structure and construct designing	85
4.4.2 Homoeolog specific expression of <i>TaCesA4</i>	86
4.4.3 Optimization of VIGS in <i>Chinese spring</i> (CS) wheat cultivar	87
4.4.4 Silencing of CesA4 gene in wheat	87
4.4.5 Analysis of cellulose content in VIGS treated plants	
4.4.6 Histological analysis of stem tissues	
4.5 Discussion	
CONNECTING STATEMENT FOR CHAPTER V	
Chapter V. Genome- wide association study reveals novel genes linked to natural variatic content in bread wheat (<i>Triticum aestivum</i> , L.)	on of cellulose
5.1 Abstract	
5.2 Introduction	
5.2.1 Hypothesis	
5.2.2 Objective I	
5.2.3 Objective II.	

5.3 Materials and methods	. 102
5.3.1 Plant material	. 102
5.3.2 Phenotypic analysis	. 103
5.3.3 Population structure and GWAS analysis	. 104
5.4 Results	. 105
5.4.1 Cellulose content	. 105
5.4.2 Principal component analysis and marker-trait associations	. 106
5.4.3 Gene identification	. 106
5.5 Discussion	. 107
5.6 Conclusion	. 110
CONNECTING STATEMENT FOR CHAPTER VI	. 116
Chapter VI. Genome-wide analysis of the <i>Cellulose synthase-like (Csl)</i> gene family in bread wheat (<i>Triticum aestivum</i> L.)	. 117
6.1 Abstract	. 117
6.2 Introduction	. 118
6.2.1 Hypothesis	. 120
6.2.2 Objective I	. 120
6.2.3 Objective II	. 120
6.3 Materials and methods	. 120
6.3.1 Data sources and sequence retrieval	. 120
6.3.2 Blast searches for wheat homologs	. 121
6.3.3 Protein structure and motif/domain identification	. 121
6.3.4 Evolutionary relationships of <i>Csl</i> genes	. 122
6.3.5 RNA-seq expression analysis	. 122
6.4. Results	. 123
6.4.1 Identification and classification of <i>Csl</i> gene family in bread wheat	. 123
6.4.2 Splice variants of <i>Csl</i> genes	. 124
6.4.3 Conserved motifs and domains	. 124
6.4.4 Phylogenetic analysis of the <i>CslD</i> subfamily	. 125
6.4.5 Gene structure and intron evolution of <i>TaCslD</i> subfamily	. 126
6.4.6 RNA-seq expression analysis of <i>TaCsl</i> genes from bread wheat	. 127
6.5 Discussion	. 128
6.6 Conclusion	. 131

CHAPTER VII. GENERAL DISCUSSION AND FUTURE STUDIES	147
7.1 General discussion	147
7.2 Future studies	152
VIII. APPENDIX	153
IX. REFERENCES	182

LIST OF TABLES

Table 3.1 CesA genes and their chromosomal locations in hexaploid wheat.

Table 3.2 Structures of the *TaCesA* genes for PCW and SCW synthesis.

Table 3.3 *TaCesA* genes and their orthologs from Arabidopsis, barley, maize, and rice involved in the formation of the primary cell wall (PCW) or secondary cell wall (SCW).

Table 4.1 Primers used for semi-qPCR, qRT-PCR and confirmation of VIGS construct.

Table 5.1 Regions of wheat genome showing significant associations with stem cellulose content

 variation based on GWAS.

Table 5.S1. Sequences of SNPs significantly associated with stem cellulose content variation.

Table 6.1 Homoeologous copies of wheat *Csl* genes with their corresponding orthologs from rice.Table 6.2 Status of splice variants of *Csl* genes in wheat genome.

LIST OF FIGURES

Fig 1.1 Schematic showing the components of plant cell wall. Adapted from Achyuthan et al. 2010.

Fig 3.1 Predicted protein features of wheat cellulose synthase genes. The numbers 1 to 8 in the purple rectangles refers to the transmembrane domains (TMDs). Black triangles localise the conserved motifs. The newly identified motifs CXXC and SXXCEXWF are highlighted in blue and previously reported motifs in black.

Fig 3.2 Structural features of the *TaCesA* genes. Drawn to scale, exons are represented by black boxes and introns by grey lines. Intron lengths are presented on top of each intron. PCW and SCW *CesA* genes are shown in blue and red colours, respectively.

Fig 3.3 Amino acid sequence alignment of wheat CESA proteins. Drawn to scale with solid lines representing conserved amino acid sequences and the gaps representing the mismatches and deletions. Corresponding phases of intron evolution (0, 1, and 2) for the CESA proteins are shown on the top of the solid lines. Primary and secondary cell wall CESAs are shown in blue and red colour, respectively.

Fig 3.4 Motifs differentiating PCW and SCW CESA orthologs from different species. Conserved and non-conserved amino acids residues are highlighted in red and green respectively. Amino acid changes in the motifs are shown in blue.

Fig 3.5 Conserved motifs differentiating the orthologs of SCW CESAs from *Triticum aestivum* (TaCESA), *Arabidopsis thaliana* (AtCESA), *Hordeum vulgare* (HvCESA), *Oryza sativa* (OsCESA), and *Zea mays* (ZmCESA). Conserved and non-conserved amino acids residues are highlighted in red and green respectively. Amino acid changes in the motifs are shown in blue.

Fig 3.6 Conserved motifs differentiating the orthologs of PCW CESAs from *Triticum aestivum* (TaCESA), *Arabidopsis thaliana* (AtCESA), *Hordeum vulgare* (HvCESA), *Oryza sativa* (OsCESA), and *Zea mays* (ZmCESA). Conserved and non-conserved amino acids residues are highlighted in red and green respectively. Amino acid changes in the motifs are shown in blue. Fig 3.7 Monocots and dicots specific motifs of CESA orthologs from *Triticum aestivum* (TaCESA), *Arabidopsis thaliana* (AtCESA), *Beta vulgaris* (BvCESA), *Eucalyptus grandis* (EgCESA), *Glycine max* (GmCESA), *Gossypium hirsutum* (GhCESA), *Hordeum vulgare* (HvCESA), *Oryza sativa* (OsCESA), *Populus trichocarpa* (PtCESA), *Solanum tuberosum* (StCESA) and *Zea mays* (ZmCESA). Conserved and non-conserved amino acids residues are highlighted in red and green respectively. Amino acid changes in the motifs are highlighted in blue.

Fig 3.8 Unrooted phylogenetic tree of the CESAs from *Triticum aestivum* (TaCESA), *Arabidopsis thaliana* (AtCESA), *Beta vulgaris* (BvCESA), *Eucalyptus grandis* (EgCESA), *Glycine max* (GmCESA), *Gossypium hirsutum* (GhCESA), *Hordeum vulgare* (HvCESA), *Oryza sativa* (OsCESA), *Populus trichocarpa* (PtCESA), *Solanum tuberosum* (StCESA) and *Zea mays* (ZmCESA). The bar provides a scale for the branch length in the horizontal dimension. The line segment with the number '0.1' means that an equal length of the branch between the CESA proteins represents a change of 0.1 AA. Color codes for different species: Red - TaCESA, blue – AtCESA, purple - HvCESA, yellow - ZmCESA, green - OsCESA, and grey – BvCESA, EgCESA, GmCESA, GhCESA, PtCESA, StCESA.

Fig 3.9 Heat map of 21 *CesA* transcripts by log2 counts per million (CPM) standard deviation in hexaploid wheat.

Fig 3.10 Map positions of *TaCesA* genes in the wheat genome. The exact locations are shown in

Fig 4.1 Schematic depicting the structure of *TaCesA4* gene and its homoeologs. Red bar indicates the 110 bp region of the *TaCesA4* gene cloned into the BSMV γ vector.

Fig 4.S1a Multiple sequence alignment of the fragment used for designing VIGS construct with other secondary cell wall related genes (*TaCesA4*, *TaCesA7* and *TaCesA8*) along with their homoeologs representing the non-conserved region.

Fig 4.S1b Multiple sequence alignment of the fragment of *TaCesA4* gene used for designing VIGS construct with its homoeologs representing the conserved region.

Fig 4.2 *In silico* expression of *TaCesA4* homoeologs in different wheat tissues, expressed as reads per kilo base of transcripts per million mapped reads (FPKM) in hexaploid wheat. Blue color bar represent *TaCesA4A*, black and green bars denotes *TaCesA4B* and *TaCesA4D* respectively.

Fig 4.3 Silencing of the *phytoene desaturase (PDS)* gene. Leaf phenotypes of wheat plants inoculated with BSMV:00 and BSMV: *TaPDS* at 21 dpi.

Fig 4.4 Semi-qPCR based expression of *TaCesA4* normalised to reference gene *TaActin* in silenced plants (BSMV: *TaCesA4*) and non-silenced plants (BSMV:00); Where L- marker, -ve- negative control. **Fig 4.5** Quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) analyses to confirm the gene knockdown as the relative transcript expression of *TaCesA4* normalized to *TaActin* mRNA in BSMV: *TaCesA4* inoculated plants as compared to control (BSMV:00) plants at 21 dpi.

Fig 4.6 Cellulose content (% w/w) in *TaCesA4* silenced (BSMV: *TaCesA4*) plants as compared to control (BSMV:00) plants.

Fig 4.7 Transverse sections of stem tissues of control (BSMV: 00) and silenced (BSMV: *TaCesA4*) wheat plants at 20X and 4X magnification; where mx is meta xylem, px is protoxylem, ph is phloem.

Fig 5.1 Density plot showing the percentage cellulose content among 288 diverse spring wheat accessions.

Fig 5.2 Principal component analysis of 288 diverse genotypes used for GWAS.

Fig 5.3 Minor allele frequency (MAF) patterns determined relative to allele calls for wheat genotypes based on 21073 SNPs.

Fig 5.4 Manhattan plot of genome-wide association study (GWAS) on stem cellulose content (mg cellulose/mg dry weight) by using the FarmCPU. The -log10(*p*-values) from GWAS are plotted against the position on each of the 42 bread wheat chromosomes. U represents unassigned chromosome scaffolds. Two loci on chromosomes 1A and 5A were identified above the Bonferroni threshold correcting genome-wide multiple tests at type I error of 0.001 (green line).

Fig 5.5 Quantile-quantile (QQ) plot showing the deviation from null hypothesis for associated SNP makers.

Fig 6.1 An unrooted phylogenetic tree representing the *Cellulose synthase-like* gene family from Arabidopsis, maize, rice and wheat using MEGA6. Tree was constructed using Neighbour joining (NJ) method with 100 bootstrap value. Different colors represent the subfamilies with orthologous CSL proteins from different species. The bar provides a scale for the branch length in the horizontal dimension. The line segment with the number '0.5' means that an equal length of the branch between the CSL proteins represents a change of 0.5 AA.

Fig 6.2 Distribution of *TaCsl* genes and their splice variants in seven subfamilies and their corresponding pfam domains used to identify *TaCsl* gene family.

Fig 6.3 An unrooted phylogenetic tree representing the *CslD* subfamily from Arabidopsis, Brachypodium, maize, rice and wheat using MEGA6. Tree was constructed using Neighbour joining (NJ) method with 100 bootstrap value. Different colors represent orthologous *Csl* genes from different species. Arabidopsis-blue, Brachypodium-purple, maize-sky blue, rice-green, wheat-red.

Fig 6.4 Structural features and phases of intron evolution of the *CslD* subfamily genes. Drawn to scale, exons are represented by red boxes and introns by back lines. Corresponding phases of intron evolution (0, 1, and 2) for the CslD genes are shown on the top of the black lines.
Fig 6.5 Heat map showing the expression profiling of wheat *Cellulose synthase-like (TaCsl)* genes at seedling, vegetative and reproductive stages. (A) *CslA* (B) *CslC* (C) *CslD* (D) *CslE* (E) *CslF* (F) *CslH* & *CslJ*. RNA-seq data 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.

Fig 6.6 Pie chart showing the percentage of *TaCsl* genes on wheat chromosomes.

LIST OF ABBREVIATIONS

AFLPs	Amplified fragment length polymorphism
AX	Arabinoxylan
BLAST	Basic local alignment search tool
BSMV	Barley stripe mosaic virus
cDNA	Complementary deoxyribonucleic acid
CEF	Cellulose elementary fibril
CesA	Cellulose synthase A
CIMMYT	International Maize and Wheat Improvement Center
CSC	Cellulose synthase complex
Csl	Cellulose synthase-like
C-SR	Class-specific region
CSS	Chromosome Survey Sequence
DFMs	Direct Functional Markers
DNA	Deoxyribonucleic acid
dpi	Days post inoculation
EST	Expressed Sequence Tag
FPKM	Fragments Per Kilo base of transcript per Million mapped reads
FarmCPU	Fixed and Random Model Circulating Probability Unification
FMs	Functional Markers
GAPIT	Genomic Association and Prediction Integrated Tool
GAX	Glucuronoarabinoxylan

GBS	Genotyping by sequencing
GH	Glycosyl Hydrolase
GS	Genomic selection
GT	Glycosyltransferase
GTMs	Gene Target Markers
GWAS	Genome-wide association study
HRS	Hard Red Spring
HWS	Hard White Spring
IFM	Indirect Functional Markers
IWGSC	International wheat genome sequencing consortium
LD	linkage disequilibrium
MAF	Minor allele frequency of
NGS	Next Generation Sequencing
PCR	Polymerase chain reaction
P-CR	Plant-conserved regions
PCW	Primary cell wall
PDS	Phytoene desaturase
PIECE	Plant Intron-Exon Comparison and Evolution database
PNW	Pacific North West
QTL	Quantitative trait loci
RAPDs	Random Amplified Polymorphic DNA
RDM	Random markers
RFLPs	Restriction fragment length polymorphism

RNA	Ribonucleic Acid
RNAi	RNA interference
SCW	Secondary cell wall
SNP	Single nucleotide polymorphism
SRS	Soft Red Spring
SSR	Simple Sequence Repeat
SWS	Soft white spring
TaCesA	Triticum aestivum Cellulose synthase A
TILLING	Targeting Induced Local Lesions IN Genomes
TMDs	Transmembrane domains
UDP	Uridine diphosphate
UGT	UDP-glucuronosyltransferase
VIGS	Virus-induced gene silencing
ZnF	Zinc-finger

ABSTRACT

Synthesis and remodelling of various cell wall components play a vital role in plant development, architecture and innate immunity. Plant cell walls are mainly composed of cellulose and hemicellulose which produce a bulk of renewable biomass vital for food, feed and biofuels. Cellulose in the primary and secondary cell wall of plants is synthesised by the family of genes called CesA (Cellulose synthase A). This study is a first report about the distinctive structural and functional motifs of primary and secondary cell wall synthesis genes. Using publicly available genomic databases and resources, 22 TaCesA genes located on A, B and D genomes of hexaploid wheat were identified. Cellulose in secondary cell walls is synthesised by three genes (*TaCesA4*, TaCesA7, and TaCesA8) co-expressing in the mature stem tissues of bread wheat. But the relative transcript abundance was found to be higher for TaCesA4 genes, which indicates its major role in the secondary cell wall cellulose synthesis. We employed the virus-induced gene silencing (VIGS) approach to functionally characterize TaCesA4 gene through silencing its three homoeologs (TaCesA4A, TaCesA4B, and TaCesA4D) collectively in bread wheat. Silenced plants showed a significant reduction in transcript abundance and cellulose content in the stem tissues. However, the anatomy of stem cross sections of silenced plants did not show any evidence of abrupt changes in the secondary cell wall of stems at the booting stage. A panel of 265 diverse wheat lines was evaluated for natural variation of cellulose content that was linked to the SNP genotyping data through genome-wide association studies (GWAS). This analysis led the identification of novel genes (β -tubulin and UDP-glycosyl transferase) associated with cellulose biosynthesis in wheat. In addition, Cellulose synthase-like (Csl) genes of wheat were explored. These genes have been known for the regulation/synthesis of hemicelluloses such as heteromannan, xyloglucan, heteroxylans, and mixed-linkage glucan. A total of 108 Csl genes were identified based on the

family specific Pfam conserved domains. Tissue-specific expression and phylogeny of *Csl* genes were also elucidated. Taken together, genome- wide exploration of *CesA* & *Csl* genes and their association with cellulose and hemicellulose biosynthesis offer a valuable resource for designing high yielding wheat varieties possessing appropriate lignocellulosic traits.

RÉSUMÉ

La synthèse et la remodelage des divers composants des parois cellulaires jouent un rôle important dans le développement, l'architecture et l'immunité innée des plantes. Les parois cellulaires sont principalement composées de cellulose et d'hémicellulose, lesquelles représentent une quantité importante de biomasse dans les aliments pour humains et bétail autant que dans les biocombustibles. La cellulose présente dans les parois cellulaires primaires et secondaires est synthétisée par des gènes de la famille CesA (Cellulose synthase A). Cette étude est la première à décrire les motifs structurels et fonctionnels caractéristiques de ces gènes de synthèse de parois cellulaires primaires et secondaires. Utilisant des ressources génétiques disponibles, 22 gènes TaCesA situés sur les génomes A, B et D du blé hexaploïde furent identifiés. La cellulose dans les parois cellulaires secondaires est synthétisée par trois gènes (TaCesA4, TaCesA7 et TaCesA8) qui sont coexprimés dans les tissus matures des tiges de blé. Cependant, les transcrits du gène TaCesA4 étaient plus abondants, ce qui indique l'importance élevée de ce gène pour la synthèse de la cellulose dans les parois cellulaires secondaires. Par biais d'une technique silençage de gène induit par virus (VIGS), nous avons caractérisé la fonctionnalité du gène TaCesA4 en désactivant tous ses trois homologues (TaCesA4A, TaCesA4B et TaCesA4D) dans le blé. Les plantes avec les gènes ainsi désactivés montrèrent une réduction significative en abondance des transcrits et en quantité de cellulose présente dans les tissus de leurs tiges. Cependant, l'anatomie des sections transversales des plantes aux gènes désactivés ne montrèrent aucune évidence de changements dramatiques dans les parois secondaires des cellules des tiges au phase de reproduction. Un ensemble de 265 diverses lignées de blé fut évalué pour caractériser la variation naturelle de la teneur en cellulose. Ces différences furent ensuite comparées avec des données de génotypage de polymorphismes mononucléotidiques par biais d'une étude d'association pangénomique. Cette analyse mena à l'identification de nouveaux gènes (β -tubulin et glycosyl transférase UDP) associés avec la biosynthèse de la cellulose dans le blé. Des gènes du blé similaires à ceux de la cellulose, *Cellulose* synthase-like (*Csl*), furent aussi explorés. Ceux-ci ont déjà été reconnus pour leur rôle dans la régulation et la synthèse des hémicelluloses tels que le l'hétéromannane, le xyloglucane, les hétéroxylanes, et les glucanes à liaisons mixtes. Un total de 108 gènes de *Csl* fut identifié grâce aux domaines Pfam conservés spécifiques à cette famille, et la phylogénie et l'expression au niveau des tissus de ceux-ci furent ensuit analysées. L'analyse en profondeur de l'architecture génétique de la biosynthèse de la cellulose et de l'hémicellulose offre un atout précieux pour l'amélioration végétale et les modifications génétiques des variétés de blé en but d'obtenir une production de biomasse désirable tout en conservant une résistance suffisante envers de divers stresses.

ACKNOWLEDGEMENT

This Dissertation is dedicated to my beloved parents

Firstly, I owe the debt to the "Almighty" Lord for showering ultimate blessings on me and helping me to become able to present this humble contribution to the knowledge of science. It is with His grace and blessings that I have been able to make another remarkable achievement in my life.

It is rightly said that "Every effort is motivated by ambitious and all ambitions have an inspiration behind." I owe this pride place to my parents, Mr Jagseer Singh and Mrs Veerpal Kaur. I am forever indebted to my parents for their understanding, endless patience and encouragement when it was most required and for providing me with the means to learn and understand. I cannot weigh my feelings with words for my dearest brothers Hardeep and Harpreet for their motivation, encouragement, everlasting love and affection and moral support.

Though the debt of learning cannot be repaid, it is my sovereign privilege to express my gratitude and moral obligation to my esteemed Supervisor, Dr Jaswinder Singh, Associate Professor, Department of Plant Science for his enlightened, invaluable and inspiring guidance. I shall remain ever indebted for his care and affection during the course of the investigation as well as in the preparation of this thesis. His multifaceted personality and commitment to work motivated and encouraged me to work, even harder and hence developed right attitude not only for my research work but also as a managed human being. I would also like to express my gratitude to the members of Singh lab: Harvinder Syan, Surinder Singh, Prabhjot Nanda, Daishu Yi, Chi-Kang Tsai, Haritika Majithia, and Rajeev Tripathi.

Words are insufficient to convey my deep sense of gratitude to members of my advisory committee, Dr Rajinder Dhindsa, Department of biology and Dr Don Smith Department of plant science for their timely and valuable help, judicious guidance, constant encouragement and constructive suggestions throughout the preparation of this manuscript.

I feel elated in expressing thanks to Dr Kanwarpal Dhugga and Dr Kulwinder Gill for their expert advice, constructive reviews and never ending cooperation from time to time in conducting the research work and for making improvements while going critically through the manuscript. I would like to gratefully acknowledge the support of Xu Zhang and Dr Zhiwu Zhang and Dr. Amita Mohan for their help and constructive discussions during the GWAS analysis, and Dr Raj Duggavathi for teaching me the use of microtome for histology.

Thanks are due to Dr Martina Stromvik, Department chair and past graduate Director of Plant Science and current graduate Director Dr Suha Jabaji for their constant support and recommendations for faculty scholarships. I also express my thanks to the supporting office staff plant science, Guy Rimmer, Ian Ritchie, Carolyn Bowes and Lynn Bachand for their continuous help during my research work. I am grateful to computer facility provided in our department by DST and Mr Serge Dernovici, technician, Institute of Parasitology, for teaching me to use confocal microscopy for histochemical studies.

I am highly obliged to McGill University for awarding the graduate research excellence awards, Graduate research enhancement and travel awards, Frederick Dimmock Memorial Fellowship in the department of plant science. I wish to mention my gratefulness to the BioFuelNet, Canada for funding my PhD project, providing me with Highly Qualified Personnel (HQP) international exchange Awards and HQP travel awards. Innovagrains Network, Quebec, Canada is fully acknowledged for awarding me the réseau Innovagrains scholarship for the year 2015. McGill University financial aid services are highly appreciated for their timely help during my PhD studies.

I have been fortunate to come across my funny & good friends without whom life would be bleak. I am happy to acknowledge the shadow support and moral upliftment showered upon me by Anant, Mamta, Maryanne, Julian, Aman, Jaskaran, Swadha, Balwinder, Chelsea, Aman, Gagan, Rizwana, Raghu and Raman. Last but not the least, I duly acknowledge my sincere thanks to all those who love and care for me. Every name may not be mentioned but none is forgotten.

PREFACE AND CONTRIBUTION OF THE AUTHORS

a. Preface

This thesis work is presented in a manuscript-based format. During the course of my PhD studies, I have conducted comprehensive analysis of the genetic variation in cellulose content in bread wheat and genes underlying this. I explored the genes involved in the synthesis of cellulose and hemicellulose of cell wall through genetic, genomics and bioinformatics approaches. The thesis contains four different studies revolving around cell wall- associated genes. In study 1, *cellulose synthase (CesA)* genes were identified and analysed for their structure, function and evolution in wheat, study II involves the functional characterization of secondary cell wall specific *CesA4* gene using virus-induced gene silencing (VIGS), study III was performed to estimate the cellulose content of wheat straw and its genetic control in diverse wheat varieties, and final study IV was the Genome-wide analysis of the *Cellulose synthase-like (Csl)* gene family in bread wheat. The results of these studies have been presented in Chapter III, IV, V and VI respectively.

The following features of this study are considered as distinctive contributions to knowledge:

- Identification of *CesA* genes in wheat and their structural, functional and evolutionary studies will lead to designing cultivars suitable for both food and fuel.
- Functional validation of SCW-forming *CesA4* gene extrapolates differential functional role in higher plants
- Discovery of novel genes and/or SNPs associated with cellulose content in wheat could be helpful in devising molecular- assisted selection strategies for enhancing the culm strength, lodging resistance and wheat stem sawfly tolerance.

• Genome-wide identification and expression studies of poorly understood *cellulose synthase-like* (*Csl*) genes underscore their role in various polysaccharide biosynthetic processes in plants

b. Contribution of the authors

This thesis involves four studies (Chapter III to VI) printed in the form of four manuscripts as per the thesis preparation guidelines provided by McGill. The research work presented here has been completely outlined by me under the guidance of my supervisor Dr Jaswinder Singh. I have performed all the experiments in the greenhouse and laboratory set up and conducted genetic, genomic, bioinformatics analyses. Under the supervision of Dr Jaswinder Singh, I have analysed the data, wrote manuscripts and the thesis. He helped in troubleshooting, provided constructive comments, suggestions and financial support to conduct the experiments. He has thoroughly edited all the manuscripts and incorporated his suggestions.

The first manuscript (Chapter III) is co-authored by Kanwarpal S. Dhugga, Kulvinder Gill, and Jaswinder Singh. Dr Dhugga thoroughly edited the manuscript and added his valuable thoughts, Dr Gill shared the ideas of representation of bioinformatics analysis. The second manuscript (Chapter IV) was co-authored by Kanwarpal S. Dhugga, Raj Duggavathi, Kulvinder Gill and Jaswinder Singh. Dr Gill provided the training for VIGS and Dr Dhugga and Dr Duggavathi provided their expert advises performing the experiments. The third manuscript (Chapter V) was co-authored by Xu Zhang, Amita Mohan, Prashant Vikram, Sukhwinder Singh, Kanwarpal S. Dhugga, Zhiwu Zhang, Kulvinder Gill and Jaswinder Singh. Prashant Vikram, and Sukhwinder Singh provided the genotyping data, Xu Zhang and Amita Mohan helped in the creating the SNP data and GWAS analysis. Dr Dhugga provided the protocols for cellulose content analysis, Dr Zhang and Dr Gill provided their expert advice and suggestions to interpret the results. The fourth manuscript (Chapter VI) was co-authored by Kanwarpal S. Dhugga and Jaswinder Singh. Dr Dhugga again provided his expert advice and edited the manuscript.

Chapter I: General introduction

The cell wall is the robust outermost layer of plant cells that covers the plasma membrane (Keegstra 2010). In the living cells, these walls not only encase the protoplasm but also act as complex and dynamic compartments with diverse and subtle functions (Fry 2004). They play a major role in plant growth, development, physical strength and innate immunity (Cosgrove 2000). Polysaccharide composition of cell walls makes them fundamentally different from cell membranes that are made up of proteins and phospholipids (Fry 2001). Cell walls usually laid down soon after the mitosis surrounding the dividing daughter cells. The thickness of the walls usually increases with the deposition of new microfibrils on the inner face of the developing cell wall (Cosgrove 2005).

Cell walls are classified into primary and secondary walls (Burton and Fincher 2014a). Primary cell walls are laid around the plasma membrane just after the cell division, allowing the cells to increase in size as they grow (Thomas et al. 2013). Whereas, secondary cell wall usually develops inner to the primary cell wall after the cell stops growing (Zhong and Ye 2014). Secondary cell walls provide greater mechanical strength to the cells and often surrounds the xylem vessels and lignin-rich woody tissues (Boerjan et al. 2003). The composition of the cell wall is fractionated into three polysaccharide classes: cellulose, hemicellulose and pectins (Achyuthan et al. 2010). In addition to these components, the cell wall matrix also contains some proteins, lignin, cutin and suberin infiltered between the microfibrils (Fry 2004).

Cell wall polymers are the end products of solar energy transformation by plants through photosynthesis. Total dry matter of plants including carbohydrate polymers (cellulose, hemicellulose and pectins) and aromatic polymers (lignin) is called lignocellulosic biomass (Guerriero et al. 2016). Beyond their fundamental significance associated with overall plant physiology, lignocellulosic cell walls represent the most abundant renewable carbon source for biofuels and biomaterial industries. Over 90% of the global plant biomass is lignocellulose which accounts for about 200×10^9 tonnes/year, of which $8-20 \times 10^9$ tonnes remains accessible every year (Saini et al. 2015).

Wheat, a major staple food of the world, which also produces a large amount of lignocellulosic straw (1–3 tonnes/acre annually), is currently an important target crop for the synthesis of bioproducts (Saini et al. 2015). However, synthesis of cell-wall components, genetic diversity and their association with polysaccharide composition are not well understood in wheat. Identification and characterization of these genes is a prerequisite in designing the crops for more desirable harvests.

Cellulose (a homopolymer of glucose) and hemicellulose (heteropolymer of pentoses and hexoses) are the major components of the lignocellulose and are synthesised by the genes of a large superfamily known as *Glycosyltransferase 2* (*GT2*) (Breton et al. 2006; Kaur et al. 2016). Within this superfamily, there are two distinct multigene families that encode the catalytic subunits for the synthesis of cellulose and hemicellulose. The group of genes that involve in the synthesis of cellulose at the plasma membrane are called *Cellulose synthase A* (*CesA*) (McFarlane et al. 2014). On the other hand, hemicelluloses are synthesised by *Cellulose synthase-like* (*Csl*) genes located in the Golgi membranes (Pauly et al. 2013).

In addition to the *CesA* genes, the genes of *Glycosyl Hydrolase 9* (*GH9*) family and sucrose synthase (Fujii et al. 2010) have been reported to be involved in the synthesis of cellulose and cell expansion (Szyjanowicz et al. 2004; Lei et al. 2014; Vain et al. 2014). This explains the complexity of the cellulose synthesis process in the plant. Many of cell wall-related genes have been reported in case of model species Arabidopsis (Turner and Somerville 1997; Arioli et al. 1998; Taylor et al.

1999; Desprez et al. 2007b) and other cereals such as rice (Wang et al. 2010a), maize (Appenzeller et al. 2004), brachypodium (Coomey and Hazen 2015) and barley (Schwerdt et al. 2015)

However, bread wheat is lagging behind in the understanding of cell wall genetic architecture due to its complex and large genome size (17 Gb) (Krasileva et al. 2017). In addition to that, the first version of the chromosome-based draft genome sequence of bread wheat (*Triticum aestivum*) has been made available to the public recently (https://www.wheatgenome.org). Reference genotype Chinese Spring has been used to sequence the whole genome by international Wheat Genome Sequencing Consortium (IWGSC) (Consortium 2014).

Recent progress in sequencing efforts and availability of extensive genomic resources have permitted the identification and isolation of candidate genes of interest. But the functional validation of these genes is a major challenge for the researchers. There are many ways to characterise genes in model crops and crops with small genome size, such as chemicals, T-DNA, stable transformation through RNAi (Chen et al. 2014). Although some of these approaches are also available for wheat but are very laborious, time-consuming and expensive. A rapid and less expensive tool has recently developed in wheat called virus-induced gene silencing (VIGS) (Stratmann and Hind 2011; Bennypaul et al. 2012a; Baenziger et al. 2014).

Therefore, bioinformatics approaches coupled with functional genomics tools such as VIGS can enable the rapid exploration of structure and function of cell wall related genes in wheat. Being a crucial polysaccharide for plants and humans, exploring whole genome targets is vital to uncover the complex mechanism of cellulose synthesis in plants. Genome-wide association studies (GWAS) has emerged as an effective way to find the novel gene-trait associations. Moreover, the screening of diverse wheat genotypes for cellulose will provide the basis of genetic manipulation of lignocellulose and stalk strength.



Fig 1.1 Schematic showing the components of plant cell wall. Adapted from Achyuthan et al. 2010.

1.1 General hypothesis

We hypothesised, cellulose content in bread wheat cultivars varies greatly which is associated with cellulose synthase, cellulose synthase like and other related genes

1.2 General objectives

I: Identification of *Cellulose synthase* (*CesA*) genes to understand their structure, function and evolution in wheat

II: Functional characterization of secondary cell wall specific *CesA* gene using virus-induced gene silencing (VIGS)

III: Estimation of the cellulose content of wheat straw and its genetic control in diverse wheat varieties

IV: Genome-wide analysis of the Cellulose synthase-like (Csl) gene family in bread wheat

Chapter II. Literature review

2.1 Future energy requirements

With the increase in global population, depleting energy sources are among the biggest concerns for humanity (Scholey et al. 2016). Use of fossil fuels as an energy source over the years is a major factor in global warming and increase in greenhouse gas emissions (Strezov and Evans 2014). Additionally, fossil fuels are a finite resource and the process of fossil fuel formation is very slow, therefore one cannot survive by solely relying on this fuel (Moriarty and Honnery 2016). There is a need for renewable energy resources to overcome the danger of depleting non-renewable energy resources and to conserve the environment. Concerns of fuel depletion and environmental safety have attracted governments and scientists to search for alternatives to fossil fuels to secure future energy requirements (Perera 2016). Lignocellulose, the most abundant renewable biomass on the earth, has tremendous potential for conversion into biofuels (Broom et al. 2013). It is estimated that 10¹¹ tonnes of cellulose are synthesised each year through the process of photosynthesis (Carroll et al. 2012). Research in the field of biofuels from lignocellulose feedstock is growing to meet the future energy requirements and check greenhouse gas emissions.

2.2 Lignocellulosic materials as bioethanol

Agricultural residues such as stems, stalks, and straws are the most abundant sources of renewable lignocellulosic biomass that can be efficiently converted to bioethanol (Hood 2016). Lignocellulosic biomass is less competitive, cheaper and has no influence on the growing demand for human food (Gabhane et al. 2014). A large part of agricultural lignocellulosic biomass comes from world's major crops such as maize, wheat, rice, and sugarcane (Chandra et al. 2012) Plant biomass is mainly composed of cell walls and quality of biomass is determined by the type of cell

walls (Sorek et al. 2014). There are two types of cell walls: primary cell walls and secondary cell walls. Deposition of primary cell wall takes place during the cell division and expansion stage whereas the secondary cell wall is deposited on the cell after expansion ceases until the cell dies (Carroll et al. 2012). All of the cell wall components cellulose, hemicellulose, pectin, lignin, and minerals, are collectively known as lignocellulose (Guerriero et al. 2016). The composition of lignocellulose varies depending on the species, cell type, environmental conditions and developmental stages of the plant (Sorek et al. 2014).

2.3 Structure and composition of lignocellulose

Lignocellulosic material is a complex network of three major cell wall components: cellulose, hemicelluloses and lignin along with other minor components. In general lignocellulose of wheat straw is comprised of cellulose (~30-40%), hemicelluloses (~20-35%) and lignin (~15-25%) (Ruiz et al. 2013). The composition of lignocellulose plays a crucial role in determining its biodegradation to bioethanol. To improve the efficiency of biofuel production, there is a need to explore the composition of lignocellulose and its genetic regulation.

2.3.1 Cellulose

Cellulose, the major component of plant cell walls, consists of a linear chain of β (1 \rightarrow 4) linked glucan (poly glucose) units. Cellulose elementary fibril (CEF) is the fibril synthesised by the cellulose synthase complex (CSC) and the bundle of these CEF are called macrofibrils. It is probable that CSC containing more than 24 isoforms of cellulose synthase can synthesise about 36-chain CEFs. Microfibrils are morphological units that can be either CEF associated with hemicelluloses or a small macrofibril. The cellulose microfibril is 2 to 50 nm in size. Physiochemical properties of cellulose varies according to the degree of polymerization, a number

of chains and the orientation of the chains which are packed together (Ding et al. 2013). Cellulose is insoluble in water and most organic solvents because of intra and intermolecular hydrogen bonding which results from its free alcoholic groups (Shaveta et al. 2014). The size of cellulose microfibrils can vary in different tissues depending upon the degree of polymerization, from 500 to 15,000 glucose molecules. Cellulose microfibrils are generally bonded to hemicellulose through hydrogen bonds (Sorek et al. 2014).

2.3.2 Hemicellulose

In addition to cellulose, there is another cell wall component made up of several heteropolymers that are called hemicellulose. Hemicellulose is made up of diverse linear and branched polysaccharides and their composition varies widely depending on tissue and species. They mainly contain a β -(1, 4)-linked glucan, xylan, galactan, mannan, or glucomannan backbone branched with glycosyl residues. In addition to these components, mixed linked (1-4), (1-3) β -glucans are also abundant in some grass species (Sorek et al. 2014). Due to the presence of heterogenous substituents or other linkages in their polymer backbone, the structure of hemicellulose is amorphous and can be easily hydrolysed as compared to cellulose. These polysaccharides can interact with cellulose chains through hydrogen bonds (Pauly et al. 2013). Hemicellulose acts as a cross-linking agent between cellulose bundles, lignin and proteins through covalent or non-covalent bonding (Sorek et al. 2014).

Xylans are the most important hemicellulose and second most abundant polymer in the plant kingdom. Glucuronoarabinoxylan (GAX) is the major hemicellulose of monocot plants' secondary cell wall (Ong et al. 2014). Agricultural crops such as sorghum, sugar cane, corn stalks are all potential sources of xylans. Xylan occurs up to 70% of the weight in some tissues of grasses and

cereals (Ebringerová et al. 2005). Another important hemicellulose class Arabinoxylans (AXs), are most commonly found in the cell walls of cereal grains (Girio et al. 2010).

2.3.3 Lignin

Lignin is one of the major components of cell walls and is responsible for making them rigid, impermeable, and resistant to microbial attack and oxidative stress. Lignin makes biomass insoluble, therefore higher the lignin content lowers the digestibility of a given biomass (Eudes et al. 2014). It is the second most abundant polymer in nature after cellulose and is comprised of amorphous, heteropolymer of phenylpropane units. Lignin in most of the angiosperm species is composed of the phenylpropanoids p-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) in different proportions (Penning et al. 2014). Lignin forms a covalent bond with hemicelluloses and occupies the spaces in the cell wall between cellulose and hemicellulose (Sorek et al. 2014).

2.4 Biofuels and plant cell walls

Plant cell walls possess polysaccharides that are a huge source of possibly fermentable sugars. Cell wall polysaccharides are catching industry attention for their use in the production of various bioproducts (Burton and Fincher 2014b). Natural variability of different components of cell wall provides an opportunity to select biomass for specific applications (Ciesielski et al. 2014). Among the major cell wall biopolymers, cellulose is the key fermentable sugar, but the productivity of biofuels is highly influenced by hydrolysis of cellulose and hemicellulose. Breakdown of these polysaccharides into fermentable sugars (Saccharification) is the major step that determines the efficiency of biofuel production. Lignocellulosic biomass in its innate shape is recalcitrant to enzymatic degradation because of complex cellulose-hemicellulose network and lignin crosslinking (Douche et al. 2013). Distortion of this interaction between lignin, cellulose and hemicellulose needs a pre-treatment step that increases the cost of converting the feedstock to biofuel. The efficiency of pre-treatment largely depends upon the presence of covalent linkages among cell wall components, the strength of hydrogen bonding between cellulose and hemicellulose, thickness of the cell wall, and accessibility of cellulose for the breakdown. Therefore optimisation of biosynthesis of cell wall components is imperative to increase the efficiency of enzymatic hydrolysis for lignocellulosic feedstock (Ong et al. 2014). Several mutant studies have been performed to identify genes involved in cellulose and hemicelluloses biosynthesis, however, our current knowledge of mechanisms involved in cell wall polysaccharides biosynthesis is still rudimentary (Burton and Fincher 2012). Understanding cell wall characteristics and its natural variability will allow the creation of biomass specifically designed for efficient biofuel production.

2.5 Functional significance and synthesis of key cell wall components

Cell walls are the most abundant renewable source on earth and play a major role in providing physical strength and innate immunity to plants (Sarkar et al. 2009; Endler and Persson 2011). Plant cell walls consist mainly of cellulose, along with different proportions of hemicellulose and lignin. Among all the components of the cell wall, cellulose is the major target of the biofuel industry and most plentiful carbohydrate and a biopolymer in nature.

2.5.1 Genetics of cellulose synthesis

The cellulose in primary and secondary cell walls of plants is synthesised by a multigene family called *cellulose synthase* (*CesA*). In higher plants, the *CesA* gene family is primarily responsible for the synthesis of cellulose. In herbaceous plants, the secondary cell wall is deposited inside the primary cell wall. Genes involved in secondary cell wall thickening are important candidates to

study the genetic variability between diverse genotypes and are valuable in breeding programmes (Tian et al. 2014). Plant cellulose synthases (CesAs) belong to a large enzyme family called glycosyltransferase 2 (GT2), which is responsible for the creation of β -linkages between the glucose molecules in cellulose (Richmond 2000). Cellulose synthesis takes place at the Golgi membrane through the action of different isoforms of CesA that are specific to primary and secondary cell wall celluloses. CESAs are intrinsic membrane proteins whose catalytic domains extend into the cytoplasm (Rayon et al. 2014). They are found as rosettes, or cellulose synthase complexes (CSC), which are composed of a hexagonal structure of six protein subunits. A recent study predicted that each of the six subunits of cellulose synthase complex is composed of 4-6 enzymatically active CESAs that lead to the formation of an elementary microfibril. This microfibril is made of 24/36 glucan chains which are arranged in a rectangular form, with eight sheets of three chains to each sheet (Burton and Fincher 2014b). Multiple CESA proteins catalyse the synthesis of cellulose microfibrils through the polymerization of glucan chains and are involved in the crystallisation process (Li et al. 2014b). UDP-glucose acts as a substrate for a single-step CESA-catalysed reaction that polymerises the glucose residues (Liu et al. 2012). The structure and composition of cellulose microfibrils in primary and secondary cell walls determine cell wall elasticity and plant growth.

Higher plant CESAs are predicted to be comprised of eight transmembrane domains that form a pore in the plasma membrane to extrude newly synthesised cellulose. A zinc finger domain on the cytoplasmic amino terminal of CESAs is thought to be involved in the protein-protein interaction and dimerization of CESA proteins (Kaur et al. 2016). Motifs are the functional units of proteins and their discovery is important for the analysis of functional variability in different genes. The highly conserved motif 'CXXC' is present within this domain and distinguishes *CesA* genes from *cellulose synthase-like* (*Csl*) genes (Richmond 2000). The zinc finger domain is followed by an acidic amino acid-rich region called the hypervariable region (Fig 4). A central domain between the second and third transmembrane domains possess most of the conserved residues of glycosyltransferases. Three conserved aspartic acid residues (D1, D2, and D3) and a QXXRW motif present in the central domain act as signature residues in all species and are probably involved in substrate binding, acceptor binding, and catalysis (Li et al. 2014).

2.5.2 Cellulose Synthase-Like (Csl) genes and their importance

Hemicellulose in plants is synthesised by a superfamily of genes called *cellulose synthase-like* (Csl) genes. These genes encode the catalytic subunit of enzymes required for hemicellulose synthesis. These genes possess the "D, D, D, QXXRW" motif that is characteristic of glycosyltransferases. Csl genes share sequence similarity with CesA genes; 30 to 50 Csl genes can be found in different plant species (Hazen et al. 2002) There are 30 known Csl genes in Arabidopsis and about 37 in rice (Hazen et al. 2002; Somerville et al. 2004). Csl genes are classified into nine subfamilies (CslA-CslH and CslJ). Among which, the CslD subfamily, is conserved in all land plants. In addition, it shares the highest sequence similarity with CesA genes. This reveals their fundamental role in plant development. Two groups of Csl genes, CslF and CslH, have evolved independently in grasses and are responsible for the biosynthesis of $(1-3), (1-4)-\beta$ -D-glucan (Burton et al. 2011b). A third group CslJ had been recently identified as grass specific (Farrokhi et al. 2006). Arabidopsis which does not make (1-3), (1-4)- β -D-glucan shows small amounts of β -Dglucan when the gene from rice was heterologously expressed (Burton et al. 2006a). Comparative genomic analysis has revealed seven CslF family members in barley i.e., HvCslF3, HvCslF4, *HvCslF6*, *HvCslF7*, *HvCslF8*, *HvCslF9* and *HvCslF10* (Burton et al. 2008).
In barley, the *CslF* gene family is located in the genomic region corresponding to a major QTL involved in the synthesis of mixed linked glycans (Burton et al. 2006a). Transcription profiles of *CslF3* to *CslF10* have been detected in barley using 16 different tissues (Burton et al. 2008). Results showed the variable expression pattern of different *CslF* genes in different types of tissues. Relatively higher expression of *CslF3* and *CslF7* was detected in stem and peduncle tissues of barley; prompting the additional analysis of the involvement of specific *CslF* genes in the synthesis of (1-3), (1-4) β -glucan (Burton et al. 2011b). To date, the functional role of *CslF4*, *CslF6* and *CslH* in the synthesis of (1-3), (1-4) β -glucan has been demonstrated (Schreiber et al. 2014b). Taketa et al. (2012) reported the role of *CslF6* genes in β -glucan biosynthesis using β -glucanless mutants (Taketa et al. 2012) and the role of these genes have additionally, functionally characterised in wheat grain (Nemeth et al. 2010).

The subfamilies *CslA*, *CslC*, and *CslD* are found in all land plants, while the *CslB* and *CslG* subfamilies are present in dicots (Dhugga 2012). Four members of *CslA* group are involved in the synthesis of mannan and/or glucomannan. Expression profiling of seed development in guar (*Cyamopsis tetragonolobus*) shows that the *CslA* gene is responsible for mannan synthesis (Dhugga et al. 2004b; Liepman et al. 2005). Reverse genetic approaches in *Arabidopsis* have revealed that the *CslA* family is responsible for glucomannan biosynthesis (Goubet et al. 2009). Recent studies indicated the role *CslD* family in mannan synthesis (Verhertbruggen et al. 2011; Yin et al. 2011). Heterologous expression studies in the case of *Pichia* revealed that the *CslC* genes are involved in the synthesis of 1-4-β-glucan backbone of xyloglucan and some other polysaccharides (Cocuron et al. 2007). Despite much progress in the identification and functional analysis of *CesA/Csl* gene families in plants, there are no *(CesA)* and very few (*Csl*) case studies in wheat on these gene families.

2.6 Wheat straw and its potential as biofuel

Wheat (Triticum aestivum) is an important food crop all over the world. It is cultivated in around 115 nations, with an annual grain harvest of nearly 700 million tonnes (Zhang et al. 2014a). Global production of wheat straw is approximately 355 million tonnes every year which has potential to yield about 104 gigaliters of bioethanol (Saini et al. 2015). Wheat straw is composed of leaf and stem residues that remain after the harvesting of grain. It is comprised of 50-60% internodes, 15-30% leaves, and 10% nodes (Motte et al. 2014). Most of this straw is discarded as waste or burnt in the fields in developing countries. This creates big environmental and economic issue that could be otherwise used as a powerful resource of energy or source of biofuels (Shaveta et al. 2014). Canada is a major wheat producer; being ranked 6th on a global scale (Zhang et al. 2012). In the Canadian prairies, wheat, barley, oat, and flax grain production resulted in 37 million tonnes (Mt) of straw annually. Wheat alone contributes 25 Mt of straw. All this straw is not always available for industrial purposes as 0.75 t/ha to 1.5 t/ha of straw is required for soil conservation, depending on soil type. Also, 13-15 Mt of straw is required for livestock. However, 15 Mt of straw remains available for industrial purposes, that varies largely between 27-2.3 Mt (Sokhansanj et al. 2006; Tumuluru et al. 2014). Biofuel from wheat straw has been considered the most effective way to reduce the greenhouse effect and to generate energy from abundant biomass (Qureshi et al. 2013). Currently, the complexity in the structure of wheat lignocellulose makes the process of ethanol production less efficient. Current varieties of wheat have not been designed for cellulosic biofuel production. However, great potential exists at genetic and genomics level to alter lignocellulose composition of wheat and other grasses (Ong et al. 2014). Our current knowledge is limited in respect to the genetic and phenotypic variation of lignocellulosic biomass. Inclusive understanding of cell wall components is necessary for the complex process of converting lignocellulose into biofuel.

2.7 Importance of wheat and its genetics

Wheat originated about 10,000 years ago in the Near Eastern Fertile Crescent (Faris 2014). Wheat provides 20% of total calories in the average human diet and feeds 40% of the world population (Gupta et al. 2008). Wheat, an allohexaploid, has the genome size of ~17 Gb of which ~80-90% are repetitive sequences. In hexaploid wheat, three homeologous sets of seven chromosomes are distributed in three A, B and D subgenomes. These subgenomes were originally derived from three diploid species, Triticum urartu (AA), an unknown close relative of Aegilops speltoides (BB), and Aegilops tauschii (DD). Tetraploid wheat Triticum turgidum L. (2n=4x=28; AABB) originated about 0.5 million years ago (MYA) through the first hybridization event between the ancestral species: Triticum urartu (2n=2x=14) and Aegilops speltoides (2n=2x=14). About 8,000 years ago, a second hybridization event between Tetraploid wheat and a wild relative: Aegilops tauschii, which contributed the DD subgenome, resulted in *Triticum aestivum* with the (2n = 6x = 42)AABBDD genome (Choulet et al. 2014b). It still behaves as a diploid because of the action of homologous pairing through the action of Ph genes. However, subgenome B possess a higher number of gene loci as compared to the A and D subgenomes. Gene sequences on subgenomes A, B and D of hexaploid wheat have more than 99% identity with their respective diploid progenitors (Mayer et al. 2014). It has been reported that present day genome of hexaploid wheat is resulted from multiple rounds of hybrid speciation (homoploid and polyploid) (Marcussen et al. 2014).

2.8 Molecular markers in wheat

2.8.1 Random Markers (RDMs)

Random markers (RDMs) are derived arbitrarily from polymorphic sites in genomic DNA and cDNA (Gupta and Rustgi 2004) and are developed using restriction enzyme based methods. The most commonly used random DNA markers are RFLPs (Restriction fragment length polymorphism), SSRs (Simple sequence repeat) and AFLPs (Amplified fragment length polymorphism). Sequence information is required for SSRs, SNPs but not for RFLPs, RAPDs (Random Amplified Polymorphic DNA), AFLPs etc. SSR and SNPs (single-nucleotide polymorphism) are markers of choice for molecular breeding (Salgotra et al. 2014) and play an important role in crop improvement. For example, they are used for gene introgression through marker assisted backcrossing/marker-assisted recurrent selection, germplasm characterization, diversity analysis, identifying polymorphisms, construction of molecular maps, QTL analysis, gene tagging, map-based cloning, and phylogenetic analysis (Varshney et al. 2007).

2.8.2 Gene Target Markers (GTMs)

Due to the availability of high-throughput sequencing platforms and genomic information, there was a shift in trends from RDM to GTMs and functional Markers (FMs); located in or near the gene of interest (Poczai et al. 2013). GTMs are developed from polymorphic sites within genes, that may or may not be involved in phenotypic trait variations (Varshney et al. 2007). These markers can also tag untranslated regions of genes (Poczai et al. 2013). These markers are developed through sequencing, expression profiling, sequence comparisons, or synteny studies (Andersen and Lubberstedt 2003).

2.8.3 Functional Markers (FMs)

Genome-wide sequencing provides a platform to mine molecular markers (Muthamilarasan et al. 2013). Functional markers are the polymorphic sites within genes that are functionally validated for phenotypic variations (Salgotra et al. 2014). These makers are further classified into two groups: indirect functional markers (IFMs) and direct functional markers (DFMs); depending upon whether the proof for their role in phenotypic trait variation is indirect or direct. Functional markers can be derived from non-redundant EST databases either by direct mapping or database mining for markers such as EST-SNP (Mochida and Shinozaki 2010). GTMs and FMs allow the detection of nucleotide diversity in the genes controlling agronomic traits. These markers are useful in predicting the genetic relationship, as well as the functional diversity of the genes in relation to adaptive variation. In contrast to RDMs, these markers are transferable to related species or genera (Varshney et al. 2007).

2.9 Comparative genomics

Comparative genomics is based on collinearity and synteny of genes or chromosomes in diverse species descended from a common ancestor (Poursarebani et al. 2013). Grass species such as rice, oats, barley, and wheat and *Brachypodium*, are derived from common ancestors, therefore, gene order in these species is highly conserved. Among all these grass species, wheat has most recently split from barley (Bolot et al. 2009). Comparative analysis shows that wheat chromosome groups 2, 3, 4, 5, 6, and 7 are syntenic to barley chromosomes 2H, 3H, 4H, 5H, 6H, and 7H respectively (Cho et al. 2006). Comparative genomics studies provide us with information about orthologous gene functions from different species that are expected to produce similar phenotypes. With the progress of sequencing facilities and the availability of whole genome sequences for major cereals

such as rice, maize and barley, it is now possible to identify genes and predict their functions in cereal crops with more limited sequencing information. Comparative genomics predicts gene function by exploring genomics and post-genomics associations for the genes, either within or between plant species and prokaryotes. Biochemical functions can also be determined using 3D structures (Bradbury et al. 2013). The availability of large-scale genomic information and conserved synteny between various grass species provides an opportunity to explore gene function and structure (Mochida and Shinozaki 2013; Molnár et al. 2016; Devos et al. 2017).

Sequence comparison using online such ensemblplants resources as (http://plants.ensembl.org/index.html) (Bolser et al. 2015), gramene (http://www.gramene.org/ and (https://phytozome.jgi.doe.gov/pz/portal.html) (Goodstein et al. 2014) are important comparative, functional genomics analysis tool for crop plants (Monaco et al. 2014). Comparative analysis was performed taking Arabidopsis as a model to identify the Sm family of RNA-binding proteins in rice and maize (Chen and Cao 2014). A Phytoene synthase (Psy) gene was identified and cloned in wheat using its ortholog from maize, using *in silico* cloning (He et al. 2008). And an Ortholog (TaGW2) of a gene involved in grain development in rice has been similarly identified in wheat (Su et al. 2011). A comparative genomic analysis resulted in the introgression of Yr5 resistance, a major resistance component against yellow rust (McGrann et al. 2014).

2.10 Functional genomics in wheat

Functional genomics is a wide approach for predicting functions and interactions of genes and their products. With the advancement of genome sequencing platforms, large numbers of plant genomes have been fully sequenced. A large-scale genomic information needs to be characterised by assigning functions to individual genes and exploring the role of non-coding sequences. Integration and analysis of the genomic data are currently the biggest challenges (Mittler and Shulaev 2013).

Several reverse genetics tools such as transposons mutagenesis, T-DNA insertion, RNA interference (RNAi), and Targeting Induced Local Lesions IN Genomes (TILLING) enable researchers to study specific genes and their functions (Chen et al. 2014). The introduction of the maize *Ac-Ds* transposable element system as a transposon tagging tool into heterologous species offers unprecedented opportunities to link genes with function by creating and characterising mutant alleles (Singh et al. 2012). Similarly, virus-induced gene silencing (VIGS) has been considered as a rapid and cost-effective functional analysis tool for complex crop species such as wheat to suppress the expression of homeologous genomes (Stratmann and Hind 2011; Baenziger et al. 2014).

2.10.1 Gene silencing approach through RNA interference

RNA interference (RNAi)-induced gene silencing is the post-transcriptional degradation of m-RNA. It is a robust functional genomics tool to suppress the expression of three homologous genes in wheat. It can be efficiently utilised for silencing multigene families and homoeologous genes in polyploids with functional redundancy. RNA interference (RNAi) induced phenotype is stably inherited, that makes it very important tool for functional analysis of genes in wheat (Baenziger et al. 2014). A gene controlling grain traits was functionally characterised through RNAi (Hong et al. 2014). A recent study led to the downregulation of gliadins wheat lines through RNAi that can be useful for production of gluten free products for the celiac community (Gil-Humanes et al. 2014).

2.10.2 Virus-induced gene silencing in wheat

VIGS has emerged as a powerful tool for plant functional genomics. VIGS involve the silencing of target gene/genes as a part of plant defence mechanism against viral attack. This is a fast and cost-effective alternative to the polyploid crops where stable transformation through RNAi is difficult to perform (Senthil-Kumar and Mysore 2011). Infection of plants by a virus engineered with fragments of the gene of interest activates the post-transcriptional gene silencing as an innate defence response. VIGS can be performed with or without the availability of sequence information as reverse and forward genetic tool for functional analysis of genes (Ramegowda et al. 2014). This technique is based on the spread of the virus in a plant upon inoculation. Multiplication of recombinant cloned virus with incorporated plant gene sequence led to the complete or partial loss of gene function through post-transcriptional gene silencing. Suppression of gene expression and phenotypic changes can be observed in VIGS treated plants (Lee et al. 2012). VIGS provides an opportunity to clone genes in genetically complex organisms such as wheat, using the candidate gene approach. Barley stripe mosaic virus (BSMV)-based VIGS system can be used to silence three homoeologous copies of each gene in wheat (Bennypaul et al. 2012b; Buhrow et al. 2016; Zhang et al. 2016).

2.11 Genomics-integrated breeding

The modern era of -omics (functional genomics, comparative genomics) and high throughput marker technology provides an opportunity to understand the functions of genes with small effects that underlie most of the important traits (Madramootoo 2015). Genome-wide markers have potential to capture all additive effects for selection of desirable genotypes. Emerging genomic-integrated breeding technologies are revolutionising the understanding of mechanisms of complex quantitative traits in time/cost efficient manner.

2.11.1 Genome-wide association (GWA)

Association mapping is an advanced tool to detect the genes/QTLs based on phenotypic and genotypic associations. It is an important strategy to identify genes underlying variations in quantitatively inherited traits. It is based on the principle of linkage disequilibrium (LD). In simple terms, linkage/LD is the deviation from Mendel's 2nd law, which explains the independent assortment of two different loci. The phenomenon of association between two loci is called linkage when the common ancestor is within the recorded pedigree. Whereas, when the common ancestor is the recorded pedigree, it is known as LD (Laird and Lange 2011). Whole genome scanning for LD between mapped marker loci and traits of interest is called Genome-wide Association (GWA). Genome-wide markers have the potential to capture additive effects and thereby aid in the selection of desirable genotypes (Neumann et al. 2010).

There are a number of factors influencing association mapping studies, such as genetic marker coverage, a number of individuals studied, and linkage disequilibrium (Cockram et al. 2010). Marker density on a genomic map should be higher than the extent of LD, which in turn depends upon the population structure, genetic diversity and number of recombination events that have occurred and have restructured that diversity (Brachi et al. 2011).

A recent study integrated the approach of sequence-based GWAS and functional genome annotation displayed the potential of matching complex traits to their causal polymorphisms in rice (Huang *et al.*, 2012). Modern maize breeding techniques have shown a remarkable increase in its productivity in the last few decades. As maize is such a diverse crop, genome-wide genetic variation pattern among various maize lines has been studied extensively. In a GWAS maize study, two candidate genes were identified that were associated with yield-related traits measured under water-stress conditions (Hu and Xiong, 2013). Nested association mapping (NAM) population of 25 RIL families was generated for quantitative trait analysis in maize (McMullen et al. 2009). Genome-wide association (GWA) study of the maize nested association mapping (NAM) panel was performed to determine the genetic basis of quantitative leaf architecture traits and identification of some of the important genes (Tian et al. 2011). Genome-wide association studies (GWAS) found a strong association between genetic loci and 14 agronomic traits in the population of *Oryza sativa* subspecies indica (Huang et al. 2010). Genetic architecture of (aluminium) Al tolerance and Al tolerance loci in rice was identified through bi-parental QTL mapping and GWAS (Famoso et al. 2011). A recent GWA study showed the involvement of *Glycosyltransferases* (*GT*) and *Glycoside hydrolases* (*GH*) along with *Cellulose synthase A* (*CesA*) in the culm cellulose content of barley (Houston et al. 2015)

2.11.2 Genomic selection (GS)

The genomic selection was first introduced by (Meuwissen et al. 2001) as a recent advancement in molecular breeding technology for the study of quantitative traits. Quantitative traits are controlled by a large number of genes, with a cumulative effect of each gene on the trait. This approach uses whole genome molecular markers (high-density markers and high throughput genotyping) to develop a prediction model for estimating a breeding value for each individual (Crossa et al. 2011). The availability of full genome sequences through NGS (Next Generation Sequencing) has provided high throughput molecular markers (Jonas and de Koning 2013).

GS based on LD can be applied to the populations having extensive phenotypic data over the years to dissect complex traits. This process also avoids the generation of special mapping populations (Xu et al. 2013). In contrast to few major genes/QTLs, thousands of molecular markers possessing strong LD with the trait of interest are used for GS. A number of simulation studies in various crops such as wheat, maize, oil palm (Bernardo and Yu 2007; Wong and Bernardo 2008; Bassi et al. 2016), and forages (Simeao Resende et al. 2014) illustrated higher genetic gain through GS as compared to MAS or phenotypic selection. GS predicts the breeding values based on phenotyping and genotyping of only a small training population and selection is based on the genotyping of breeding population at early stages without phenotyping (Battenfield et al. 2016; Michel et al. 2016).

CONNECTING STATEMENT FOR CHAPTER III

Chapter III entitled Novel structural and functional motifs in *Cellulose synthase A (CesA)* genes of bread wheat (*Triticum aestivum*, L.) authored by Simerjeet Kaur, Kanwarpal S. Dhugga, Kulvinder Gill, and Jaswinder Singh was published in "*PLOS ONE*" *.

Based on the literature review in chapter II, Cellulose is the key fermentable sugar found as the major proportion of plant cell walls. Cellulose in the primary and secondary cell wall of plants is synthesised by the family of genes called *CesA* (*Cellulose synthase A*) (Haigler et al. 2016). The structure, function, and evolution of *CesAs* are poorly understood in wheat. This study is a first report about the distinctive structural and functional motifs of primary and secondary cell wall synthesis genes in wheat. Using available genomic resources, this study in chapter III describes the identification of 22 *TaCesA* genes located on A, B and D genomes of hexaploid wheat. A thorough analysis was performed to investigate their structure, motif & domain architecture, evolution and expression patterns. Newly identified motifs were found to act as signature residues for the specificity of different *CesA* genes. A detailed information about these genes is discussed in chapter III.

*Kaur S, Dhugga KS, Gill K, Singh J. (2016) Novel Structural and Functional Motifs in *Cellulose synthase A (CesA)* Genes of Bread Wheat (*Triticum aestivum*, L.). *PLOS ONE* 11 (1): e0147046. doi:10.1371/journal.pone.014704.

Chapter III. Novel structural and functional motifs in *Cellulose synthase A* (CesA) genes of bread wheat (*Triticum aestivum*, L.)

Simerjeet Kaur¹, Kanwarpal S. Dhugga^{2, 4}, Kulvinder Gill³, Jaswinder Singh^{*1}

¹Department of Plant Science, McGill University, Sainte Anne de Bellevue, QC, Canada

² Genetic Discovery, DuPont Pioneer, 7300 NW 62nd Avenue, Johnston, IA, USA

³Department of Crop and Soil Science, Washington State University, Pullman, WA, USA

⁴Current address: International Maize and Wheat Improvement Center (CIMMYT), El Batán, Texcoco, Estado de México

*Corresponding Author

3.1 Abstract

Cellulose is the primary determinant of mechanical strength in plant tissues. Late-season lodging is inversely related to the amount of cellulose in a unit length of the stem. Wheat is the most widely grown of all the crops globally, yet information on its *CesA* gene family is limited. We have identified 22 *CesA* genes from bread wheat, which include homoeologs from each of the three genomes, and named them as *TaCesAXA*, *TaCesAXB* or *TaCesAXD*, where X denotes the gene number and the last suffix stands for the respective genome. Sequence analyses of the CESA proteins from wheat and their orthologs from barley, maize, rice, and several dicot species (Arabidopsis, beet, cotton, poplar, potato, rose gum and soybean) revealed motifs unique to monocots (Poales) or dicots. Novel structural motifs CQIC and SVICEXWFA were identified, which distinguished the CESAs involved in the formation of primary and secondary cell wall (PCW and SCW) in all the species. We also identified several new motifs specific to monocots or

dicots. The conserved motifs identified in this study possibly play functional roles specific to PCW or SCW formation. The new insights from this study advance our knowledge about the structure, function and evolution of the *CesA* family in plants in general and wheat in particular. This information will be useful in improving culm strength to reduce lodging or alter wall composition to improve biofuel production.

3.2 Introduction

Cellulose is the primary determinant of mechanical strength in plants (Appenzeller et al. 2004; Ching et al. 2006). It is also the world's most abundant renewable carbon source (Dhugga 2001; Dhugga 2007). In plants, the secondary cell wall is deposited inside the primary wall and, because of its greater thickness, it generally constitutes a majority of the vegetative biomass (Tian et al. 2014). The primary cell wall is deposited during cell division and expansion stages, whereas the secondary cell wall begins to form as cell expansion approaches cessation (Carroll et al. 2012). Cellulose in plants is synthesised by multimeric protein complexes, which consist of hexameric, rosette-like structures in the plasma membrane (McFarlane et al. 2014). Individual members of each of the hexameric components are referred to as *Cellulose synthase A (CesA)*, where the letter A stands for the catalytic subunit (Dhugga 2001). Arabidopsis (Arabidopsis thaliana) genome contains at least 10 CesA genes, which cluster into six groups (Richmond 2000; Richmond and Somerville 2000; Hamann et al. 2004). Mutational genetics established that six of the ten genes each had a nonredundant function in primary or secondary cell wall (PCW or SCW) formation. Three of the genes, AtCesA1, AtCesA3, and AtCesA6, are involved in PCW formation and another three, AtCesA4, AtCesA7 and AtCesA8 in SCW formation (Endler and Persson 2011; Hill et al. 2014). The remaining genes, AtCesA2, 5 and 9, are partially redundant with AtCesA6 (Desprez et al. 2007a; Persson et al. 2007). AtCesA10 remains uncharacterized. Maize and rice possess 13 and 11 *CesA* genes, respectively (Wang et al. 2010a; Zhang et al. 2014b). Barley has nine genes, of which *HvCesA1*, *HvCesA2*, *and HvCesA6* make PCW, and *HvCesA4*, *HvCesA7*, and *HvCesA8* from SCW (Houston et al. 2015). *HvCesA3*, *5* and *9* are different from both the groups because of their unique tissue-specific transcript levels (Burton et al. 2004). Mapping studies from Arabidopsis, maize and rice revealed that the members of the *CesA* gene family were spread across the genome although some genes were clustered together (Holland et al. 2000; Wang et al. 2010a).

Plant CESAs belong to family 2 of glycosyltransferases (GT2), which catalyse beta linkage between the glycosyl residues. CESAs are intrinsic plasma membrane proteins with their catalytic domains extending into the cytoplasm (Rayon et al. 2014). Each of the six subunits of a cellulose synthase complex (CSC) is believed to be composed of 6 enzymatically active CESAproteins. Each of the CESA proteins catalyses the synthesis of an individual β -1, 4-linked chain (Morgan et al. 2013; Slabaugh et al. 2014). Multiple chains extruded from the CSC then polymerise through hydrogen bond formation into a microfibril outside the plasma membrane.

A CESA protein of higher plants possesses eight transmembrane domains (TMDs), which are believed to form a pore in the plasma membrane to allow extrusion of the newly synthesised glucan chain. Two zinc-finger domains (ZnF), which are highly homologous to the RING-finger motif, are present on the cytoplasmic face close to the amino terminus (Kurek et al. 2002). The central or the catalytic domain is located between the second and third TMDs (Li et al. 2014a). Three aspartyl residues (referred to as D1, D2, and D3) and a QXXRW motif in the catalytic domain of the CESA proteins are conserved across all the species studied thus far. The D1 and D2 residues are believed to coordinate UDP binding while D3 provides a catalytic base for glucan chain extension (Saxena et al. 1995) The QXXRW motif acts as a binding site for the terminal disaccharide of the glucan (Morgan et al. 2013). Motifs are the conserved groups of residues in proteins, which can be associated with structural and functional variability across species. A highly conserved motif, CXXC, which is located within the ZnF, distinguishes CESAs from the CSL (cellulose synthase-like) proteins (Richmond 2000; Richmond and Somerville 2000). Crystal structure of the cellulose synthase subunit A (BcsA) and accessory protein BcsB of *Rhodobacter sphaeroides* demonstrated the involvement of a single catalytic site in the formation of the β -1,4-glycosidic bond of the glucan chain (Morgan et al. 2013). Computationally predicted model of GhCESA1 revealed two class-specific regions (C-SR-I and C-SR-II), which distinguished different CESAs, and two plant-conserved regions (P-CR), which were absent in the bacterial BcsA but highly conserved in all the plant CESAs (Ranik and Myburg 2006; Sethaphong et al. 2013; Lin et al. 2014; Slabaugh et al. 2014). The P-CR might be potentially involved in the multimerization of the plant CESA polypeptides, leading to the formation of rosettes. C-SRs are probably responsible for regulating cellulose synthesis at different developmental stages.

The *CesA* gene family has not yet been compiled from wheat, the most widely grown crop in global agriculture. Functional classification of the *CesA* genes in cereal crops has proved helpful in associating various genes with culm or stalk strength (Appenzeller et al. 2004; Houston et al. 2015). In this report, we present the *CesA* gene family from wheat. To understand the involvement of the different *CesAs* in primary or secondary wall formation in grasses or dicot plants, we have identified unique sequence motifs. Sequence comparisons of the PCW and SCW *TaCesA* genes were performed at both the DNA and protein levels. Phases of intron evolution were predicted and compared between the groups of the *TaCesA* genes involved in the formation of PCW or SCW. Unique motifs were identified among the representative monocot and dicot species. RNA-seq expression profiling of the *TaCesA* genes revealed unique, homoeolog-specific expression patterns in different tissues.

3.2.1 Hypothesis Genes involved in cellulose synthesis in primary and secondary cell walls possess unique structural and functional motifs

3.2.2 Objective I. In silico identification of true orthologs of CesA genes in wheat

3.2.3 Objective II. Comparative analysis of structural and functional conservation between genes involved in cellulose synthesis in primary and secondary cell walls

3.3 Methods and materials

3.3.1 Identification of CesAs in wheat and their true orthologs from different species

The conserved cellulose synthase domains from barley CESA proteins was used as a query to perform the tBLASTn search with Chromosome Survey Sequence (CSS) (http://plants.ensembl.org/Triticum_aestivum/Info/Index) generated by the International Wheat Genome Sequencing Consortium (IWGSC) (Mayer et al. 2014). Availability of whole genome sequence of barley (http://webblast.ipk-gatersleben.de/barley/) made it possible to isolate fulllength barley CesA sequences (Burton et al. 2004). Genome databases of Triticum urartu and Aegilops tauschii, A and D genome progenitors of wheat, respectively, were also explored to identify full-length CesA genes for the sequences missing in hexaploid wheat. The homoeologs were first identified from Ensembl Plant database followed by amino acid sequence alignment for the presence of conserved motifs and domains. Highly variable class-specific regions (C-SRs) present in different CesAs were used to differentiate the homoeologous genes from each other (Fig. 1).

Orthologs of various *CesA* genes were identified through alignment of the wheat *CesAs* with those from Arabidopsis, barley, rice and maize. The ortholog of each gene was selected based on the sequence identity and query coverage, presence of all domains and motifs similar to the query sequence, Amino acid content/size and distance among various new motifs identified in this study relative to the query sequence. Arabidopsis, rice and maize *CesA* sequences were retrieved from Phytozome v9.1: Home (http://www.phytozome.net/) (Goodstein et al. 2012).

3.3.2 Gene structure analysis

Although in this study we identified 22 *TaCesA* genes, comparative studies for gene structure were performed only for the genes that were specific for PCW and SCW cellulose synthesis. Based on analysis of the orthologs, *TaCesA4*, 7 and 8 were characterised as one-to-one orthologs of SCW-specific, *TaCesA1*, 2, and 6 as PCW-specific, and *TaCesA3*, 5 and 9 as partially redundant to the PCW *CesAs*. The homoeologous copies of each gene shared 95-99% sequence identity in addition to all the motifs and domains. Therefore only one copy among the three homoeologs was used for comparative analysis. Intron-exon boundaries and translation start and stop sites were predicted through alignments of full-length genomic copies of *TaCesA* genes with their corresponding cDNA sequences. The introns and exons were drawn to scale for all the genes as indicated by the cDNA-genomic sequence comparisons. Phases of intron evolution were predicted using Plant Intron-Exon Comparison and Evolution database (PIECE) (http://wheat.pw.usda.gov/piece/) (Wang et al. 2013)

3.3.3 Protein structure and motif identification

Amino acid sequence similarity of TaCESA protein sequences was determined by multiple sequence alignment (http://www.genome.jp/tools/clustalw/). Colour Align Conservation tool (http://www.bioinformatics.org/sms2/color_align_cons.html) was used to differentiate the conserved patterns of aligned sequences. Conserved domains and motifs were identified by manual search in the aligned sequences.

3.3.4 Phylogenetic analysis

22 newly identified wheat CESA proteins were used to deduce their phylogenetic relationships. Protein sequences for *Arabidopsis thaliana* (AtCESA), *Beta vulgaris* (BvCESA), *Eucalyptus grandis* (EgCESA), *Glycine max* (GmCESA), *Gossypium hirsutum* (GhCESA), *Hordeum vulgare* (HvCESA), *Oryza sativa* (OsCESA), *Populus trichocarpa* (PtCESA), *Solanum tuberosum* (StCESA), *Zea mays* (ZmCESA) were retrieved from NCBI (www.ncbi.nlm.nih.gov) (Kaur et al. 2013). An unrooted phylogenetic tree was constructed with bootstrap analysis over 1000 replicates, using the Neighbor-Joining method using the MEGA6 program (Saitou and Nei 1987; Tamura et al. 2013). Evolutionary distances were computed using Poisson correction method (Zuckerkandl and Pauling 1965). All positions containing gaps and missing data were eliminated.

GenBank accession numbers for CESA amino acid sequences used to generate the phylogenetic tree are: AtCESA1, AF027172; AtCESA2, AF027173; AtCESA3, AF027174; AtCESA4, AB006703; AtCESA5, AB016893; AtCESA6, AF062485; AtCESA7, AF088917; AtCESA8, AL035526; AtCESA9, AC007019; AtCESA10, At2G25540; ZmCESA1, AF200525; ZmCESA2, AF200526; ZmCESA3, NP_001292792.1; ZmCESA4, AF200528; ZmCESA5, AF200529; ZmCESA6, AF200530; ZmCESA7, AF200531; ZmCESA8, AF200532; ZmCESA9, AF200533; ZmCESA10, AY372244; ZmCESA11, AY372245; ZmCESA12, AY372246;

ZmCESA13, KJ874174; OsCESA1, AF030052; OsCESA2, D48636, OsCESA3, BAD30574; OsCESA4, AK100475; OsCESA5, BAD30574; OsCESA6, XM_477282; OsCESA7, XM_477282; OsCESA8, XM_477093; OsCESA9, XM_477093; OsCESA10, LOC_O-42g29300; OsCESA11, LOC_OS06g39970; HvCESA1, AY483150; HvCESA2, AY483152; HvCESA3, AY483151; HvCESA4, AY483154; HvCESA5/7, AY483153; HvCESA6, AY483155; HvCESA8, AY483156; HvCESA9, AK367031; PtCESA6, XP_002319002; EgCESA5, XP_010063196; StCESA3, XP_006354075; GmCESA2, XP_003531396; GhCESA5, AFB18634 and BvCESA2, XP_010678670.

3.3.5 RNA-seq expression profiling of TaCesA genes

Gene expression profiling for 21 of the wheat *CesA* genes was performed using publicly available RNA-seq data from two different databases (http://wheat-urgi.versailles.inra.fr/Seq-Repository/RNA-Seq) at McGill University and Genome Quebec Innovation Center. First dataset was a non-oriented library with five wheat organs analysed in duplicates at three development stages for each of the organs. The five organs taken into consideration with respect to developmental stages were root (at seedling, three leaves, and meiosis stages), leaf (seedling, three tillers, and 2 days after anthesis), stem (spike at 1 cm, 2 nodes, and anthesis), spike (2 nodes, meiosis, and anthesis) and grain (2, 14, and 30 days after anthesis). The second dataset was the oriented library with five wheat organs (root, leaf, stem, spike, and grain) with five conditions pooled for 4 lines per organ (Pingault et al. 2015).

The abundance of transcripts from RNA-Seq data was reported using the estimated counts quantified by a programme Kallisto (v0.42.1) (Bray et al. 2015). Counts-per-million reads were obtained using Bioconductor's edgeR (Robinson et al. 2010). Ward's linkage method was applied

to the matrix of Pearson's correlation distances to for cluster analysis. Heat map of the candidate transcripts was reported by log2 counts per million (CPM) standard deviation (Bolger et al. 2014).

3.4 Results

3.4.1 Identification and mapping of CesA gene family in wheat

We queried the Chromosome Survey Sequence (CSS) (http://plants.ensembl.org/Triticum_ aestivum/Info/Index) generated by the International Wheat Genome Sequencing Consortium to identify the orthologs of various CesA genes from bread wheat corresponding to the barley CesA sequences [32]. Twenty-two TaCesA genes were isolated, six of which were partial (S1 Text). The identified genes were named following the nomenclature of barley, which shares synteny with wheat. To simplify the nomenclature, we attached a suffix corresponding to the specific wheat genome identifier (A, B, or D) at the end of the gene number. For example, CesA1 in genomes A, B, and D is named as TaCesA1A, TaCesA1B, and TaCesA1D, respectively. As expected, we found three copies for a majority of the nine CesA orthologs corresponding to the barley genes. For CesA6, 7, and 8 we were able to find only two CesA homoeologs. Only one copy was identified for TaCesA9. The missing homoeolog of CesA6 belonged to the D genome but we obtained it from the D genome progenitor Aegilops tauschii. The TaCesA7 homoeolog, which was absent in the A genome, was recovered from the A genome progenitor Triticum urartu. We were unable to find the A genome copy of *TaCesA8* from bread wheat as well as the A genome donor, *Triticum urartu*. The three homoeologous copies of each of the CesA genes shared 95-99% sequence identity. Different CesA genes within a species possessed two highly variable class-specific regions (C-SR-I and C-SR-II) that differentiated them from each other. The wheat orthologs of the CESA proteins of other species exhibited a similarity of 70-80% at the amino acid level with Arabidopsis and 9095% with rice and barley. The *TaCesA* genes ranged from 4044 to 5251 bp in length and contained 9-13 introns. The ensembl IDs of all the newly identified wheat *CesA* genes are given in Table 1.

The newly identified wheat *CesA* genes were mapped to respective chromosomes based on the physical mapping information available in the wheat IWGSC survey sequence annotation database (http://www.wheatgenome.org/). As expected the chromosomal locations of different *CesA* genes followed the trend reported earlier in the syntenic species barley (Burton et al. 2004). *TaCesA4A*, *B*, and *D* mapped in respective genomes to chromosome 1; *TaCesA7B* and *D* to chromosome 3; and *TaCesA8B* and *D* to chromosome 5. Similarly, the homoeologs of *TaCesA1*, *2*, *3*, *5*, and *6* mapped to chromosomes 2, 5, 5, 1 and 6 of the respective genomes. However, *TaCesa9B* mapped to chromosome 2B, unlike its ortholog from barley, which is located on chromosome 6. The approximate location of *TaCesA* genes and their homoeologs is presented in Table 1.

3.4.2 DNA sequence comparison of primary and secondary cell wall TaCesA genes

On average, a PCW forming gene was longer than the one involved in SCW formation. The longest gene, *TaCesA6*, was 5251 base pairs (bp) and the shortest, *TaCesA4*, was 3923 bp in length. The size variations among different *CesA* genes arose mainly from the number and length of introns (Table 2). *TaCesA1*, 2, and 6 had 13 introns each, whereas *TaCesA4*, 7 and 8 had 7, 12, and 9 introns, respectively (Fig 2).

The introns in PCW *TaCesA1*, *2*, and *6* accounted for 1732-2026 bp of the genes, approximately double that of the 791 and 879 bp for the SCW *TaCesA4* and 8 genes. One of the SCW genes, *TaCesA7*, possessed a large total intronic region of 2095 bp, which was similar to the PCW *TaCesA* genes. Exonic regions in all the PCW forming genes (~3.2 kb) were similar in length

to those of the SCW forming genes (2.9-3.2kb). Exon-intron boundaries were random in all the genes studied, which was in contrast to the conserved boundaries reported in other species (Endo et al. 2002). The PCW and SCW genes, across groups, were 45-52% similar. Sequence similarity within the PCW and SCW groups was 54-56% and 46-63% respectively.

3.4.3 Evolution of introns in *TaCesA* gene family

Three different phases of intron evolution were predicted. Phase 0, 1, or 2 referred to the insertion of an intron between two consecutive codons, between the first and the second base or second and the third base of a codon, respectively (Csuros et al. 2011). In PCW *TaCesA* genes, all of the introns had identical phase distributions: introns 1, 3, 7, 8, 9, 10, 12, and 13 occurred in 0 phase, introns 2, 4, and 11 were in phase 1, and introns 5 and 6 occurred in phase 2. In contract, SCW *TaCesA* genes exhibited variable patterns of intron phase distribution. Introns 2, 5, 6, and 7 in *CesA4* had 0 phase distribution, introns 1 and 3 had 1, and intron 4 had a phase distribution of 2. *TaCesA7* also had introns with all three types of phase distribution; introns 2, 6, 7, 8, 9, 11, 12 were in phase 0, introns 1, 3, and 10 in phase 1, and introns 4 and 5 in phase 2. *CesA8* similarly had introns 1, 4, 5, 6, 8, and 9 in phase 0, introns 2 and 7 in phase 1, and intron 3 in phase 2 (Fig 3). The largest proportion of introns (57-66%) in all the studied genes was found to be in phase 0, followed by phase 1 (22-28%) and phase 2 (11-16%).

3.4.4 Amino acid variability of predicted TaCESA proteins

The predicted size of PCW and SCW TaCESAs ranged between 1075-1091 and 991-1055 amino acids (AA), respectively. To identify group-specific changes in primary and secondary cell wall CESA proteins, AA sequences from all TaCESAs were aligned. All the complete CESA proteins

possessed the already known, specific CESA domains, such as a ZnF (CX2-CX14-ACX2-CX4-CX2-CX7-GX3-CX2-C) near the N-terminus of the derived amino acid sequence (S2 Text). All the TaCESAs possessed eight TMDs; two towards the N-terminus and six near the C-terminus, as well as the conserved D, DXD, D, QXXRW signatures (Fig 1).

Major differences among TaCESAs resulted from the deletion of up to 45 AAs in hypervariable regions. The N-terminal of the PCW TaCESAs possessed more highly conserved motifs and fewer deletions in comparison to the SCW TaCESAs. ZnF consisted of 46 AAs in the predicted TaCESAs, with the exception of an 8 AAs deletion in TaCESA7 and its homoeologs, resulting in the following domain: CX2-CX6-ACX2-CX4-CX2-CX7-GX3-CX2-C as compared to the known domain (CX2-CX12-FXACX2-CX2PXCX2-CXEX5-GX3-CX2C), where X is any amino acid (Cosgrove 2005). Four of the TaCESAs out of 22 were missing the ZnF as did TaCESA9 because they were incomplete on the N-terminal end.

3.4.5 New motifs distinguishing PCW CESAs from SCW CESAs

A new motif distinguishing the PCW CESAs from the SCW CESAs was found within the ZnF. The motif, CQIC, was identified within the small motif, CXXC, reported earlier for differentiating CESAs from the CSL genes [8]. This motif was present in all the PCW TaCESAs. Although SCW TaCESAs also possessed a "CXXC" motif, the two middle amino acids in these proteins were variable. In the SCW TaCESA4, the polar amino acid glutamine was replaced by the negatively charged amino acid, glutamate; in TaCESA7, both the amino acids were replaced by the marginally hydrophobic amino acid alanine; and in TaCESA8, glutamine was replaced by a highly basic (positively charged) amino acid, arginine, and isoleucine was replaced by a relatively conservative substitution of alanine (Fig 4). Another conserved motif, SVICEXWFA, was located

within the second transmembrane domain in all the PCW CESAs. In the SCW-specific CESAs, TaCESA4, 7, and 8 this motif was variable but all the amino acid replacements were conservative. For example, isoleucine, a hydrophobic amino acid next to glutamate was replaced by an iso-amino acid, leucine, in CESA4; alanine was replaced by glycine, both somewhat hydrophobic, in CESA7; and valine and isoleucine, both hydrophobic amino acids, switched places in CESA8.

3.4.6 Conservation of motifs in monocots and dicots

The two motifs, CQIC and SVICEXWFA, distinguished the PCW from the SCW CESAs (Fig 4). That these motifs were conserved was confirmed by analysing the CESA proteins in the PCW and SCW groups from dicot (Arabidopsis) and monocot (barley, maize, rice, and wheat) species. Alignment results demonstrated that the CQIC and SVICEXWFA motifs were completely conserved only in the PCW-specific CESAs in all the plant species studied. The completely conserved amino acid residues in each motif across all the CESA proteins were CXXC and SXXCEXWF (Fig 1).

3.4.7 Unique motifs conserved among the CESA orthologs from different species

Motif analysis was performed by aligning CESA proteins from Arabidopsis, barley, maize, rice and wheat. Arabidopsis CESA4 and its orthologs from wheat, barley, maize, and rice exhibited 73-74% sequence similarity. In the case of SCW, nine motifs ranging from 2-15 amino acid residues in length provided ortholog-specific identity to the SCW CESAs from different species (Fig 5). These motifs were highly conserved among the orthologs from the five species analysed in this study. Only one gene from each species, with the exception of maize which had two closely related copies for one of the three SCW genes (CESA12 and 13), shared these motifs including a dicot, Arabidopsis. This suggests that the genes for SCW had already duplicated before the separation of monocots and dicots. The number of amino acid residues among most of these motifs was also conserved among different species (Fig 5). CESA7 and 8 from wheat showed 71-75 % and 77-79 % sequence similarity with the corresponding orthologs from different species, respectively. Although the motifs were unique for CESA4, 7 and 8, they were highly conserved among the orthologs from different species (Fig 5).

Two PCW CESAs, AtCESA1, 3 and their orthologs from other species differed from AtCESA6 and its orthologs in structural features. AtCESA1 and 3 were highly similar (77-79%) to the corresponding orthologs from barley, maize, rice and wheat. Four motifs in TaCESA6 and three in TaCESA1 orthologs differentiated them from each other and all other CESAs (Fig 6).

3.4.8 Motifs differentiating CESAs from monocots and dicots

Arabidopsis CESA6 and its orthologs from other species in this study exhibited 68-70% sequence similarity but lacked any specific patterns that could differentiate them from the other CESAs. However, this group possessed motifs that were only conserved in the orthologs from monocots (grasses) but not in Arabidopsis. To confirm the specificity of these motifs for grasses, we retrieved the sequences of TaCESA2 orthologs from seven dicot species: *Arabidopsis thaliana* (AtCESA6), *Beta vulgaris* (BvCESA2), *Eucalyptus grandis* (EgCESA5), *Glycine max* (GmCESA2), *Gossypium hirsutum* (GhCESA5), *Populus trichocarpa* (PtCESA6) and *Solanum tuberosum* (StCESA3). The CESA2 and its orthologs from grasses were compared with its orthologs from dicot species. For this particular gene, nine motifs were highly conserved in the orthologs from grasses (Fig 7). But in dicots, these motifs were replaced by variable amino acid residues.

3.4.9 Phylogenetic analysis

The evolutionary history of the CESAs was inferred from the analysis involving 70 CESA protein sequences from different species. An unrooted phylogenetic tree revealed that the orthologs from Arabidopsis, barley, beet, cotton, maize, poplar, potato, rice, rose gum, soybean and wheat were grouped together. Branch lengths, which are indicative of the evolutionary distances were used to interpret the phylogenetic tree (Fig 8). The paralogs from various species were grouped in different clades from those of the orthologs. This suggests, again, that divergence of the *CesA* genes had occurred prior to the separation of monocots and dicots.

3.4.10 RNA-seq analysis of *TaCesA* genes

Gene expression of 21 of the 22 *TaCesA* genes was studied in five organs at three development stages. We left out the *TaCesA9* gene because it was represented by a highly truncated cDNA. A heat map displaying transcript abundance of the *CesA* genes from different wheat tissues is shown in Fig 9.

Transcript abundance data revealed the presence of two distinct groups. Group, I consisted of *TaCesA4A*, *B*, *D*, *TaCesA7B*, *D* and *TaCesA8B*, *D* genes, all involved in SCW synthesis. These genes were highly expressed in the mature tissues, for example, stem collected soon after anthesis, and at very low levels in the PCW formation (Fig 9). For example, *TaCesA7B*, *D* and *TaCesA8B*, *D* genes were expressed at extremely low levels in the spike and grain tissues (Fig 9).

Group II comprised the PCW *TaCesA* genes: *TaCesA1*, 2, 3, 5 and 6 along with their homoeologs from A, B and D genomes. These genes were expressed at lower levels in the mature tissues and at relatively high levels in the PCW forming cells (Fig 9). For example, all three

homoeologous copies of the *TaCesA3* gene were expressed in the grain and the leaf tissues. These genes were expressed moderately in the developing grain, which agrees with grain having a relatively low cell wall fraction. The expression of the *TaCesA5A* and *B* genes was highest in the grain tissues from 14 and 30 DAAs, whereas the *TaCesA5D* was moderately expressed in these tissues. The expression of *TaCesA5D* homoeolog was dramatically lower in the leaf tissues at 2 days after anthesis (DAA), whereas *TaCesA1D* was expressed at higher level. The transcript abundance of *TaCesA1A* was highest in the grain tissues at 2DAAs whereas *TaCesA6B* homoeolog was moderately expressed. The expression level of *TaCesA1B* was moderate in the root and grain tissues.

3.5 Discussion

Cellulose consists of paracrystalline microfibrils of multiple, unbranched β -1, 4-glucan chains, which are synthesised by the individual CESA polypeptides in the plasma membrane-localized rosette. *CesA* is a multigene family consisting of more than eight members in higher plants (Suzuki et al. 2006). Structure and function of the *CesA* genes in wheat remain undocumented. Most studies about structural and functional characterization of *CesAs* have been performed in Arabidopsis (Arioli et al. 1998; Richmond and Somerville 2000; Taylor et al. 2003), maize (Holland et al. 2000; Appenzeller et al. 2004), and rice (Tanaka et al. 2003; Wang et al. 2012a). Bread wheat, an allohexaploid, has a complex genome, ~17 Gb in size, ~80-90% of which consists of repetitive DNA (Mayer et al. 2014). The availability of large-scale genomic sequence information and conserved synteny between barley and wheat is valuable in exploring wheat gene function and structure (Mochida and Shinozaki 2013). In barley, the *CesA* gene family consists of nine genes (*HvCesA1* to *HvCesA9*. Three genes, *HvCesA1*, *HvCesA2*, and *HvCesA6*, are expressed during

primary wall formation, and another three, *HvCesA4*, *HvCesA7*, and *HvCesA8*, during secondary wall formation (Burton et al. 2004). In this report, we document 22 *CesA* genes from wheat, which we identified using a comparative genomics approach using barley sequences as anchors. As expected, most of the *TaCesA* genes each have three paralogs in the homoeologous genomes A, B and D. Four of the 22 genes deviated from this pattern: only one paralog was identified for *TaCesA9*, and two each for *TaCesA6*, *7*, and *8*. One of the genes, *TaCesA2*, had two paralogous copies on chromosomes 5B and 5D but the third on chromosome 4A, which was most likely because of a translocation between chromosomes 5A and 4A (Table 1) (Ma et al. 2013).

All the CESAs possess domains known to be highly conserved among all the plant species studied thus far (Richmond and Somerville 2000). Sequences in the non-conserved domains, however, are useful for the identification of the orthologs of individual *CesA* genes (Table 3). In the case of gene families, it is often difficult to determine true orthology among different species solely based on sequence similarity. Many previous studies reported *CesA* orthologs based on phylogenetic analyses (Burton et al. 2004; Wang et al. 2010a). We supplemented the phylogenetic analysis as a tool for the identification of the *CesA* orthologs by searching for the conserved motifs in addition to the ones already known (Ma et al. 2013).

Knowledge about the conserved structural motifs that can distinguish *CesA* genes involved in PCW and SCW formation as well as *CesAs* between monocots and dicots is limited. Distinct patterns of intron placement, removal, and the phases of insertion in *TaCesA* genes suggested that the phases of intron insertion remained conserved during the evolution of these genes (Trapp and Croteau 2001). Deviation of phase distribution from the expected 33% suggested a bias in intron insertions towards the 0 phase, that is, between the codons rather than within the codons (Csuros et al. 2011).

The motif CQIC in ZnF distinguishes the PCW and SCW CESAs from both the monocots and dicots. Distinct CSCs for the synthesis of primary and secondary cell walls have been reported (Arioli et al. 1998; Tanaka et al. 2003; Taylor et al. 2003). The high level of conservation of the CQIC motif suggests that it is possibly related to cellulose synthesis. This concurs with the observation in other major gene families, where domains and motifs were conserved during the evolution (Arioli et al. 1998; Taylor et al. 2003).

A similar trend of intron phase distribution and motif conservation was observed when we compared CESA1 of *Arabidopsis thaliana* with its orthologs from angiosperms (*Arabidopsis lyrata*, *Aquilegia coerulea*, *Brachypodium distachyon*, *Carica papaya*, *Citrus clementina*, *Citrus sinensis*, *Cucumis sativus*, *Eucalyptus grandis*, *Glycine max*, *Manihot esculenta*, *Medicago truncatula*, *Mimulus guttatus*, *Oryza sativa*, *Populus trichocarpa*, *Physcomitrella patens*, *Prunus persica*, *Ricinus communis*, *Setaria italica*, *Sorghum bicolor*, *Vitis vinifera*, *Zea mays*), Chlorophytes (*Chlamydomonas reinhardtii*, *Volvox carteri*), and pteridophyte (*Selaginella moellendorffii*).

We also identified new, highly conserved motifs among the CESA orthologs of five species (Arabidopsis, barley, maize, rice and wheat). Despite the variable protein sequence of each member of the CESA family among the orthologs from various species, the organisation of the motifs remained conserved.

RNA-seq expression profiling revealed that the three SCW genes (*TaCesA4*, *7*, *8*) and their homoeologs were co-expressed in the mature stem tissues (Fig 9). This observation provided support for these genes being functionally orthologous to the secondary wall-forming genes from other species, for example, Arabidopsis (Arioli et al. 1998; Richmond and Somerville 2000; Taylor

et al. 2003), barley (Burton et al. 2004), maize (Holland et al. 2000; Appenzeller et al. 2004), and rice (Tanaka et al. 2003; Wang et al. 2012a). Five genes (*TaCesA1, 2, 3, 5, and 6*) and their homoeologs from the A, B and D genomes of wheat constituted a second group involved in PCW synthesis.

Most of the *TaCesA* genes were differentially expressed among three different genomes of bread wheat, which is a common phenomenon in hexaploid wheat (Mochida et al. 2004). This differential expression pattern is attributable to the genetic divergence of paralogous genes during the evolution (Takata and Taniguchi 2015). *TaCesA* genes are distributed across the wheat genome (Fig 10). Similar distribution patterns were observed in Arabidopsis, barley and maize (Holland et al. 2000; Burton et al. 2004).

Our study compiles a list of the *CesA* genes in bread wheat, classified them into PCW and SCW formation, and maps them to the chromosomes. This information will be useful in breeding wheat for culm strength and biofuel-related traits.

3.6 Conclusion

We have identified 22 *CesA* genes from bread wheat and compared them with their orthologs from Arabidopsis, barley, maize, and rice. New structural motifs were identified, which allowed differentiation of the CESA proteins for their roles in primary or secondary wall (PCW or SCW) formation in higher plants. Further characterization of the motifs would be needed, however, to establish their respective biological roles. Several new motifs identified in this study would be useful as signatures for the identification of orthologs of the *CesA* genes from various species. The compilation of the *CesA* gene family in bread wheat along with the expression patterns and

genomic map positions of individual members will be helpful in improving culm strength for reduced lodging as well as improving the straw for biofuels.

Fig 3.1 Predicted protein features of wheat cellulose synthase genes. The numbers 1 to 8 in the purple rectangles refers to the transmembrane domains (TMDs). Black triangles localise the conserved motifs. The newly identified motifs CXXC and SXXCEXWF are highlighted in blue and previously reported motifs in black.



Fig 3.2 Structural features of the *TaCesA* genes. Drawn to scale, exons are represented by black boxes and introns by grey lines. Intron lengths are presented on top of each intron. PCW and SCW *CesA* genes are shown in blue and red colours, respectively.



Fig 3.3 Amino acid sequence alignment of wheat CESA proteins. Drawn to scale with solid lines representing conserved amino acid sequences and the gaps representing the mismatches and deletions. Corresponding phases of intron evolution (0, 1, and 2) for the CESA proteins are shown on the top of the solid lines. Primary and secondary cell wall CESAs are shown in blue and red colour, respectively.



Fig 3.4 Motifs differentiating PCW and SCW CESA orthologs from different species. Conserved and non-conserved amino acids residues are highlighted in red and green respectively. Amino acid changes in the motifs are shown in blue.

PCW	AtCESA1	38CQIC264SVICEIWFA
	TaCESA6	18CQIC285SVICEIWFA
	HvCESA6	40CQIC257SVICEIWFA
	OsCESA1	40CQIC259SVICEIWFA
	AtCESA3	19CQIC267SVICEIWFA
	TaCESA1	33CQIC280SVICEIWFA
	HvCESA1	18CQIC285SVICEIWFA
	OsCESA8	18CQIC285SVICEIWFA
	ZmCESA4	18CQIC283SVICEIWFA
	ZmCESA9	18CQIC285SVICEIWFA
	TaCESA2	38CQIC266SVICEIWFA
	HvCESA2	38CQIC266SVICEIWFA
	OsCESA3	38CQIC268SVICEIWFA
	OsCESA5	38CQIC267SVICEIWFA
	ZmCESA7	38CQIC262SVICEIWFA
SCW	AtCESA4	22CKVC220SVICEIWFA
	TaCESA4	13CRAC220SVICELWFA
	HvCESA4	10CRAC220SVICELWFA
	OsCESA7	17CRVC223SVICELWFA
	ZmCESA10	40CRVC220SVICELWFA
	AtCESA7	36CEIC229SVICEIWFA
	TaCESA8	36CEIC257SIVCEIWFA
	HvCESA8	36CEIC259SIVCEIWFA
	OsCESA9	36CEIC258SIICEIWFA
	ZmCESA12	36CEIC256SIICEIWFA
	ZmCESA13	36CEIC256SIICEIWFA
	AtCESA8	08CNTC203SVICEIWFA
	TaCESA7	08CAAC210SVICEIWFG
	OsCESA4	08CAAC208SVICEIWFG
	ZmCESA11	09CAAC201SVICEIWFG

Fig 3.5 Conserved motifs differentiating the orthologs of SCW CESAs from *Triticum aestivum* (TaCESA), *Arabidopsis thaliana* (AtCESA), *Hordeum vulgare* (HvCESA), *Oryza sativa* (OsCESA), and *Zea mays* (ZmCESA). Conserved and non-conserved amino acids residues are highlighted in red and green respectively. Amino acid changes in the motifs are shown in blue.

AtCESA4 138.SVLGKDFEAER.6.EWKERVDK.6.KRG.27.LWR.27.RFR.182.KKP.25.VYL.80.KKL.143.YDEL

TaCESA4 128.SVAGKELEAER.6.EWKERIDK.6.KRG.28.LWR.27.RFR.182.KKP.25.VYL.80.KKL.145.YDEL

HvCESA4 125.SVAGKELEAER.6.EWKDRIDK.6.KRG.28.LWR.27.RFR.182.KKP.25.VYL.80.KKL.145.YDEL

OsCESA7 136.SVAGKDLEQER.6.EWKDRIDK.6.KRG.27.LWR.27.KFR.182.KKP.25.VYL.80.KKL.157.YDEL

ZmCESA10 156.SVAGKDLEAER.6.EWKDRIDK.6.KRG.27.LWR.27.KFR.182.KKP.25.VYL.80.KKL.152.YDEL

AtCESA7 36.CEIC.71.NIE.11.AMLYGK.15.PPVI.20.LHKRVHPYP.11.ERMDD.486.EGGVPPSSS.89.HSPLW

TaCESA8 36.CEIC.71.NIE.11.AMLYGK.15.PPVI.20.LHKRVHPYP.19.ERMDD.485.EGGVPPSSS.89.HSPLL

HvCESA8 36.CEIC.71.NID.17.AMLHGK.21.PPII.26.LHKRIHPYP.19.ERMDD.485.EGGVPPSSS.89.HSPLL

OsCESA9 36.CEIC.71.NID.17.AMLHGK.23.PPII.26.LHKRIHPYP.18.ERMDD.485.EGGVPPSSS.89.HSPLL

ZmCESA12 36.CEIC.71.NID.19.AMLHGR.18.PPII.26.LHKRIHPYP.18.ERMDD.487.EGGVPPSSS.89.HSPLL

ZmCESA13 36.CEIC.71.NID.19.AMLHGR.18.PPII.26.LHKRIHPYP.18.ERMDD.487.EGGVPPSSS.89.HSPLL

ZmCESA13 36.CEIC.71.NID.19.AMLHGK.18.PPII.26.LHKRIHPYP.18.ERMDD.487.EGGVPPSSS.89.HSPLL

ZmCESA13 36.CEIC.71.NID.18.AMLHGK.18.PPII.26.LHKRIHPYP.18.ERMDD.487.EGGVPPSSS.89.HSPLL

ZmCESA13 103.ELNDE.3.PIWKNRVESWKDKK.82.SAYG.173.TP.32.ARD.74.DVC.107.NYDEY.21.IE.15.ST

TaCESA7 118.ELNDE.3.PIWKNRVESWKEKK.74.SAFG.172.TP.32.ARD.74.DVC.108.NYDEY.21.IE.15.ST

GCESA4 116.ELNDE.3.PIWKNRVESWKEKK.74.SAFG.172.TP.32.ARD.74.DVC.111.NYDEY.21.IE.15.ST

JCESA1 110.ELNDE.3.PIWKNRVESWKEKK.74.SAFG.172.TP.32.ARD.74.DVC.108.NYDEY.21.IE.15.ST

Fig 3.6 Conserved motifs differentiating the orthologs of PCW CESAs from *Triticum aestivum*

(TaCESA), Arabidopsis thaliana (AtCESA), Hordeum vulgare (HvCESA), Oryza sativa

(OsCESA), and Zea mays (ZmCESA). Conserved and non-conserved amino acids residues are

highlighted in red and green respectively. Amino acid changes in the motifs are shown in blue.



Fig 3.7 Monocots and dicots specific motifs of CESA orthologs from *Triticum aestivum* (TaCESA), *Arabidopsis thaliana* (AtCESA), *Beta vulgaris* (BvCESA), *Eucalyptus grandis* (EgCESA), *Glycine max* (GmCESA), *Gossypium hirsutum* (GhCESA), *Hordeum vulgare* (HvCESA), *Oryza sativa* (OsCESA), *Populus trichocarpa* (PtCESA), *Solanum tuberosum* (StCESA) and *Zea mays* (ZmCESA). Conserved and non-conserved amino acids residues are highlighted in red and green respectively. Amino acid changes in the motifs are highlighted in blue.

TaCESA2 125.ESML.22.PNV.7.MVDD.61.QKQER.112.FDK.307.PPSR.45.AYAL.16.IVNQQ.251.ELYTF HvCESA2 125.ESML.22.PNV.7.MVDD.61.QKQER.112.FDK.307.PPSR.45.AYAL.16.IVNQQ.251.ELYTF MONOCOTS Oscesa3 125.esml.23.PNV.7.MVDD.61.QKQER.113.FDK.307.PPSR.45.AYAL.16.IVNQQ.251.ELYTF OsCESA5 125.ESML.22.PNV.7.MADD.61.QKQER.113.FDK.307.PPSR.45.AYAL.16.IVNQQ.251.ELYTF OsCESA6 129.ESML.18.PNV.7.MVDD.64.QKQER.111.FDK.307.PPSR.44.AYAL.16.IVNQQ.251.ELYTF ZmCESA6 095.ESML.22.PNV.7.MVDD.61.QKQER.109.FDK.307.PPSR.46.AYAL.16.IVNQQ.251.ELYTF ZmCESA7 124.ESML.22.PNV.7.MVDD.61.QRQER.109.FDK.307.PPSR.44.AYAL.16.IVNQQ.251.ELYTF ZmCESA8 131.ESML.19.PNV.7.MVDD.64.QKQER.110.FDK.307.PPSR.44.AYAL.16.IVNQQ.251.ELYTF PtCESA6 130.LGGP.18.PQV.7.MVPS.81.QKQDN.109.YEK.307.PPTR.42.ALEG.14.VTSEQ.251.ELYAF DICOTS EgCESA5 130.EAML.20.PQV.7.MVDD.67.QKQEK.113.YEK.307.PPTR.45.PLEG.12.PTPQH.251.ELYAF StCESA3 129.DYFE.12.PQV.7.MHYH.65.KKQEK.107.YEK.307.APSR.37.SLAL.13.LISDH.251.ELYAF GmCESA2 127.ESLY.28.SDI.7.EDPE.62.RQSDK.114.YEK.307.PPSK.41.ALEN.14.NLTQT.251.ELYIF GhCESA5 124.EAML.29.SQI.7.EHSE.62.WQNEK.114.YEK.307.PPGK.42.ALEN.14.EASQI.251.ELYLF AtCESA6 126.EGMS.19.SQI.7.EDVE.63.KQNEK.111.YEK.307.GPRK.40.ALEN.16.EAMQM.251.DLYLF BvCESA2 126.EAIY.28.SEQ.7.EDTG.62.RQNDR.114.YEK.307.PIGK.49.ALEN.14.LMPQV.251.DLYLF Fig 3.8 Unrooted phylogenetic tree of the CESAs from *Triticum aestivum* (TaCESA),

Arabidopsis thaliana (AtCESA), *Beta vulgaris* (BvCESA), *Eucalyptus grandis* (EgCESA), *Glycine max* (GmCESA), *Gossypium hirsutum* (GhCESA), *Hordeum vulgare* (HvCESA), *Oryza sativa* (OsCESA), *Populus trichocarpa* (PtCESA), *Solanum tuberosum* (StCESA) and *Zea mays* (ZmCESA). The bar provides a scale for the branch length in the horizontal dimension. The line segment with the number '0.1' means that an equal length of the branch between the CESA proteins represents a change of 0.1 AA. Color codes for different species: Red - TaCESA, blue – AtCESA, purple - HvCESA, yellow - ZmCESA, green - OsCESA, and grey – BvCESA, EgCESA, GmCESA, GhCESA, PtCESA, StCESA.


Fig 3.9 Heat map of 21 *CesA* transcripts by log2 counts per million (CPM) standard deviation in hexaploid wheat.



Fig 3.10 Map positions of *TaCesA* genes in the wheat genome. The exact locations are shown in Table 1.



CesA gene	Map position (MB)	Ensembl ID
TaCesA1A	NA	Traes_2AS_665AF9500.1
TaCesA1B	23.50	Traes_2BS_064B02A89.3
TaCesA1D	18.70	Traes_2DS_C80293002.1
TaCesA2A	176.7	Traes_4AL_941C0E3EF.2
TaCesA2B	262.70	Traes_5BL_3A1A752B7.1
TaCesA2D	151.78	Traes_5DL_3B0E69498.2
TaCesA3A	125.80	Traes_5AL_E176291CC.1
TaCesA3B	247.07	Traes_5BL_CFCBFDA99.2
TaCesA3D	144.92	Traes_5DL_BBFD06D43.1
TaCesA4A	93.16	Traes_1AL_F420A1BBE.1
TaCesA4B	48.70	Traes_1BL_B34FCB150.1
TaCesA4D	NA	Traes_1DL_129574E44.1/ EMT11949
TaCesA5A	29.31	Traes_1AS_10C467127.1
TaCesA5B	103.00	Traes_1BS_64E9CC6E0.1
TaCesA5D	NA	Traes_1DS_65C1FDCD8.2
TaCesA6A	9.23	Traes_6AS_CF6D8CD28.2
TaCesA6B	25.84	Traes_6BS_8DA635027.1
TaCesA7B	514.04	TRAES3BF028900030CFD_t1
TaCesA7D	42.80	Traes_3DL_B2FD2FBFA
TaCesA8B	163.48	Traes_5BL_51C858A97.1
TaCesA8D	60.35	Traes_5DL_E82D6D246.2
TaCesA9B	NA	Traes_2BS_9B34A7A43.2

Table 3.1 CesA genes and their chromosomal locations in hexaploid wheat.

NA- Precise location of these genes on the respective chromosomes is not known because of the incomplete assembly of the wheat genome.

~ .	PCW or SCW	Gene length (nt)	Introns (#)	ORF length	Map to
<i>CesA</i> gene				(AA)	Chromosome
TaCesA1	PCW	5175	13	1080	2
TaCesA2	PCW	5005	13	1091	5
TaCesA3	PCW	5127	13	1105	5
TaCesA4	SCW	3923	7	1044	1
TaCesA5	PCW	4,085	14	1078	1
TaCesA6	PCW	5251	13	1075	6
TaCesA7	SCW	5072	12	991	3
TaCesA8	SCW	4044	9	1055	5
TaCesA9	PCW	2184	5	537	2

Table 3.2 Structures of the *TaCesA* genes for PCW and SCW synthesis.

Table 3.3 *TaCesA* genes and their orthologs from Arabidopsis, barley, maize, and rice involved in the formation of the primary cell wall (PCW) or secondary cell wall (SCW).

Gene Function	Wheat	Barley	Maize	Rice	Arabidopsis
PCW	<i>CesA5</i> , 6 and 9	CesA6 and 9	<i>CesA1</i> , 2 and 3	CesA1	CesA1 and 10
	CesA1 and 3	CesA1 and 3	CesA4, 5, and 9	CesA2, 8,10 and 11	CesA3
	CesA2	CesA2	<i>CesA6</i> , 7, and 8	<i>CesA3, 5,</i> and 6	CesA2, 5, 6, and 9
SCW	CesA4	CesA4	CesA10	CesA7	CesA4
	CesA8	CesA8	<i>CesA12</i> and <i>13</i>	CesA9	CesA7
	CesA7	CesA5 and 7	CesA11	CesA4	CesA8

CONNECTING STATEMENT FOR CHAPTER IV

Chapter IV, entitled "Functional characterization of secondary cell wall specific *CesA4* gene in bread wheat using Virus-Induced Gene Silencing (VIGS)" authored by Simerjeet Kaur, Kanwarpal S. Dhugga, Raj Duggavathi, Kulvinder S. Gill, and Jaswinder Singh has been submitted to "*Cellulose*".

As discussed in chapter III, Cellulose Synthase Complexes (CSCs) in secondary cell walls of wheat plants are composed of three genes *TaCesA4*, *TaCesA7*, and *TaCesA8*. These three genes co-expressed in the mature stem tissues of bread wheat. But the relative transcript abundance was found to be higher for *TaCesA4* genes, which indicates its major role in the secondary cell wall cellulose synthesis. However, the function of this gene requires further attention which could provide further understanding of cellulose synthesis in secondary cell walls. In study IV, the biological role of *TaCesA4* gene has been functionally evaluated using Virus-induced gene silencing (VIGS) approach. Three homoeologs (*TaCesA4A*, *TaCesA4B*, and *TaCesA4D*) were silenced collectively in bread wheat using the *TaCesA4* specific oligo designed from the conserved region of these homoeologs. Silenced plants showed a significant reduction in transcript abundance and cellulose content in the stem tissues. However, the anatomy of stem cross sections of silenced plants did not show any evidence of abrupt changes in the secondary cell wall of stems at the booting stage.

Chapter IV. Functional characterization of secondary cell wall specific *CesA4* gene in bread wheat using virus-induced gene silencing (VIGS).

Simerjeet Kaur¹, Kanwarpal S. Dhugga², Raj Duggavathi⁴ Kulvinder Gill³, Jaswinder Singh^{*1}

¹Department of Plant Science, McGill University, Sainte Anne de Bellevue, QC, Canada

² International Maize and Wheat Improvement Center (CIMMYT), El Batán, Texcoco, Estado de México

³Department of Crop and Soil Science, Washington State University, Pullman, WA, USA ⁴Department of Animal Science, McGill University, Sainte Anne de Bellevue, QC, Canada *Corresponding Author

4.1 Abstract

Plant cell walls produce a bulk of renewable biomass vital for food, feed and biofuels. Cell wall consists of three layers including middle lamella, secondary cell wall and primary cell wall. The secondary cell wall is a thicker layer developed inside the primary cell wall. Cellulose is the main carbohydrate of the primary and secondary cell walls and involved in the shape, structure and strength of the plant. Recently, three major genes (TaCesA4, TaCesA7 and TaCesA8) have been identified in wheat which are appeared to play important roles in the synthesis of the secondary cell wall. In this study, efforts have been made to functionally characterize TaCesA4 using Virus-induced gene silencing (VIGS) approach. Inoculations were performed at the booting stage of the bread wheat (Chinese spring) plants grown under the temperature regime of 220C days and 180C nights with 23-50% relative humidity and 16 hrs of light. Quantification of the transcript at 21 dpi (Days post inoculation) revealed 87.17% decrease of TaCesA4 transcripts in the first internode of

silenced plants. Around 30% decline in the cellulose content was observed in silenced plants as compared to controls. However, negligible anatomical differences in the shape of cells and the arrangement of vascular bundles were observed between the stem cross-sections of silenced and control plants.

4.2 Introduction

In plants, cellulose microfibrils are known to be synthesized by a heteromeric rosette complex known as cellulose synthase complex (CSC). Each subunit of CSC has six cellulose synthase (CESA) isoforms that bound to the plasma membrane and catalyze the polymerization of β -1, 4glucans using UDP-glucose as a substrate. CESA isoforms have been encoded by different cellulose synthase A (CesA) genes that play an important role in the synthesis of cellulose in primary and secondary cell wall (Endler and Persson 2011). In the case of Arabidopsis thaliana, CesA1, CesA3, and CesA6 are involved in primary wall and CesA4, CesA7 and CesA8 appear to be required for cellulose synthesis in the secondary wall (Endler and Persson 2011). Cellulose in the primary cell wall determines the shape of cells, which is laid down during plant growth (Wasteneys 2004). The cellulose in secondary cell walls is deposited after the cell stop growing, because of its greater thickness, constitutes the bulk of terrestrial biomass (Joshi and Mansfield 2007). Moreover the higher degree of polymerization increase microfibril crystallinity of cellulose in secondary cell wall which determines the physical strength of the plant (Saxena and Brown 2005). Genes involved in secondary cell wall thickening are important candidates to study the genetic variability between diverse genotypes (Tian et al. 2014). Based on the structural, evolutionary and expression analysis of CesA genes in wheat, TaCesA4, TaCesA7 and TaCesA8 genes are appeared to play distinctive roles in the synthesis of the secondary cell wall (Taylor et al. 2003). Among secondary cell wall forming CesAs, higher transcript abundance of TaCesA4

has been observed in mature stem tissues of wheat (Kaur et al. 2016). Genetic evidence for the role of these genes in secondary cell wall formation came from the Arabidopsis irregular xylem mutants, *irx1 (AtCesA8)*, *irx3 (AtCesA7)*, and *irx5 (AtCesA4)*, showing collapsed mature xylem cells due to lowered content of secondary cell wall cellulose (Taylor et al. 2003). Several brittle culm retrotransposons and EMS mutants for secondary cell wall CesAs in rice and a spontaneous brittle stalk-2 mutant in maize showed a significant reduction in cellulose content as compared to wild-type plants (Ching et al. 2006; Zhang et al. 2009; Kotake et al. 2011; Wang et al. 2012a; Kaur et al. 2016). Gene expression studies coupled with reverse genetic approaches is a preferred method to functionally and rapidly annotate a particular gene (Held et al. 2008). In the current study, efforts have been made to functionally validate the role of *CesA4* gene in the wheat tissues using Virus-Induced Gene Silencing (VIGS) approach. VIGS is one of the powerful plant functional genomics tools (Singh et al. 2006; Bennypaul et al. 2012a; Singh et al. 2013) that exploits an RNA-mediated antiviral defence mechanism and triggers targeted gene silencing. VIGS is a fast and cost effective alternative to examining the function of uncharacterized genes, especially in polyploid crops, where stable transformation through RNAi is difficult to perform (Senthil-Kumar and Mysore 2011; Bhullar et al. 2014). Infection of plants by a virus engineered with fragments of a gene of interest activates post-transcriptional gene silencing as an innate defence response. Barley stripe mosaic virus (BSMV) is a single-stranded RNA virus consisting of tripartite α , β and γ genome. DNA plasmids carrying full-length cDNA clones of these three RNAs were constructed from BSMV strain ND18 (Petty et al. 1989). The insertion of a 178-bp fragment of the barley phytoene desaturase (PDS) gene into γ construct of BSMV resulted in the silencing of *PDS* with obvious phenotype after infection (Holzberg et al. 2002). The BSMV-based VIGS system was previously shown to silence the three homoeologous copies of a gene in wheat

efficiently (Bennypaul et al. 2012b). A similar approach has been utilized here to characterize *CesA4* genes in wheat.

4.2.1 Hypothesis *Cellulose synthase A 4 (CesA4)* gene in wheat is required for the deposition of cellulose in mature stem tissues.

4.2.2 Objective I. Generation of appropriate constructs for VIGS

4.2.3 Objective II. Functional validation of *CesA4* gene in wheat using the VIGS system

4.3 Materials and methods

4.3.1 TaCesA4 gene structure analysis

Full gene sequences of three homoeologs of *TaCesA4* were downloaded from Ensemblplants (http://plants.ensembl.org/Triticum_aestivum) (Kaur et al. 2016). Multiple sequence alignments were performed using Clustal omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) (Sievers et al. 2011). The sequence manipulation suite: Color align conservation (http://www.bioinformatics.org/sms2/color_align_cons.html) was used to highlight the conserved regions of *TaCesA4* used to synthesize VIGS construct (Stothard 2000). Gene structure was predicted using the gene structure display server 2.0 (http://gsds.cbi.pku.edu.cn/) via the genomic and cDNA sequences of *TaCesA4* homoeologs.

4.3.2 In silico expression analysis of TaCesA homoeologs

Publicly available RNA-seq data generated from hexaploid bread wheat (var. *Chinese spring*) was used to predict the expression of *TaCesA4* homoeologs. The data was compiled from five different wheat tissues (Spike, root, leaf, grain and stem) collected at three stages of wheat development. The developmental stages with respect to each organ were reported in Zodoks scale; spike_z32

(two nodes), spike_z39 (meiosis), spike_z65 (anthesis), root_z10 (seedling), root_z13 (three leaves), root_z39 (meiosis), leaf_z10 (seedling), leaf_z23 (three tillers), leaf_z71 (2 days after anthesis), grain_z71 (2 days after anthesis), grain_z75 (14 days after anthesis), grain_z85 (30 days after anthesis), stem_z30 (spike at 1 cm), stem_z32 (two nodes), stem_z65 (anthesis). The relative expression of each *TaCesA4* homoeolog was presented as fragments per kilo base of transcript per million mapped reads (FPKM) (Choulet et al. 2014a).

4.3.3 Preparation of VIGS-construct

For the transient gene silencing experiment, 110 bp fragment of wheat *cellulose synthase A* (TaCesA4) gene was selected from a region conserved among homoeologous genes but unique to all other genes in wheat. The fragment was scanned for specificity to avoid off target genes using BLAST against GenBank database with siRNA search and Scan tool а (http://bioinfo2.noble.org/RNAiScan.htm). The fragment was cloned into pUC57 vector (GenScript, NJ, USA) and the sequence was confirmed. The Plasmid was then digested using the restriction enzymes NotI and PacI (New England Biolabs, MA, USA) to generate NotI and PacI ends in the cloned DNA fragment. The cDNA fragment was subsequently ligated to the pSL038-1 vector of BSMV γ genome (Scofield et al. 2005). The plasmids were linearized using *Mlul* restriction enzyme for BSMV α , pySL038-1 whereas, *Spe1* enzyme for BSMV β .

4.3.4 In vitro transcription of VIGS plasmids and rub inoculation

Infectious BSMV RNA was prepared from the linearized plasmids by *in vitro* transcription using a T7 DNA-dependent RNA polymerase (Thermo Fisher Scientific Inc., CA, USA) according to manufacturer's instructions. Three *in vitro* transcripts, BSMV α , β and γ (BSMV: 00/BSMV: *TaCesA4*/ BSMV: *TaPDS*) in ratio 1:1:1 (2.5 μl each) were mixed with 22.5 μl of abrasive FES buffer to facilitate the viral entry (Bennypaul et al. 2012a). Ten plants were separately inoculated with each of test (BSMV: *TaCesA4*), positive control (BSMV: *TaPDS*) and negative control (BSMV: 00). FES buffer (1% sodium pyrophosphate, 1% bentonite, 1% celite in 0.1 M glycine, 0.06 M dipotassium phosphate) was used as abrasive buffer for rub inoculation. Plants were infected by rub-inoculating the flag leaf of each plant 2 to 3 at times at booting stage. Inoculated plants were grown at 22^oC days and 18^oC nights with 23-50% relative humidity and 16hrs light for 2-3 weeks, considered optimal for VIGS experiments.

4.3.5 RNA Isolation and cDNA synthesis

Stem tissues (internodes below the peduncle) of 21dpi (days post inoculation) plants, was collected and immediately placed in liquid nitrogen. For RNA extraction, samples were homogenised using a TissueLizer and then incubated in 1 ml of TRIzol reagent (Invitrogen, USA). Total RNA was extracted following the manufacturer's recommendations. Samples were treated with DNase1 (Promega Corp., WI, USA) and incubated in a 37^{0} C water bath for 30 minutes. cDNA was synthesized from 2 µg of RNA using an iScript cDNA synthesis kit (Bio-Rad, ON, Canada).

4.3.6 Real-time PCR

Primers for semi-qPCR and qRT-PCR were designed from the region unique to TaCesA4 and outside the region which was used for making VIGS construct using clone manager suite software 6.0 (Table 1). The semi-qPCR and qRT-PCR were performed with three biological and three technical replicates as described previously (Bregitzer et al. 2007; Singh et al. 2013). Amplification was performed using a reaction volume of 25 µl containing 2 µl of cDNA template

and SYBR Green II master mix (Stratagene, Cedar Creek, USA) following the manufacturer's recommendations. The cycling conditions are as follows: five minutes of activation at 95°C, followed by 30 cycles of 95°C for 20 sec, 52°C for 40 sec, and 72°C for 40 sec, followed by a dissociation curve cycle of 95°C for 1 min, 52°C for 40 sec, and 95°C for 40 sec using an Mx3005p PCR machine (Stratagene, Cedar Creek, USA). Gene silencing was expressed as a ratio of *TaCesA4* mRNA (normalized to *TaActin* mRNA) in BSMV: *TaCesA4* (test/silenced) inoculated plants to that inBSMV:00 (control/non-silenced) plants. qRT-PCR generated data was examined with Realtime PCR Miner (http://www.miner.ewindup.info/version2) and JMP software (version 3.2.2, SAS Institute Inc., Cary, NC, USA).

4.3.7 Estimation of cellulose content

Cellulose content was estimated as described by (Kaur et. al. unpublished) for three plants each from BSMV:00 and BSMV: *TaCesA4* inoculated plants. The first internode of the main tiller of each mature plant was taken and dried at 80°C. Dry sample (45-55 mg) was filled into a pre-weighted 2 ml Eppendorf tubes with a screw cap. 1.5 ml of a mixture of acetic acid: water: nitric acid (8:2:1) was added to each tube and vortexed (Appenzeller et al. 2004). All tubes were transferred to a steel rack and placed in a boiling water bath for four hours. Tubes were removed from the water bath and allowed to cool to room temperature. After the tubes reached room temperature they were placed in a swing-out rotor and centrifuged at 10,000 rpm for 10 minutes. The supernatant was aspirated off, washed with distilled water four times and finally washed with 90% ethanol. After each wash, the tubes were vortexed and centrifuged at 10,000 rpm for 10 minutes to aid in the formation of solid pellets. The caps were removed after the final wash and

the tubes were placed in the oven for drying at 80°C. The final weight of the tubes was used to calculate the percentage cellulose content on a dry matter basis.

% cellulose = Cellulose weight (final pellet dry weight)/Initial sample dry weight X100

4.3.8 Microscopic analysis of stem sections

Five stem samples (second internode from the top) per treatment (BSMV:00 and BSMV: TaCesA4) were taken at 21 dpi to analyse the anatomical features. The stem tissues were embedded in cryomolds containing Shandon CRYOMATRIX (Richard-Allan Scientific, Kalamazoo, USA) for cryosectioning. Stem cross sections (15 µm) were prepared using a cryotome (Leica, CM1850, Canada) machine at -10 °C. Cross-sections were stained with 5% toluidine blue (Sigma-Aldrich, Canada) for 5 min and washed with distilled water three times before mounting on glass slides (O'brien et al. 1964). Stained samples were observed under a fluorescence microscope (Nikon, Eclipse E800, USA).

4.3.9 Statistical analysis

Data from qPCR and cellulose content study was analyzed statistically using one-way analysis of variance (ANOVA) followed by student t-test (Cohen 1992).

4.4 Results

4.4.1 TaCesA4 gene structure and construct designing

Three paralogs of *TaCesA4* gene were obtained corresponding to the three homoeolog genomes of hexaploid wheat (A, B and D). The genomic copies of three *TaCesA4* homoeologs were variable

in size; *TaCesA4A* (4614bp), *TaCesA4B* (3181bp), *TaCesA4D* (3063bp), however their coding DNA sequences (CDS) shared 98% identity. Multiple sequence alignment of genomic copies with their corresponding CDS revealed the intron-exon boundaries and translation start and stop sites. There were nine exons in *TaCesA4A*, whereas other two homoeologs (*TaCesA4B* and *TaCesA4D*) were found to possess only four exons. Although the exon-intron boundaries were highly conserved among the homoeologs, *TaCesA4B* and *TaCesA4D* were missing their first five exons from 5' end (Fig 1).

To confirm the functional role of *TaCesA4* gene, we designed VIGS construct to target all three homoeologous copies (*TaCesA4A*, *TaCesA4B* and *TaCesA4D*). A VIGS construct was designed such that it possess at least 95% nucleotide similarity among three homoeologs. It also contained at least one stretch of 21 nucleotides showing 100% nucleotide identity towards the target gene and this criterion did not meet by any non-target gene. Such unique region was selected from the C-SR-II (Class-Specific Region-II), upstream of DXD motif of *TaCesA4* gene (Kaur et al. 2016). This region is highly variable among different *CesA* genes (Fig S1a), however, this is highly conserved among the homoeologous copies of *CesA4* genes in bread wheat (Fig S1b). The C-SR-II for *TaCesA4* gene is approximately 400bp long, which comprises exon 7 of *TaCesA4A* (2600 to 3000), and exon 2 of *TaCesA4B* (1241 to1646 bp) and *TaCesA4D* (1073 to 1478 bp) respectively. A 110 bp fragment from this region was cloned into the γ vector of BSMV genome.

4.4.2 Homoeolog specific expression of TaCesA4

In silico gene expression of three *TaCesA4* homoeologous genes was examined in five organs at three development stages (Choulet et al. 2014a). A bar graph displaying transcript abundance of the *TaCesA4* homoeologs from different wheat tissues (spike, root, leaf, grain and stem) at three

stages of wheat development is shown in Fig 2. Transcript abundance data revealed relatively higher expression of *TaCesA4A* as compared to that of *TaCesA4B* and *TaCesA4D* in all 5 tissue samples. These genes were highly expressed in the mature stem tissues collected soon after anthesis. However, significantly lower expression levels were observed in grain, leaf, root and spike tissues during different developmental stages (Fig 2).

4.4.3 Optimization of VIGS in Chinese spring (CS) wheat cultivar

The silencing of *PDS* (*phytoene desaturase*) gene triggered the photobleaching of leaves due to loss of chlorophyll pigments. This photobleaching effect was used as the visual marker to optimize the VIGS system in the *Chinese spring* variety of hexaploid wheat. An intense effect of *PDS* gene silencing (BSMV: *TaPDS*) was observed in wheat plants (booting stage) grown in the greenhouse under the temperature regimen 22°C day/18°C night. Ten plants were inoculated among which eight plants showed intense symptoms of photo-bleaching while others showed mild phenotypes at 21 dpi. There were no symptoms of photo-bleaching in the plants inoculated with BSMV:00. These plants were morphologically similar to the un-inoculated *Chinese spring* plants (Fig 3).

4.4.4 Silencing of *CesA4* gene in wheat

At 21 dpi, plants inoculated with the BSMV: TaCesA4 were phenotypically similar to the control (BSMV:00) plants as well as to the plants that were not inoculated. RNA was extracted from three plants each of BSMV:00 and BSMV: TaCesA4 inoculated plants to confirm the transient silencing via their relative transcript abundance. As per semi-qPCR (Fig 4) and qRT-PCR analysis, relative transcript expression of TaCesA4 normalised to reference gene TaActin in silenced plants (BSMV: TaCesA4) showed significant (P=0.0065) reduction (87.17%) compared to non-silenced

plants (BSMV:00), confirming the successful silencing of the target gene in the wheat stem (Fig 5).

4.4.5 Analysis of cellulose content in VIGS treated plants

Cellulose content was measured for five plants each of control (BSMV:00) and silenced plants (BSMV: *TaCesA4*). The percentage content of cellulose was significantly lower (29.27%) in the *TaCesA4* silenced plants as compared to the control plants at P=0.0041. An average percent cellulose content of control (BSMV:00)plants was 45.1% whereas the silenced plants (BSMV: *TaCesA4*) showed 31.9% cellulose in their stem tissue (Fig 6).

4.4.6 Histological analysis of stem tissues

To analyse the morphological characteristics of stem tissues of control (BSMV: 00) and silenced plants (BSMV: *TaCesA4*) plants, 15 μ m transverse sections of second internode of were stained with toluidine blue to visualize the tissue architecture. In the wheat stem, vascular bundles consisting of xylem (tracheids) and phloem (sieve tube elements) were clearly observed. A hollow cavity inside the stem called internodal cavity was lined by the parenchyma cells. Xylem cells were large and thick-walled as compared to small phloem cells (Fig 7). It was observed that the xylem and phloem cells of intermodal and nodal tissues of the stem were intact in the silenced plants. The organization and appearance of cells in silenced plants were also similar to that of control plants. This confirmed that the silencing of *TaCesA4* gene at booting stage has no effect on the shape and arrangement of cells.

4.5 Discussion

Plant cell wall polysaccharides are getting attention more recently due to their extensive use as dietary fibres, food additives, a raw material for biofuels, and fodder for livestock (Taylor-Teeples et al. 2015). In addition, to providing mechanical support and a barrier against pathogen invasion, secondary cell wall accounts for the bulk of renewable cellulosic biomass. Cellulose, hemicellulose and lignin are the major constituents of the secondary cell wall, among which cellulose is the main load bearing network. Cellulose in the secondary cell wall is synthesised by a complex containing three CESA subunits. The cells of dicots and most of the monocots possess Type I cell walls whereas commelinoid monocots possess type II cell walls (Carpita 1996). The major cereals such as barley, oat, wheat, maize, and rice, as well as the C4 grasses, comes under commelinoid monocots (Vogel 2008). More than 1500 genes have been reported for cell wall related function in Arabidopsis, rice and maize (http://cellwall.genomics.purdue.edu), and over 1000 unannotated genes are estimated for their probable role in cell wall biogenesis (Yong et al. 2005).

Presently, the major aim of cell wall research is to assign specific functions to this large collection of genes at different developmental stages of plant and to understand the regulatory networks responsible for cell wall biosynthesis. Although bioinformatics approaches have been remained quite supportive in providing tentative functions to these genes (Holland et al. 2000; Burton et al. 2004; Yin et al. 2009; Wang et al. 2010b; Liepman and Cavalier 2012; Liu et al. 2012; Schreiber et al. 2014b), only a few of them have been characterised for their specific functional role (Dhugga et al. 2004b; Burton et al. 2006a; Cocuron et al. 2007; Burton et al. 2011b; Taketa et al. 2012). To explore the function of genes in cell wall biosynthesis, a vast majority of mutant resources are available for Arabidopsis, however, there are very limited cell wall mutants in grass

species. For example *irregular xylem* (*irx*) mutants, *irx1* (*AtCesA8*), *irx3* (*AtCesA7*) and *irx5* (*AtCesA4*) of Arabidopsis unveiled a collapsed xylem phenotype indicating requirement of these genes for due to secondary cell wall formation (Hernández-Blanco et al. 2007)

In the case of bread wheat, the genes of this complex are named as *TaCesA4*, *TaCesA7*, and *TaCesA8* (Kaur et al. 2016). *In vitro* expression studies in wheat showed the highest transcript abundance of *TaCesA4* in mature stem tissues. These expression patterns are supported by their involvement in the formation of secondary cell wall formation, which is laid down in the mature cells (Taylor-Teeples et al. 2015). Wheat genome comprises three homoeologs of *CesA4* gene, which are structurally different; *TaCesA4A* homoeolog possesses 9 exons while other two homoeologs have 4 exons. Interestingly, an ortholog of wheat *CesA4* gene in rice (*OsCesA7*) also possesses 9 exons (Wang et al. 2010b; Kaur et al. 2016). Although first five exons from 5' end have been found to be missing in *TaCesA4B* and *D*, yet their expression has been observed in the mature stem tissues. Nonetheless, the expression of these two homoeologs was not as prominent as of *TaCesA4A* homoeolog. Despite an essential part of plant's basic structural unit, functional characterization of *TaCesA4* has not been reported in wheat.

In the current study, we have employed VIGS to understand the role of *TaCesA4* in the secondary cell wall of wheat. Transient gene knockdown through VIGS has been successfully described in wheat for assigning functions to different genes (Scofield et al. 2005; Tai et al. 2005; Bennypaul et al. 2012b). VIGS has also been employed in (*Nicotiana benthamiana*) for the functional analysis of *CesA* genes inserted in potato X virus vectors (Burton et al. 2000). A reduction in transcript levels and cellulose content was recorded for the infected plants. VIGS with BSMV has also been shown as an effective means of transient gene silencing in wheat and barley (Holzberg et al. 2002; Scofield et al. 2005; Bennypaul et al. 2012a). Silencing of *PDS* gene

encoding a phytoene desaturase through VIGS was used as a positive control which leads to visual photo-bleaching symptoms (Ruiz et al. 1998; Bennypaul et al. 2012a). The efficiency of transient knockdown through VIGS is dependent on cultivars and growth conditions (Bennypaul et al. 2012a). Inoculation of BSMV: *CesA4* at booting stage of *Chinese spring* plants successfully silenced *TaCesA4*. (Cakir and Tör 2010; Bennypaul et al. 2012a). We observed significant difference (87.17 %) in *TaCesA4* expression of control and silenced plants using qRT-PCR. Also, there was 29.27% decrease in the cellulose content of silenced plants as compared to control plants. The comparable impact of *CesA* gene silencing was observed on the gene expression and cellulose content in tobacco (*CesA1* and *CesA2*) (Burton et al. 2000), barley (*CesA6*) (Held et al. 2008), and flax (*CesA4* and *CesA8*) (Chantreau et al. 2015).

Reduction of cellulose content was also recorded for the brittle culm mutants of rice (Tanaka et al. 2003; Taylor et al. 2003; Kotake et al. 2011; Wang et al. 2012a). Cellulose synthesis was inhibited in *CesA4* (irx5), *CesA7* (irx3), and *CesA8* (irx1 mutants (Hernández-Blanco et al. 2007). Although phenotypes of mutant plants varied in different plant species, but the reduction in cellulose content was a common phenomenon in all these studies. The anatomical changes in the stem sections of flax plants were more pronounced for the VIGS of genes related to primary cell wall *CesAs* (*CesA1*, *3*, and *6*) and as compared to secondary cell wall *CesAs* (*CesA4* and *8*) (Chantreau et al. 2015). Similar to these observations, no obvious anatomical changes were observed in the stem sections of wheat plants after the silencing of *TaCesA4* gene. A possible explanation of this may be the growth stage of cells and tissues at the time of viral infection which largely determine anatomical features of cross sections. If the cells have grown to their full size and have adequate primary and secondary cell wall before the inoculation, they will appear normal after viral infection. Viral infection, in this case, can minimise or stop the further deposition of

cellulose due to the gene knockdown, but the decrease in the cellulose levels may not necessarily impact the cell shape and integrity.

Fig 4.1 Schematic depicting the structure of *TaCesA4* gene and its homoeologs. Red bar indicates the 110 bp region of the *TaCesA4* gene cloned into the BSMV γ vector.



Fig 4.S1a Multiple sequence alignment of the fragment used for designing VIGS construct with other secondary cell wall related genes (*TaCesA4, TaCesA7* and *TaCesA8*) along with their homoeologs representing the non-conserved region.

TaCesA7 3DL	-AAGAAAAAGGTTGAAAAAACTGAGAAGGAAATGCACAGAGA	1773
TaCesA7 3B	-AAGAAAAAGGTTGAAAAAACTGAGAAAGAAATGCACAGAGACTCCAGAAAGA	1770
TaCesA8 5BL	GAAGCGAAAGGGCGGCAAGGATGGGGTGCCGGAGCGC	2010
TaCesA8 5DL	GAAGCGAAAGGGCGGCAAGGATGGGCTGCCGGAAGGCCTGCCGGAAGGC	2013
TaCesA8 5AL	GAAGCGAAAGGGCGGCAAGGATGGG	2013
TaCesA4 1DL	GCACCGCAAGTCGAGCAAGGACAAGAAGGGCGGCGGCGGCGGCGACGATGAGGCGCGCGC	907
TaCesA4 1BL	GCACCGCAAGTCAAACAAGGAGAAGAAGGGCGGCGGCGGCGGCGACGACGACGGGCGCGCGC	904
TaCesA4 1AL	GCACCGCAAGTCGGACAAGGACAAGAAGGAGGCGGCGACGACGACGACGCGCGCCGCGGCTCCTCGGGTTCTACA	1885
VIGS_Construct	GCGACGACGACGACGCCCCCCCCCCCCCCCCCCC	39
TaCesA7 3DL	E	1791
TaCesA7 3B	e	1788
TaCesA8 5BL	eTeccegate	2020
TaCesA8 5DL	c	2023
TaCesA8 5AL	c	2023
TaCesA4 1DL	AGAACCGGGCAAGAAGGATAAGCTCGGCGGCGCCCGAAGAAGGGGTCGTACCGGAAGCAGCAGCGGGGTACGAGCTG	987
TaCesA4 1BL	AGAACCGGGGCAAGAAGGACAAGCTCGGCGGCGCCCCGAACAAGGGGTCGTACAGGAAGCAGCAGCGGGGTACGAGCTG	984
TaCesA4 1AL	AGAACCGGGGCAAGAAGGACAAGCTCGGCGGCGCGCGCGC	1965
VIGS_Construct	AGAACCGGGGCAAGAAGGACAAGCTCGGCGGCGGCGGGGGAAGAAGGGGTCGTACAGGAAGCAGCAGCGCGGG	110
TaCesA7 3DL	ATTTTCAATCTACGGGAAATCGACAACTACGACGAGTATGAGCGGTCCATGCTTATCTCCCAGATGAGCTTTGAGAAGTC	1871
TaCesA7 3B	ATTTTCAATCTACGGGAAATCGACAACTACGACGACGATGAGCGGTCCATGCTTATCTCCCAGATGAGCTTTGAGAAGTC	1868
TaCesA8 5BL	GAGGAATGACGGCGACAAGGAGCAGATGTCCCAGATGAACTTTGAGAAGCG	2075
TaCesA8 5DL	GAGGAATGACGCCGCCGACAAGGAGCAGATGTCCCAAATGAACTTCGAGAAGCG	2078
TaCesA8 5AL	GAGGAATGACGCCGCCACAAGGAGCAGATGTCCCAGATGAACTTCGAGAAGCG	2078
TaCesA4 1DL	GAGGAGATCGAGGAGGGGATAGAGGGGTACGACGAGGGGGGGCGCTCGTCGCCATGTCGCAGAAGAGCTTCCAGAAGCG	1067
TaCesA4 1BL	GAGGAGATCGAGGGGCATCGAGGGGTACGACGAGCTGGAGCGCTCGTCGCTCATGTCGCAGAAGAGCTTCCAGAAGCG	1064
TaCesA4 1AL	GAGGAGATCGAGGGGCATCGAGGGGTACGACGAGCTGGAGCGCTCCTCGCTCATGTCGCAGAAGAGCTTCCAGAAGAG	2045
VIGS_Construct		110

Fig 4.S1b Multiple sequence alignment of the fragment of *TaCesA4* gene used for designing VIGS construct with its homoeologs representing the conserved region.



Fig 4.2 In silico expression of TaCesA4 homoeologs in different wheat tissues, expressed as

reads per kilo base of transcripts per million mapped reads (FPKM) in hexaploid wheat. Blue color bar represent *TaCesA4A*, black and green bars denotes *TaCesA4B* and *TaCesA4D* respectively.



Fig 4.3 Silencing of the *phytoene desaturase (PDS)* gene. Leaf phenotypes of wheat plants inoculated with BSMV:00 and BSMV: *TaPDS* at 21 dpi.



BSMV: 00 BSMV: TaPDS

Fig 4.4 Semi-qPCR based expression of TaCesA4 normalised to reference gene TaActin in silenced

plants (BSMV: TaCesA4) and non-silenced plants (BSMV:00); Where L- marker, -ve- negative control.



Fig 4.5 Quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) analyses to confirm the gene knockdown as the relative transcript expression of *TaCesA4* normalized to *TaActin* mRNA in BSMV: *TaCesA4* inoculated plants as compared to control (BSMV:00) plants at 21 dpi.



Fig 4.6 Cellulose content (% w/w) in *TaCesA4* silenced (BSMV: *TaCesA4*) plants as compared to control (BSMV:00) plants.



Fig 4.7 Transverse sections of stem tissues of control (BSMV: 00) and silenced (BSMV: *TaCesA4*) wheat plants at 20X and 4X magnification; where mx is meta xylem, px is protoxylem, ph is phloem.



Table 4.1 Primers used for semi-qPCR, qRT-PCR and confirmation of VIGS construct.

		Primers			
Experiment	Name	Forward	Reverse		
VIGS fragment amplification	Gamma	TGATGATTCTTCTTCCGTTGC	TGGTTTCCAATTCAGGCATCG		
and sequencing					
VIGS gene expression	TaActin	TGTGCTTGATTCTGGTGATGGTGTG	CGATTTCCCGCTCAGCAGTTGT		
VIGS gene expression	TaCesA4	CCGAAGAAGGGGGTCGTACAG	CTCTTCTGCGACATGAGCGA		

CONNECTING STATEMENT FOR CHAPTER V

Chapter V, entitled "Genome-Wide Association study (GWAS) revealed novel genes linked to natural variability of cellulose content in Bread Wheat (*Triticum aestivum*, L.)" authored by Simerjeet Kaur, Xu Zhang, Amita Mohan, Prashant Vikram, Sukhwinder Singh, Kanwarpal S. Dhugga, Zhiwu Zhang, Kulvinder Gill and Jaswinder Singh has been submitted to "*Frontiers in Plant Science*".

In chapter III and IV, we have explored the genes that are the major players for cellulose biosynthesis in wheat. These studies led to the identification of 22 CesA genes based on the comparative genomics approach. A gene (TaCesA4) expressing in the mature stems was validated for its contribution towards cellulose synthesis through VIGS. But our current knowledge is limited about the genetic associations of the existing natural variation in cellulose content. In chapter IV, we performed a comprehensive study about such genetic connections. We have evaluated 284 diverse wheat lines to estimate natural variation of cellulose content in the straw. This phenotypic variability was further linked to the SNP genotyping data generated by GBS (genotyping by sequencing). Genome-wide Association Studies (GWAS) led us to identify novel genetic association (β-tubulin and UDP-glycosyl transferase (UGT) family) linked to cellulose content in wheat straw. β -tubulin genes were previously reported to synthesise the microtubules that are associated with the delivery of CESA complexes to the plasma membrane (Gutierrez et al. 2009). The UGT family genes are known for the transfer of UDP-glucose to the catalytic sites for the synthesis of cellulose (Lairson et al. 2008). These novel associations will be valuable to devise marker-assisted/genomic selection strategies to monitor cellulose content in wheat breeding populations.

Chapter V. Genome- wide association study reveals novel genes linked to natural variation of cellulose content in bread wheat (*Triticum aestivum*, L.)

Simerjeet Kaur¹, Xu Zhang², Amita Mohan², Prashant Vikram³, Sukhwinder Singh³, Kanwarpal S. Dhugga³, Zhiwu Zhang², Kulvinder Gill² and Jaswinder Singh¹*

¹Department of Plant Science, McGill University, Sainte Anne de Bellevue, QC, Canada
 ²Department of Crop and Soil Science, Washington State University, Pullman, WA, USA
 ³International Maize and Wheat Improvement Center (CIMMYT), El Batán, Texcoco, Estado de México

*Corresponding Author

5.1 Abstract

Plant cell wall provides dynamic structure and shape to the cells. Cell wall formation is a complex, coordinated and developmentally regulated process. Synthesis and remodelling of various cell wall components play a vital role in plant development and architecture. Cellulose is the most abundant biopolymer on earth and the most dominant constituent of plant cell walls. Because of its paracrystalline structure, cellulose is the main determinant of mechanical strength of plant tissues. As the most abundant polysaccharide on earth, it has been the main focus of cellulosic biofuel industry. It is important, thus, to explore the underlying mechanism of cellulose biosynthesis. This report presents results on the analysis of the stem cellulose content of 288 diverse wheat accessions and genome-wide association study (GWAS). The germplasm showed 6.56% coefficient of variation (CV) in cellulose content among diverse wheat accessions. Genotypic data comprising 21,073 SNPs was used to establish genome-wide marker-trait associations. The analysis led to the

identification of nine SNPs, which associated significantly (p<1E-05) with cellulose concentration. Four strongly associated (p<8.17E-05) SNP markers were linked to wheat unigenes. These unigenes were annotated using BLASTn search against various plant databases. Genes including β -tubulin, Auxin-induced protein 5NG4 and a putative transmembrane protein of unknown function were found to be associated with cellulose content. Associated genes may be directly or indirectly involved in the synthesis of cellulose in wheat but further investigations are necessary to establish their respective involvements. GWAS results from this study have the potential for genetic manipulation of bread wheat and other small grain cereals to enhance culm strength.

5.2 Introduction

The increasing world population demands a sustainable increase in the production of food, feed and fuel crops (Scholey et al. 2016). Bread wheat (*Triticum aestivum*) occupies more agricultural area than any other food crop worldwide (http://www.wheatinitiative.org/). In addition to grain production, the annual worldwide production of wheat straw is around 350 million tons, which is used as cattle fodder in developing countries and is a potential feedstock for cellulosic ethanol production (Singhania et al. 2014). The use of grain for food and feed and straw residue for fuel could make wheat a dual purpose crop. Wheat straw, which is comprised of cellulose (~40%), hemicelluloses (~35%) and lignin (~25%), is one of the most abundant lignocellulosic raw materials in the world (Ruiz et al. 2013). Cellulose, a paracrystalline polysaccharide, is the main determinant of mechanical strength, which has implications in crop lodging, biotic and abiotic stresses. Cellulose amount in a unit length of the stem explains most of the variation for mechanical strength (Appenzeller et al. 2004). The proportion of cellulose in cell wall also affects the total sugar release during the process of enzymatic hydrolysis (FAN et al. 2012; Lindedam et al. 2012). An understanding of the natural variability of cellulose in plants and its association with chromosomal regions could provide markers for enhancing grain and biomass yield (Ciesielski et al. 2014).

Cellulose consists of a linear chain of β (1 \rightarrow 4) linked glucan (polyglucose) known to be synthesised by the members of superfamily *Glycosyltransferase 2 (GT2)* called *Cellulose synthase* A (CesA) (Fujii et al. 2010; Kumar et al. 2016). Twenty-two CesA genes have been reported in hexaploid wheat (Kaur et al. 2016). In addition to the CesA genes, the Glycosylhydrolase 9 (GH9) family genes are known to have an impact on the synthesis of cellulose in plants (Kotake et al. 2011). Based on the mutant analysis in Arabidopsis, a member of GH9 family called KORRIGAN1 (KOR1) has been reported to be involved in cellulose synthesis, cell expansion and intracellular trafficking of cellulose synthase complex (CSCs) (Szyjanowicz et al. 2004; Lei et al. 2014; Vain et al. 2014). Investigation of *brittle culm 1* mutants in rice and *brittle stalk 2* mutant in maize revealed the association of COBRA-like proteins with the cellulose microfibrils (Ching et al. 2006). Involvement of Sucrose synthase (SuSy) in channelizing substrate to cellulose synthase has also been reported (Fujii et al. 2010). Similarly several other proteins affect cellulose synthesis, including chitinase-like 1 (CSI1) (Sánchez-Rodríguez et al. 2012), companion of cellulose synthase (CC) (Endler et al. 2015), tracheary element differentiation-related (TED) 6 and 7 (Rejab et al. 2015).

The involvement of several genes for cellulose synthesis highlights the complexity of the process, which needs further nvestigation to better comprehend the underlying mechanism (Kotake et al. 2011). Also, the variation for the proportion of cellulose in cell wall among wheat varieties is not been well understood. This study was planned to identify the genomic regions affecting the variability of cellulose content among diverse spring wheat genotypes through GWAS.

Genes associated with cell wall have been previously explored through GWAS in miscanthus (Slavov et al. 2014), Populus (Porth et al. 2013) maize (Li et al. 2016) and barley (Houston et al. 2015). In the case of barley genes of *Glycosyltransferase 2* and *Glycosylhydrolase* families were found to be associated with culm cellulose variation. However, none of the genes found in maize through GWAS of stalk cellulose content was specifically involved in the cellulose biosynthesis pathway. In the present study, the stem internodes of 288 spring wheat varieties were analysed for variation in cellulose content. Utilizing the 21,073 SNPs generated by DArT-seq GBS and cellulosic content, GWAS was performed by the fixed and random model circulating probability unification (FarmCPU) method (Liu et al. 2016). Genes, which were not reported previously for their role in cellulose formation, were identified as associated with the culm cellulose content. Gene-trait associations identified in this study might be useful in altering the lignocellulose composition of wheat and other grasses at a genetic level.

5.2.1 Hypothesis Variability of culm cellulose content in diverse wheat genotypes is linked to specific genomic regions.

5.2.2 Objective I. Analysis of cellulose content for diverse wheat lines

5.2.3 Objective II. GWA Study to identify novel genes linked to cellulose content in wheat

5.3 Materials and methods

5.3.1 Plant material

A worldwide collection of 288 diverse spring growth-habit wheat germplasm was used for the phenotypic and genotypic analysis. The collection included cultivars from different regions of United states, the International Maize and Wheat Improvement Centre (CIMMYT), Mexico, and historical lines dating back to 1871 (Mohan et al. 2013). The wide span of our collection was

intended to capture the maximum variation possible while maintaining a manageable population size. This worldwide collection also represents the various market classes of wheat based on color, hardiness and shape of the kernel: i.e. soft white spring (SWS), soft red spring (SRS), hard red spring (HRS), hard white spring (HWS), and club wheat cultivars (Mohan et al. 2013). The plants were grown in the greenhouse of the Plant Growth Facilities, Washington State University, Pullman at 22°C/18°C day/night temperature with 16 hours of light in 2014-15. Seeds were planted with randomised design to accommodate the effect of light.

5.3.2 Phenotypic analysis

The analysis on percentage cellulose was performed for 288 diverse spring wheat genotypes, with three replicates per genotype. The first internode (from the base) of the main tiller of each mature plant was taken and dried at 80°C. Measured amount of dried sample (45-55 mg) was put into a pre-weighted 2 ml Eppendorf tubes with a screw cap. A mixture of acetic acid: water: nitric acid (8:2:1) was added to each tube (1.5ml) and vortexed (Appenzeller et al. 2004). All tubes were transferred to a steel rack and placed in a boiling water bath for four hours. After four hours, tubes were removed from the water bath and allowed to cool at room temperature. After the tubes reached room temperature they were placed in a swing-out rotor and centrifuged at 10,000 rpm for 10 minutes. The supernatant was aspirated off, washed with distilled water four times and finally washed with 90% ethanol. After each wash, the tubes were vortexed and centrifuged at 10,000 rpm for 10 minutes to aid in the formation of solid pellets. The caps were removed after the final wash and the tubes were placed in the oven for drying at 80°C. The final weight of the tubes was used to calculate the percent cellulose content on a dry matter basis using the formula: % cellulose = Cellulose weight (final pellet dry weight)/Initial sample dry weight x100

5.3.3 Population structure and GWAS analysis

The population structure was represented by the first Principal Components (PCs) calculated from all the SNPs. The three PCs were fitted as covariates in both the fixed effect model and the mixed linear model to eliminate the non-genetic effect confounded with population structure. The two models were iterated until converge on the estimated QTNs (Lipka et al. 2012; Ahmad et al. 2015; Tang et al. 2016).

A total of 21,073 SNP markers were obtained by analysing genomic DNA with the Genotyping-By-Sequencing (GBS) based approach (Mohan et al. unpublished). In brief, genotyping was carried out at DArT Pyt Ltd in Canberra-Australia, using a combination of HiSeq 2000 (Illumina) next-generation sequencing with DArT-seq GBS technology (called DArTseq TM). This method follows two-step complexity reductions by using two enzymes, PstI/HpaII and PstI/HhaI, along-with TaqI restriction enzyme to eliminate subsets of PstI -HpaII and PstI-HhaI fragments, respectively. The polled barcoaded samples were run in a single lane on an Illumina Hiseq 2000 instrument for sequencing. A proprietary analytical pipeline developed by DArT Pyt Ltd was used to obtain the DArT score and SNP tables (http://www.diversityarrays.com/). GWAS was conducted using a recently developed method, FarmCPU (Fixed and Random Model Circulating Probability Unification) (Liu et al. 2016) in R version 2.15.3. The model controls both non-genetic effects that confound with population structure, and genetic effects that confound with genetic loci having no genetic linkage with the test SNPs.

The confounded genetic effects controlled by Quantitative Traits Nucleotides (QTNs) were estimated using an algorithm named SUPER (Settlement of MLM under Progressively Exclusive Relationship). The whole genome was divided into bins. Each bin was represented by the most significant SNP within each bin. The bin size and significant threshold were optimized by using the restricted maximum likelihood (REML) in a mixed linear model with kinship among individual lines calculated from the candidate bins. The set of bins with the optimum REML were used as the estimated QTNs. The estimated QTNs were directly fitted as covariates for testing SNPs in a fixed effect model to control the genetic effects confounded to the test SNPs. A Manhattan plot was generated using the $-\log_{10}(p)$ values for each SNP with 1% Bonferroni test threshold (Team 2014). The significance of the genome-wide association between SNP marker and cellulose content was tested at FDR p <0.001. 5.3.4 Gene annotation.

The SNPs containing sequences were mapped against wheat unigenes downloaded from the NCBI database. The significant SNPs with associated unigese were annoted using BLASTn with the International Wheat Genome Sequencing Consortium (IWGSC) (Mayer et al. 2014) reference Sequence v1.0 (<u>https://www.wheatgenome.org</u>) posted on May 30, 2017. The functions to associated unigenes were also searched in orthologs found in another species.

5.4 Results

5.4.1 Cellulose content

A set of 288 diverse wheat lines was analysed for native variation in cellulose content (Appendix 5.1). Significant differences in percent cellulose content of wheat lines on a dry matter basis were identified. The coefficient of variation for cellulose content is 6.56% among the wheat lines. The cellulose content of germplasm ranged from 0.32 to 0.52 mg cellulose/mg of dry weight with an average of 0.45 mg cellulose/mg of dry weight). The wheat population showed a trend of a normal

distribution with respect to the cellulose variation and the density plot for the cellulose analysis is shown (Fig 1).

5.4.2 Principal component analysis and marker-trait associations

Principal component analysis (PCA) was performed to investigate the population structure. The first two PCs explained 8.13 and 4.90% variation in the population. The collection showed two distinct clusters, a minor and a major one. To simplify the population structure, the minor cluster containing 20 genotypes was removed from the final analysis and first PC was used as covariate while conducting GWAS (Fig 2).

A total of 21073 SNP markers with minor allele frequency (MAF) above 5% and the cellulose content data from 268 lines were used for GWAS analysis (Fig 3). Using the GWAS analysis, we found nine significant marker-trait associations with p values of less than 1E-05. The most significant correlation in our analysis corresponded to wheat chromosome 5AL with p-value 1.86E-07. The second most significant SNP being on chromosome 1AL with a p-value of 2.24E-07. In addition, we found significant SNPs corresponding to chromosome 1AL, 6BS, 1DL, 2DS, 4DL, 5BL, and 3B with p values <1E-05 respectively (Table 1). The quantile–quantile (QQ) plot drawn for calculated p-values was used to check spurious associations. The deviation of relatively a few markers from null expectations in the QQ plot is evidence for significant associations to be present (Fig 5).

5.4.3 Gene identification

Significant SNP markers resulting from GWAS were mapped to the wheat unigene database, their corresponding unigene identified. These unigenes were used to provide the most likely annotation through the NCBI BLAST and EnsemblPlant database. The searches resulted in the identification

of genes corresponding to these hits. The first SNP marker was found to be on the gene TRIAE_CS42_5AL_TGACv1_376159_AA1232950 and the second SNP marker corresponded to a genomic region containing unigene gnl|UG|Ta#S52545076. The third and fourth significant SNPs corresponded to the genes TRIAE CS42 2DS TGACv1 179544 AA0607850 and TRIAE_CS42_3B_TGACv1_224721_AA0800650.1 respectively. The gene TRIAE_CS42_5AL_TGACv1_376159_AA1232950 is uncharacterized in wheat as well as other plant species. The unigene gnl|UG|Ta#S52545076 showed 60% amino acid identity and 85% coverage with a gene in the *Tubulin* superfamily, *Tubulin* β -1 chain of *Triticum urartu*. TRIAE CS42 2DS TGACv1 179544 AA0607850 showed 82% identity and 97% coverage with the Auxin-induced protein Aegilops 5NG4of tauschii, whereas TRIAE_CS42_3B_TGACv1_224721_AA0800650.1 was annotated based on 51% amino acid identity and 97% coverage with a putative transmembrane protein of *Medicago truncatula* (Table 1).

5.5 Discussion

From a larger set of 288 diverse bread wheat lines, we used 268 well-structured accessions to describe the genetic association of cellulose content variation. The most appropriate model was selected to obtain a higher level of confidence in our association results. We employed GBS for genome-wide SNP genotyping and conducted a comprehensive phenotypic analysis for multiple replications of 288 diverse wheat lines, to capture the variability in cellulose content. Phenotypic data was then combined with genotypic screening to implement Genome Wide Association Studies (GWAS) using Fixed and Random Model Circulating Probability Unification (FarmCPU); a new and more efficient method has been recently published that accounts for fixed and random effects to control false positives (Liu et al. 2016). The fixed effects include testing

SNPs and population structure represented by the first three principal components calculated from all the SNPs. The random effects were the genetic effect of individuals lines with variance and covariance structure defined by the kinship calculated from the estimated Quantitative Traits nucleotides (QTNs). Most of the GWA mapping studies in wheat has been employed for the identification of genes or QTLs related to agronomic performance (Lopes et al. 2015; Jaiswal et al. 2016), grain yield (Sukumaran et al. 2015), disease resistance (Kollers et al. 2013; Gurung et al. 2014). To our knowledge, this is the first GWAS analysis related to the natural variation of cellulose content in wheat. Cellulose is a key component of plant cell walls and involved in mechanical strength in plants (Appenzeller et al. 2004). It is well documented that the CesA genes are involved in the synthesis of cellulose and recently a total of 22 CesA genes have been reported in wheat which differentially expresses in primary and secondary cell wall (Kaur et al. 2016). We have identified two significant associations for cellulose content in spring wheat. Although there were approximately 9 SNP markers that were associated $[-\log_{10}(p)=7 \text{ to } -\log_{10}(p)=5]$ with cellulose content (Table S1), we were able to map only four of these to the wheat unigene database. Greater marker density and population size used here provides higher confidence about these hits (Wang et al. 2012b).

The corresponding genomic regions for the SNP markers showing significant association with stem cellulose content were explored and the gene annotations were derived from the EnsemblPlant database. The fact that these genes showed significant association with cellulose content suggests that they may play a role in controlling the natural variation of cellulose in wheat lines. The involvement of many genes other than *CesAs* in controlling cellulose synthesis provides the evidence for the complexity of the process (Kotake et al. 2011). But there are still some missing links to completely understand the complex mechanism of cellulose synthesis.
Only a few studies have been performed to explore the additional genes involved in the cellulose biosynthesis pathway (Porth et al. 2013; Slavov et al. 2014; Houston et al. 2015; Li et al. 2016). Recently a GWA study in barley, a species syntenic to wheat, showed the involvement of genes co-expressing with *CesA* genes in culm cellulose content variation. Cellulose content was analysed for 288 two-rowed and 288 six-rowed spring type barley accessions genotyped with 3072 SNPs. GWAS results showed the significant hits involving genes mainly from *Glycosyltransferase* and *Glycosylhydrolases* (Houston et al. 2015). Similar to barley GWAS hits, our results also showed the involvement of *GT* gene family. However, we have encountered some unique hits that were probably missing in barley study because of the lower number of SNP markers used for the analysis. The present study has shown statistical evidence for marker-trait associations, which will add to our present knowledge of cell wall genetic architecture.

Our results pointed to the involvement of β -tubulin in the regulation of cellulose content. β -tubulins are proteins that form heterodimers with α -tubulins to form microtubules. These microtubules showed a closed association with cellulose microfibril deposition and formation of the secondary cell wall (Rao et al. 2016). There are many studies that have shown the functional association of cortical microtubules with cellulose synthase complexes, most of which were studied in Arabidopsis (Paredez et al. 2006; Chan et al. 2007; Wightman and Turner 2008; Crowell et al. 2009; Gutierrez et al. 2009; Chan et al. 2010).

Another important hit in our analysis is the Auxin-induced protein *5NG4*. This gene is a member of the plant drug/metabolite exporter (P-DME) (TC 2.A.7.4) family, also called WALLS ARE THIN1 (WAT1)-related proteins. Mutant studies in Arabidopsis have revealed its involvement in secondary cell wall formation in fibres. Comparative transcriptomics and metabolomics demonstrate the synchronised downregulation of the secondary cell wall *CesAs*

(CesA8, CesA7 and CesA4) and auxin metabolism genes (auxin-responsive genes and auxin influx transporter genes) in *wat1* mutants (Ranocha et al. 2010). The RNA-seq expression profiling of Chinese fir (*Cunninghamia lanceolate*) has also revealed the higher expression of PIN-like auxin efflux carrier and auxin-induced protein 5NG4 genes in relation to both cell division and cell expansion (Qiu et al. 2013). Our results also indicated the possible involvement of an uncharacterized gene (TRIAE_CS42_5AL_TGACv1_376159_AA1232950) in cellulose biosynthesis. This gene can be further explored for its specific role in the cell wall synthesis. The last hit in our analysis is a putative transmembrane protein of unknown function. Functional validation of these novel identified associations will further strengthen our understanding of their biological role in cellulose content variation found in the wheat stems. Though we have not yet drawn conclusions regarding the differences in cellulose content between different varieties of same species, our results indicate that additional genes are likely involved in the mechanisms responsible for the cellulose content variation in diverse wheat varieties.

5.6 Conclusion

Cellulose content in the culms of bread wheat varies from 0.32 to 0.52 mg /mg dry weight) in bread a diverse set of 288 genotypes. Genome-wide association analysis of 21073 SNPs with cellulose content variation helped identify 4 *de novo* genetic associations, which have the potential as molecular markers for manipulating cellulose content in wheat with the goal of improving culm strength.

Fig 5.1 Density plot showing the percentage cellulose content among 288 diverse spring wheat accessions.



Fig 5.2 Principal component analysis of 288 diverse genotypes used for GWAS.



Fig 5.3 Minor allele frequency (MAF) patterns determined relative to allele calls for wheat genotypes based on 21073 SNPs.



cellulose

Fig 5.4 Manhattan plot of genome-wide association study (GWAS) on stem cellulose content (mg cellulose/mg dry weight) by using the FarmCPU. The -log10(*p*-values) from GWAS are plotted against the position on each of the 42 bread wheat chromosomes. U represents unassigned chromosome scaffolds. Two loci on chromosomes 1A and 5A were identified above the Bonferroni threshold correcting genome-wide multiple tests at type I error of 0.001 (green line).



Fig 5.5 Quantile-quantile (QQ) plot showing the deviation from null hypothesis for associated SNP makers.



SNP ID	Allele	CHR	Scaffold:Position	P value	MAF	Unigene	Candidate	Gene ID (Ensembl)
							annotation	
1096787 F 040	C>T	5AL	376159:25309	1.86E-07	0.323	gnl UG Ta#S13258805	Uncharacterized gene	TRIAE_CS42_5AL_TGACv1_ 376159_AA1232950S
1018641 F 062	T>C	1AL	138:45403	2.24E-07	0.285	N/A		
100315676 F 050	T>C	1AL	1074:43532	2.05E-06	0.402	gnl UG Ta#S52545076	Tubulin β-1 chain	TRIUR3_05395
1080815 F 044	T>C	6BS	514572:36113	3.18E-05	0.202	N/A		
3026141 F 05	A>C	1DL	63549:20036	3.72E-05	0.394	N/A		
1018617 F 035	C>T	2DS	179544:14866	4.02E-05	0.489	gnl UG Ta#S65598833	Auxin-induced protein 5NG4	TRIAE_CS42_2DS_TGACv1_ 179544_AA0607850
1245047 F 039	C>T	4DL	344580:40916	4.12E-05	0.070	N/A		
1069330 F 06	T>A	5BL	406565:38744	5.21E-05	0.189	N/A		
2249069 F 014	G>A	3B	224721:15888	8.17E-05	0.177	gnl UG Ta#S61725485	Transmembrane protein, putative	TRIAE_CS42_3B_TGACv1_2 24721_AA0800650.1

Table 5.1 Regions of wheat genome showing significant associations with stem cellulose content variation based on GWAS.

Table 5.S1. Sequences of	SNPs significantly	associated with stem	cellulose content variation.
--------------------------	--------------------	----------------------	------------------------------

SNP ID	Allele
1096787 F 040	CTTGCCACGACCGATTATCACCAACGACTGACAAGCCACGCCCCATTTTGGGCTGCCCTGCGCG
1018641 F 062	TCCAGCAACAAATGACTTGGTTGTATAGTCCGTAGGCACATCGGGAGTTGTTTCTTGTTGTAGT
100315676 F 050	CTCATGTCGTCTAGCACGTCGAACACCTCGGAGATGAGCCCCGTGTGGTCGGCGCTCGTCAGCT
1080815 F 044	CAGTTACACTAGAGAGTTGGATAAAAGCTTCTGCTATTTTCAAAGAAAATCGGTCACTTTGGAG
3026141 F 05	ACCGTGCGTGCCCGTGCACGTGTCCGTGCCGCCCGAGATCGGAAGAGCGGTTCAGCAGGAATGC
1018617 F 035	GATGCTCATGGTGATGGCTCCCCCAGGCACAGAAGGGTCCCCACTATCTTGGCTCTTGTGTAC
1245047 F 039	GCAAGCTCTTGGGTTTCTTGGTTTCTAACAGAGGCATTGAAGCTAACCCGAGATCGGAAGAGCG
1069330 F 06	TTTTTCCAAAATTATGGTATTTTCTCTGCTTATAAAAAAGAACCCCCGACCTCTTTTTTAAAAC
2249069 F 014	CGTCCTCATGTGCGCGCTGCTCTACTTCCTCGACACCTCCGCGGACTACGCCAAGGGGATACAG

CONNECTING STATEMENT FOR CHAPTER VI

Chapter VI, entitled "Genome-wide analysis of the *Cellulose synthase-like* (*Csl*) gene family in bread wheat'' authored by Simerjeet Kaur, Kanwarpal Dhugga, and Jaswinder Singh has been submitted to "*BMC Plant Biology*".

In chapters V, GWAS was performed for the identification novel genes controlling the cellulose content variation in diverse wheat genotypes. In addition to cellulose, hemicellulose is also an important component plant cell walls comprising roughly one-third of cell wall biomass. This is composed of several heteropolymers that interact with cellulose microfibrils through hydrogen bonds. Despite their major contribution towards the biomass and infrastructure cell walls, the synthesis of hemicelluloses is poorly understood in wheat. In this study, we have explored the *Cellulose synthase-like (Csl)* members, which have been known for the regulation/synthesis of hemicelluloses such as heteromannan, xyloglucan, heteroxylans, and mixed-linkage glucan. We have identified a total of 108 *Csl* genes using the gene family specific Pfam conserved domains. The classification of these genes based on phylogenetic analysis and tissue-specific expression has been discussed in chapter VI.

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

Simerjeet Kaur1, Kanwarpal S. Dhugga2, Jaswinder Singh*1

¹Department of Plant Science, McGill University, Sainte Anne de Bellevue, QC, Canada

² International Maize and Wheat Improvement Center (CIMMYT), El Batán, Texcoco, Estado de México

*Corresponding Author

6.1 Abstract

Hemicelluloses are a diverse group of complex non-cellulosic polysaccharides, which constitute approximately one-third of the plant cell wall. Despite their extensive 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. In this study, we have isolated the gene sequences for the *cellulose synthase-like* (*Csl*) family from wheat. A total of 108 genes (hereafter referred to as *TaCsl*) including two to three homoeologous copies for each were identified and named as *TaCslXY_ZA*, *TaCslXY_ZB*, or *TaCslXY_ZD*, where X denotes the subfamily, Y as the gene number and Z stands for chromosome number on the respective genomes of bread wheat. One-fourth of these genes had 2 to 3 splice variants, resulting in a total of 137 putative proteins. Close to 45% of *TaCsl* genes were found to be located on chromosomes 2 and 3. To gain insight into the potential functional role of this gene family, we performed *in silico* expression analysis in different tissues using a publically available dataset. Although most of the genes were expressed ubiquitously, 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. This study provides new insights into the structure and function of the *Csl* gene family in hexaploid wheat.

6.2 Introduction

Non-cellulosic plant cell wall matrix polysaccharides generally referred to as hemicellulose, exhibit diverse linear or branched structures (Pauly and Keegstra 2008). These mainly encompass 1-4- β -glucan, 1,3;1,4- β -glucan, galactan, or glucomannan in grasses (Sorek et al. 2014). In addition, glucuronoarabinoxylan is a major grass well constituent. Because of the presence of heterogeneous substituents or other linkages in their polymer backbone, the structure of hemicellulose is non-crystalline and can be comparatively readily hydrolysed in comparison to cellulose. These polysaccharides can interact with cellulose chains through hydrogen bonds (Pauly et al. 2013).

Hemicellulosic polysaccharides 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*) (Richmond and Somerville 2000). These genes encoding these enzymes share sequence similarity with the *cellulose synthase A* (*CesA*) gene family known to form cellulose throughout the plant kingdom (Kaur et al. 2016). A variable number of *Csl* genes ranging from 30 to 50 have been identified from different plant species and are classified into nine subfamilies (*CslA–CslH* and *CslJ*) (Hazen et al. 2002). Cereals generally lack *CslB* and *CslG* families. Among the remaining families, *CslA, CslC*, and *CslD* are conserved in all land plants, whereas *CslF, CslH*, and *CslJ* are restricted to grasses (Farrokhi et al. 2006; Burton et al. 2011b). The subfamilies *CslB* and *CslG* were previously reported to be present only in dicots (Dhugga 2012). However, a recent study revealed the presence of *CslB* subfamily in monocots as well (Yin

et al. 2014). Diverse groups of *Csl* gene family have been reported to be involved in the biosynthesis of different cell wall polysaccharides. For example, subfamily *CslA* has been implicated in the biosynthesis of the β -1,4-mannan backbone of galactomannan and glucomannan (Dhugga et al. 2004a; Liepman et al. 2005). Similarly, *CslF* and *CslH* groups were reported to mediate the biosynthesis of 1-3;1-4- β -glucan in grasses (Burton et al. 2006b; Doblin et al. 2009) whereas *CslC* genes have been reported to be involved in the synthesis of the 1-4- β -glucan backbone of a xyloglucan and some other polysaccharides (Cocuron et al. 2007).

Wheat is a major cereal crop, which is grown on largest arable land in the world, is second only to maize in grain production, and feeds approximately 40% of the world population (Gupta et al. 2008). It has a large genome size (~17 Gb), of which ~80-90% is repetitive (Mayer et al. 2014). Because of its large genome size and hexaploid nature, *Csl* genes have not been well defined in wheat. Furthermore, the full genome sequence of bread wheat was not available until recently (Consortium 2014), which posed a challenge in exploring this complex gene family. Bread wheat possesses three homoeologous sets of seven chromosomes each distributed in three subgenomes (A, B and D). In general, homeologous copies of most of the genes are located on chromosomes of each genome. Moreover, *Csl* genes share a large sequence similarity with each other or within the subgroup, which makes it a challenging task to identify and characterise these genes in hexaploid wheat.

In the present study, we have explored the recently available resources to retrieve wheat genomic sequence. Comprehensive and large-scale data mining was performed using the Pfam domain models for the identification of *Csl* gene family in wheat. *TaCslD* has been studied in more detail for its gene structure and intron evolution, because of its evolutionary and structural

proximity to *CesA* genes and its probable role in cellulose or mannan synthesis (Verhertbruggen et al. 2011; Wang et al. 2011).

6.2.1 Hypothesis Orthologs of higher plants *Cellulose synthase-like* (*Csl*) genes are present one the A, B, D, homeolog genomes of bread wheat

6.2.2 Objective I. Identification of homeologus copies of *Csl* genes in wheat

6.2.3 Objective II. Phylogenetic and expression analysis of Csl genes

6.3 Materials and methods

6.3.1 Data sources and sequence retrieval

Wheat genome data was downloaded from the Ensembl Plants <u>FTP server</u> (ftp://ftp.ensemblgenomes.org/pub/current/plants/fasta/triticum_aestivum/), generated by the International Wheat Genome Sequencing Consortium (IWGSC) [29] and converted into a local BLAST database using the UNIX pipeline. BLAST analyses (BLASTN as well as BLASTP) were performed using the stand-alone command line version of NCBI (National Center for Biotechnology Information) blast 2.2.28+ (<u>ftp://ftp.ncbi.nih.gov/blast/executables/LATEST/</u>), 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 (Finn et al. 2016). 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) (Kim et al. 2007b). Previously known *Csl* sequences from Arabidopsis, rice, and maize were downloaded from the Cell Wall Navigator database. For Brachypodium, sequences were retrieved from phytomine.

6.3.2 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 hits below the cut-off value, CD-hit program (Li and Godzik 2006) was used to get non-redundant sequences. The genes obtained were further filtered by confirming the presence of the conserved domains *Cellulose_synt/GT2* using a batch blast search at the CDD (conserved domain database) of NCBI. Homoeologous genes from each of the three genomes were named as *TaCsIXA*, *TaCsIXB* or *TaCsIXD*, where *X* denotes the gene number and the last suffix stands for the respective genome. Alignment of the sequences of all newly identified wheat *Csl* genes is given in additional file 1.

6.3.3 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) [29]. Multiple protein sequence alignments were performed using Clustal omega (<u>http://www.ebi.ac.uk/Tools/msa/clustalo/</u>) (Sievers et al. 2011). The resulting alignments were analysed for the presence of conserved catalytic motifs (DXD and D, D, QXXRW) of the *GT2* superfamily. The conserved patterns of aligned sequences were highlighted using the sequence manipulation suite: Color align conservation (<u>http://www.bioinformatics.org/sms2/color_align_cons.html</u>) (Stothard 2000). The conserved domains were predicted using CCD database (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) (Kaur et al. 2013; Marchler-Bauer et al. 2014). Due to the resemblance of *CslD* with *CesA* genes and its probable role in cellulose synthesis, we specifically focused on *TaCslD* subfamily. Gene structures and intron evolution of *TaCslD* were predicted using the gene structure display server 2.0 (<u>http://gsds.cbi.pku.edu.cn/</u>) via the genomic and cDNA sequences.

6.3.4 Evolutionary relationships of Csl genes

A total of 215 CSL proteins from Arabidopsis, Brachypodium, maize, rice and wheat were used to predict the phylogenetic history. The phylogeny of the *CslD* subfamily was also determined separately from these species. For phylogenetic analysis, the amino acid sequences of CSL proteins were aligned using the MUSCLE (Edgar 2004) and the evolutionary history was inferred using Neighbor-Joining methods (Saitou and Nei 1987). 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 (Zuckerkandl and Pauling 1965) and are given as the number of amino acid substitutions per site. The rate of variation among sites was modelled with a gamma distribution (shape parameter = 1) and all positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013). A file containing the FASTA sequences of 215 CSL proteins is provided as the text file S_{-1} .

6.3.5 RNA-seq expression analysis

Publicly available RNA-seq data generated from hexaploid bread wheat (var. *Chinese spring*) was used to predict the expression of newly identified wheat *Csl* genes. The data was compiled from five different wheat tissues (leaf, spike, root, grain, and stem) collected at seedling, vegetative and

reproductive stages of development (Choulet et al. 2014a). The relative expression of each *TaCsl* subfamily was presented as a heat map generated from transcript per 10 million reads for each gene using wheat expression browser powered by expVIP (http://www.wheat-expression.com).

6.4. Results

6.4.1 Identification and classification of Csl gene family in bread wheat

Our search resulted in the identification of 108 cellulose synthase-like (TaCsl) genes from bread wheat using the conserved domain models PF00535 and PF03552, which are specific to the GT2superfamily. These genes include 2-3 homoeologous copies of each gene from the A, B and D genomes. To characterize the newly identified genes, a phylogenetic tree was constructed using multiple sequence alignments of full-length derived protein sequences from Arabidopsis, Brachypodium, maize, rice and wheat. An unrooted neighbor-joining (NJ) tree for the 215 Csl genes from these species is shown in Fig 1. TaCsl genes were grouped into seven subfamilies including TaCslA (32 genes), TaCslC (13 genes), TaCslD (12 genes), TaCslE (10 genes), TaCslF (29 genes), TaCslH (8 genes), and TaCslJ (4 genes) (Fig 2). The TaCslA and TaCslC sub-families were closely related as shown by their taxonomic distribution and phylogenies. These subfamilies were found to be highly conserved in all the land plants analysed in this study. A strong resemblance between TaCslF and TaCslD was observed, although TaCslF is specific only to grasses and TaCslD is present in all plants (Yin et al. 2014). Subfamilies TaCslE, TaCslH, and TaCslJ were phylogenetically diverse, however, TaCslJ was found to be closer to TaCslA and TaCslC subfamilies. 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) have been used

for each gene name. 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.

6.4.2 Splice variants of Csl genes

Among *Csl* genes, 22 genes appeared to encode two or more proteins because of the presence of alternative splicing sites, as predicted by Ensembl database, which resulted in a total of 137 probable *Csl* protein products. Splice variants were discovered in all subfamilies of the *TaCsl* genes except *TaCslD* (Table 2). In subfamily *TaCslA*, 6 genes alternatively spliced to form 13 proteins whereas, in subfamily *TaCslC*, 5 genes were alternatively splicing resulting in 14 proteins. Similarly, for subfamilies *TaCslE* and *TaCslF*, alternate splicing resulted in 7 and 10 splice variants respectively. Similarly, alternative splicing of 1 and 2 genes respectively generated 3 and 4 proteins in the *CslH* and *CslJ* subfamilies (Fig 2). Of all the splice variants, 51% stemmed from the exon skipping, ~24% from the selection of alternative 5' and 3' splice sites and the rest, ~24%, from intron retention (Table 2).

6.4.3 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). Subfamilies *TaCslA* and *TaCslC* contained the *GT* 2_3 and *CslD*, *CslE*, *CslF*, *CslH*, *CslJ* subfamilies contained the *cellulose_synt* domain (Fig 2). All *TaCsl* genes possessed the motifs D, D, DXD and QXXRW except eight truncated genes that possessed either of these four motifs (*TaCslA7_2DS*, *TaCslD4_1BS*, *TaCslD4_5BS*, *TaCslF2_7BL*, *TaCslF6_7AL*, *TaCslF6_7DL*, *TaCslH3_3AS*, *TaCslH2_3B*). The motifs DXD and

QXXRW were diverse in different subfamilies of *Csl* genes, such as for *TaCslA* (DMD, QQH/FRW); *TaCslC* (DMD, QQHRW); *TaCslD* (DCD, QVLRW); *TaCslE* (DCD, QHKRW); *TaCslF* (DC/GD, QI/VL/VRW); *TaCslH* (DCD QF/YKRW); *TaCslJ* (DCD, QNKRW). These motifs are highlighted in alignment files in the Appendix 6.

6.4.4 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 (Saitou and Nei 1987), in MEGA6 (Tamura et al. 2013) and the orthologs from various species were grouped into different clades (Fig 3). This was based on the rice *Csl* genes because complete nomenclature of rice genes is well documented. All the genes were divided into three clades. The first clade contained CslD2 and CslD1 genes from rice and their orthologs from different species. The tree homoeologous genes of wheat branched together with OsCslD1, wheat genes under this clade were named TaCslD1_1AL, TaCslD1_1BL, and TaCslD1_1DL from each of the 1AL, 1BL, and 1DL genomes. The second clade was branched into two subgroups containing the orthologs of rice genes CslD3 and CslD5 from different species. First subgroup of wheat genes were designated as TaCslD3_2AS, TaCslD3_2BS, and TaCslD3_2DS. The genes of the second subgroup were named TaCslD5_7AL, TaCslD5_7BL, and TaCslD5_7DL. The last clade was composed of the orthologs of the rice CslD4 genes and wheat genes, named TaCslD4_5BS, TaCslD4_1BS and TaCslD4_5DS. Here we found only two homoeologs of TaCslD4, but a gene from the 1BS genome (TaCslD4_1BS) of wheat grouped together with TaCslD4 genes (bootstrap = 100) (Table 1). This gene shared sequence identity of 85% with TaCslD4_5BS at amino acid level. OsCslD genes shared 73-86 % sequence identity with the corresponding wheat orthologs.

6.4.5 Gene structure and intron evolution of *TaCslD* subfamily

A total of 12 TaCslD genes were found in bread wheat. The length of CslD subfamily genes ranged from 1519-5864 bp. The *TaCslD4_1BS* gene was the shortest and *TaCslD1_1AL* was the longest. Homoeologous copies of all genes shared sequence identity ranging from 87-94% at the genomic scale. The variation in size among different genes was primarily due to the number and length of introns (Fig 4). Intron number in all the genes varied from 2 to 4. Two homoeologs: TaCslD1_1AL and TaCslD1_1BL each had three introns, however, a third homoeolog (TaCslD1_1DL) had four. Genes TaCslD3, TaCslD4 and their homoeologs had three introns each, except TaCslD4 1BS with only two introns. TaCslD5 and its homoeologs also had two introns each. Here we have predicted three different phases of intron evolution as 0, 1, or 2; referring to the insertion of an intron between two consecutive codons, between the first and the second base or second and the third base of a codon, respectively (Dhaliwal et al. 2014; Kaur et al. 2016). Genes from the *TaCslD* subfamily exhibited variable patterns of intron phase distribution. Introns 1, 2 and 3 of TaCslD1_1AL, TaCslD1_1BL and TaCslD1_1DL had 2, 0, and 0 phase distribution, the 4th intron of TaCslD1 1DL had a phase distribution of 0. Introns 1 and 2 of TaCslD3 2AS, TaCslD3 2BS and TaCslD3 2DS both had phase distribution of 0. The third intron of these genes was in phase 2, 1 and 2 respectively. Genes TaCslD4 5BS, TaCslD4 5DS, TaCslD5 7AL, TaCslD5 7BL and TaCslD5_7DL had introns 1 and 2 in phase 2 and 0 and the third intron of TaCslD4_5BS and TaCslD4_5DS were in phase 0 and 2, respectively. TaCslD4_1BS had introns 1 and 2 in phases 1 and 0. Among all the studied genes, the largest proportion of introns (60%) was found to be in phase 0, followed by phase 2 (33.3%) with very few in phase 1 (6%).

6.4.6 RNA-seq 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 different tissues of wheat including root, leaf, stem, spike, and grain. In the case of TaCslA genes, we have retrieved the expression of 32 TaCslA genes excluding splice variants. 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 revealed exceptionally higher expression in reproductive leaf tissue while the transcript abundance of the remaining genes was very low (Fig 5A). The 13 genes of the *TaCslC* subfamily were expressed highly in root and spike tissues. Two genes, TaCslC1 and TaCslC7 and their homoeologs displayed moderate to higher expression in all the tissues at seeding and vegetative stage. One gene (TaCslC10_5DL) exhibited moderate to high expression levels in all the tissues studied except reproductive stem and grain tissues (Fig 5B). Most of the genes of *TaCslD* subfamily revealed moderate to a high expression level in spike and root tissues and their expression was very low in all other tissues (Fig 5C). Three of the ten TaCslE subfamily genes (TaCslE2_6AL, TaCslE2_6BL and TaCslE3) showed moderate to a high expression in all tissues. The remaining genes were expressed at a very low level in all tissues (Fig 5D). A mixed pattern of expression was observed in the large *TaCslF* subfamily. Three genes (*TaCslF6_7AL*, *TaCslF6_7BL* and *TaCslF6_7DL*) demonstrated higher expression in all the tissues except leaves at reproductive stage. Two genes (TaCslF4_2BS and TaCslF4_2DS) indicated higher expression in stem tissues, while low to moderate in all other tissues. All other genes revealed low to moderate expression in one or more tissues (Fig 5E). In the TaCslH subfamily, one out of eight genes (TaCslH1_2BL) showed moderate to high expression levels in leaves, stem and spike tissues. The remaining genes also unveiled low to moderate expression in

all the tissues (Fig 5F). Three out of four members of the subfamily *TaCslJ* possessed low to moderate expression levels in leaves and root tissues while one gene (*TaCslJ1_3DS*) was poorly expressed in all the tissues (Fig 5G).

6.5 Discussion

The grass cell walls are composed of about 20-40% non-cellulosic polysaccharides, while the amount and composition of these polysaccharides vary widely in different plant species (Saxena and Brown 1995). Several genes of the *Csl* family have been reported to encode the corresponding enzymatic proteins hemicellulose synthesis (Liepman et al. 2005; Burton et al. 2006a; Cocuron et al. 2007; Doblin et al. 2009; Goubet et al. 2009; Yin et al. 2009; Wang et al. 2010b). As a detailed understanding of the identity of *Csl* genes in wheat was lacking, thus we undertook this study to fill this gap in wheat cell wall formation.

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 (Yin et al. 2014). Around a quarter of the identified *Csl* genes were alternatively spliced, resulting in 29 splice variants. A recent study revealed that the alternative splicing is common in plants and accounts for about 20% of the loci transcribed in the leaf and spike tissues of *Aegilops tauschii* (Iehisa et al. 2017). This phenomenon is apparently meant to increase the diversity of gene products to increase the fitness of an organism (Zhou et al. 2003).

Physical mapping revealed the distribution of TaCsl genes on all wheat chromosomes except one, chromosome 4 (Fig S1). A similar trend of *Csl* gene distribution was observed in barley (Burton et al. 2008; Schreiber et al. 2014a; Schwerdt et al. 2015). More than half the *TaCsl* genes were located on chromosomes 2 (32%) and chromosome 3 (22%). These two chromosomes appear to be *TaCsl* hotspots and can be targeted in breeding efforts for altering cell wall composition. Five of nine *CslF* genes in barley were located on chromosome 2H. A similar cluster of *CslF* genes was also detected in the conserved syntenic regions of Brachypodium, rice and sorghum on chromosomes 1, 7 and 2, respectively (Schwerdt et al. 2015).

In silico expression analyses across different tissues suggested that of the 32 genes from the subfamily *CslA* only half or so were expressed at varying levels. Moreover, there was no commonality seen between these genes based on their transcript abundance, as different genes of the same subfamily express differently in the root, leaf, stem, spike, and grain tissues during vegetative and reproductive growth stages. Reverse genetic and biochemical approaches in *Arabidopsis* have associated the *CslA* genes with glucomannan biosynthesis (Goubet et al. 2009).

In the case of subfamilies *TaCslC* and *TaCslD*, most of the genes showed relatively higher expression levels in root and spike tissues during the vegetative as well as reproductive phases. Heterologous expression studies in the case of *Pichia* revealed that the *CslC* genes are involved in the synthesis of the 1-4- β -glucan backbone of the xyloglucan and some other polysaccharides (Cocuron et al. 2007). Of all *Csl* genes, the *CslD* subfamily is conserved in all land plants and most closely related to the *CesA* gene family, between 40-50% amino acid sequence similarity (Doblin et al. 2001) Similar to *CesAs*, the *CslD* subfamily is ubiquitous in all plant genomes examined to date, unlike other, taxa-specific *Csl* subfamilies (Hunter et al. 2012). Previous reports also showed the involvement of certain members of the *CslD* subfamily in tip growth, development of root hairs (Kim et al. 2007a; Yuo et al. 2011), normal plant growth (Li et al. 2009; Hunter et al. 2012), pollen tube growth, and meristem morphology and architecture (Bernal et al. 2007; Li et al. 2009). More recently, their role in resistance against biotic stresses has been described (Douchkov et al. 2016). Adding to this discussion, our *in silico* expression analysis sheds light on the involvement of certain *TaCslD* genes during spike development. These results are relevant to the reduction in the number and width of spikelets shown by mutant *slender leaf 1 (sle1)* that encodes the rice CSLD4 protein (Yoshikawa et al. 2013).

Two groups of *Csl* genes, *CslF* and *CslH*, have evolved independently in grasses (Burton et al. 2011a). A third group *CslJ* had been recently identified as grass specific (Farrokhi et al. 2006). Although *TaCslF6* gene showed higher expression in all the studied tissues except reproductive leaf tissue, it was the only member of the *TaCslF* subfamily which expressed highly in grain tissue. Several studies have demonstrated the functional role of *CslF6* and *CslH* in the synthesis of (1-3), (1-4) β -glucan (mixed-linked glucan or MLG) (Doblin et al. 2009; Nemeth et al. 2010; Taketa et al. 2012; Schreiber et al. 2014a). Of all the genes in these families, only *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. 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 homoeologous copies from three different genomes of bread wheat, which agree with the studies reporting the unequal contributions of the three genomes towards the gene expression (Mochida et al. 2004; Hu et al. 2013; Tanaka et al. 2015). Interestingly, the homoeologous copies of *TaCslD* genes also differed from each other in terms of intron phase evolution; indicating structural and functional divergence of homoeologous gene copies (Fig 4). The three homoeologs of each gene were not observed for all genes identified here. This could be because of the incomplete sequencing information or because of the silencing of the genes during the evolution of allohexaploid wheat (Bottley et al. 2006; Aramrak et al. 2015; Jordan et al. 2015).

6.6 Conclusion

We have identified 108 *TaCsl* genes in bread wheat and classified them into seven subfamilies (*CslA, CslC, CslD, CslE, CslF, CslH, CslJ*). In most cases, two to three homoeologs of each gene were identified as was expected for a hexaploid crop like bread wheat. These genes were located on all the wheat chromosomes except chromosome 4, whereas chromosome 2 and 3 contained approximately half of all the *Csl* genes. Only of the homeoalleles of a single *CslF* gene, *CslF6* were expressed in the grain, suggesting its key role in mixed-linked glucan formation. Neither *CslJ* or *CslH* were expressed in the grain. Information in this report will be helpful in designing experiments to alter wall composition in wheat for various purposes.

Fig 6.1 An unrooted phylogenetic tree representing the *Cellulose synthase-like* gene family from Arabidopsis, maize, rice and wheat using MEGA6. Tree was constructed using Neighbour joining (NJ) method with 100 bootstrap value. Different colors represent the subfamilies with orthologous CSL proteins from different species. The bar provides a scale for the branch length in the horizontal dimension. The line segment with the number '0.5' means that an equal length of the branch between the CSL proteins represents a change of 0.5 AA.



Fig 6.2 Distribution of *TaCsl* genes and their splice variants in seven subfamilies and their corresponding pfam domains used to identify *TaCsl* gene family.



Fig 6.3 An unrooted phylogenetic tree representing the *CslD* subfamily from Arabidopsis, Brachypodium, maize, rice and wheat using MEGA6. Tree was constructed using Neighbour joining (NJ) method with 100 bootstrap value. Different colors represent orthologous *Csl* genes from different species. Arabidopsis-blue, Brachypodium-purple, maize-sky blue, rice-green, wheat-red.



Fig 6.4 Structural features and phases of intron evolution of the *CslD* subfamily genes. Drawn to scale, exons are represented by red boxes and introns by back lines. Corresponding phases of intron evolution (0, 1, and 2) for the CslD genes are shown on the top of the black lines.



Fig 6.5 Heat map showing the expression profiling of wheat *Cellulose synthase-like (TaCsl)* genes at seedling, vegetative and reproductive stages. (A) *CslA* (B) *CslC* (C) *CslD* (D) *CslE* (E) *CslF* (F) *CslH* & *CslJ*. RNA-seq data 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.

Growth stages	Seed	lling		Vege	tative			Rep	roduc	tive		
Gene Name	Root	Leaf	Root	Leaf	Stem	Spike	Root	Leaf	Stem	Spike	Grain	
TaCslA1_6AS												High
TaCslA1_6BS												
TaCslA2_1AS												
TaCslA2_2AL												
TaCslA2_2BL												
TaCslA2_2DL												
TaCslA3_7AS												
TaCslA3_7BS												
TaCslA3_7DS												
TaCslA4_6AS												
TaCslA4_6BS												
TaCslA4_6DS												
TaCslA5_2AS												
TaCslA5_2BS												
TaCslA5_2DS												
TaCslA6_3AL												
TaCslA6_3B												
TaCslA6_3DL												
TaCslA7_2AS												
TaCslA7_2DS												
TaCslA8_3B												
TaCslA8_3DS												
TaCslA9												
TaCslA9_7AL												
TaCslA9_7BL												
TaCslA10												
TaCslA10_7AL												
TaCslA10_7BL												
TaCslA10_7DL												
TaCslA11_3AS												
TaCslA11_3B												
TaCslA11_3DS												LOW

Growth stages	Seed	Seedling		Vegetative			Reproductive					
Gene Name	Root	Leaf	Root	Leaf	Stem	Spike	Root	Leaf	Stem	Spike	Grain	
TaCslC1_3AL												High
TaCslC1_3DL												
TaCslC3_3AS												
TaCslC3 3B												
TaCslC3_3DS												
TaCslC7_1AL												
TaCslC7_1BL												
TaCslC7 1DL												
TaCslC9 1BL												
TaCslC9 1DL												
TaCslC10 5AL												
TaCslC10 5BL												
TaCslC10 5DL												Low

Growth stages	Seed	dling		Vege	tative	e		Rep	rodu	ctive		
Gene Name	Root	Leaf	Root	Leaf	Stem	Spike	Root	Leaf	Stem	Spike	Grain	
TaCslD1_1AL												High
TaCslD1_1BL												
TaCslD1_1DL												
TaCslD3_2AS												
TaCslD3_2BS												
TaCslD3_2DS												
TaCslD4_1BS												
TaCslD4_5BS												
TaCslD4_5DS												
TaCslD5_7AL												
TaCslD5_7BL												
TaCslD5_7DL												Low

Growth stages	See	dling		Vegetative			Reproductive					
Gene Name	Root	Leaf	Root	Leaf	Stem	Spike	Root	Leaf	Stem	Spike	Grain	
TaCslE1_5BL												High
TaCslE1_5DL												
TaCslE2_6AL												
TaCslE2_6BL												
TaCslE2_6DL												
TaCslE3												
TaCslE4_6DS												
TaCslE6 5AL	-											
TaCslE6_5BL												
TaCslE6 5DL												Low

Growth stages	See	dling		Vegetative				Reproductive				
Gene Name	Root	Leaf	Root	Leaf	Stem	Spike	Root	Leaf	Stem	Spike	Grain	
$T_{a}C_{s}IF1 2AI$		_			•1	•1	_		•1	•1	•	Hiał
TaCsIF1 2RL												11181
TaCsIF1 2DL												
TaCsIF1 2DL												
TaCsIF 2 74L												
TaCsIF2 7BL												
TaCsIF2 7DL												
TaCsIF3 24S												
TaCsIF3 2BS												
TaCsIF 3 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												
TaCslF10 1BS												Low

Growth stages	Seed	lling	٦	Veget	tative	•	Reproductive						
Gene Name	Root	Leaf	Root	Leaf	Stem	Spike	Root	Leaf	Stem	Spike			
TaCslH1_2AL													
TaCslH1_2BL													
TaCslH1_2DL													
T G 1112 AD													

Low

Fig 6.6 Pie chart showing the percentage of *TaCsl* genes on wheat chromosomes.



Grain

High

 Table 6.1 Homoeologous copies of wheat Csl genes with their corresponding orthologs from rice.

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_592860_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_6BS_TGACv1_513376_AA1639390.1	TaCslA4_6BS	CslA10/4/2
13	TRIAE_CS42_2BS_TGACv1_146583_AA0468630.1	TaCslA5_2BS	CslA5/7
14	TRIAE_CS42_2AS_TGACv1_113418_AA0355820.1	TaCslA5_2AS	CslA5/7
15	TRIAE_CS42_2DS_TGACv1_177473_AA0578070.1	TaCslA5_2DS	CslA5/7
16	TRIAE_CS42_3DL_TGACv1_249033_AA0835410.1	TaCslA6_3DL	CslA11
17	TRIAE_CS42_3B_TGACv1_221079_AA0729630.1	TaCslA6_3B	CslA11
18	TRIAE_CS42_3AL_TGACv1_197519_AA0666560.1	TaCslA6_3AL	CslA11
19	TRIAE_CS42_2AS_TGACv1_113300_AA0354190.1	TaCslA7_2AS	CslA5/7
20	TRIAE_CS42_2DS_TGACv1_177798_AA0584795.1	TaCslA7_2DS	CslA5/7
21	TRIAE_CS42_3B_TGACv1_220828_AA0720500.1	TaCslA8_3B	CslA11
22	TRIAE_CS42_3DS_TGACv1_273022_AA0927600.1	TaCslA8_3DS	CslA11
23	TRIAE_CS42_U_TGACv1_642146_AA2112270.1	TaCslA9	CslA9
24	TRIAE_CS42_7BL_TGACv1_579090_AA1903960.1	TaCslA9_7BL	CslA9
25	TRIAE_CS42_7AL_TGACv1_558725_AA1795700.1	TaCslA9_7AL	CslA9
26	TRIAE_CS42_U_TGACv1_642146_AA2112290.1	TaCslA10	CslA9

27	TRIAE_CS42_7DL_TGACv1_602617_AA1962870.1	TaCslA10_7DL	CslA9
28	TRIAE_CS42_7AL_TGACv1_557254_AA1778850.1	TaCslA10_7AL	CslA9
29	TRIAE_CS42_7BL_TGACv1_578444_AA1895100.1	TaCslA10_7BL	CslA9
30	TRIAE_CS42_3AS_TGACv1_210508_AA0674280.1	TaCslA11_3AS	CslA11
31	TRIAE_CS42_3DS_TGACv1_272005_AA0912960.1	TaCslA11_3DS	CslA11
32	TRIAE_CS42_3B_TGACv1_223332_AA0780350.1	TaCslA11_3B	CslA11
33	TRIAE_CS42_3DL_TGACv1_251593_AA0882850.1	TaCslC1_3DL	CslC1
34	TRIAE_CS42_3AL_TGACv1_197197_AA0665370.1	TaCslC1_3AL	CslC1
35	TRIAE_CS42_3DS_TGACv1_271926_AA0910940.1	TaCslC3_3DS	CslC3
36	TRIAE_CS42_3B_TGACv1_220758_AA0718310.1	TaCslC3_3B	CslC3
37	TRIAE_CS42_3AS_TGACv1_211225_AA0686890.1	TaCslC3_3AS	CslC3
38	TRIAE_CS42_1DL_TGACv1_061928_AA0205730.1	TaCslC7_1DL	CslC7
39	TRIAE_CS42_1BL_TGACv1_030750_AA0099830.1	TaCslC7_1BL	CslC7
40	TRIAE_CS42_1AL_TGACv1_001272_AA0028090.1	TaCslC7_1AL	CslC7
41	TRIAE_CS42_1DL_TGACv1_062162_AA0209740.1	TaCslC9_1DL	<i>CslC10/9</i>
42	TRIAE_CS42_1BL_TGACv1_030501_AA0092480.1	TaCslC9_1BL	<i>CslC10/9</i>
43	TRIAE_CS42_5BL_TGACv1_404820_AA1311790.1	TaCslC10_5BL	CslC10/9
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	TaCslD3_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_425241_AA1392650.1	TaCslD4_5BS	CslD4
54	TRIAE_CS42_5DS_TGACv1_457675_AA1488780.1	TaCslD4_5DS	CslD4
55	TRIAE_CS42_7BL_TGACv1_577301_AA1871610.1	TaCslD5_7BL	CslD5
56	TRIAE_CS42_7AL_TGACv1_559436_AA1799630.1	TaCslD5_7AL	CslD5

57	TRIAE_CS42_7DL_TGACv1_603510_AA1985050.1	TaCslD5_7DL	CslD5
58	TRIAE_CS42_5DL_TGACv1_433536_AA1415830.1	TaCslE1_5DL	CslE6/1
59	TRIAE_CS42_5BL_TGACv1_406235_AA1342600.1	TaCslE1_5BL	CslE6/1
60	TRIAE_CS42_6DL_TGACv1_526558_AA1687090.1	TaCslE2_6DL	CslE2
61	TRIAE_CS42_6AL_TGACv1_471004_AA1500600.1	TaCslE2_6AL	CslE2
62	TRIAE_CS42_6BL_TGACv1_499967_AA1596110.1	TaCslE2_6BL	CslE2
63	TRIAE_CS42_U_TGACv1_683314_AA2158770.1	TaCslE3	CslE6/1
64	TRIAE_CS42_6DS_TGACv1_543277_AA1737920.1	TaCslE4_6DS	CslE6/1
65	TRIAE_CS42_5DL_TGACv1_433536_AA1415840.1	TaCslE6_5DL	CslE6/1
66	TRIAE_CS42_5BL_TGACv1_406235_AA1342610.1	TaCslE6_5BL	CslE6/1
67	TRIAE_CS42_5AL_TGACv1_376126_AA1232370.1	TaCslE6_5AL	CslE6/1
68	TRIAE_CS42_2DL_TGACv1_159781_AA0542640.1	TaCslF1_2DL	CslF1/2/4
69	TRIAE_CS42_2AL_TGACv1_094713_AA0301960.1	TaCslF1_2AL	CslF1/2/4
70	TRIAE_CS42_2DL_TGACv1_160109_AA0546890.1	TaCslF1_2DL	CslF1/2/4
71	TRIAE_CS42_2BL_TGACv1_130934_AA0420130.1	TaCslF1_2BL	CslF1/2/4
72	TRIAE_CS42_7BL_TGACv1_580651_AA1914920.1	TaCslF2_7BL	CslF1/2/4
73	TRIAE_CS42_7AL_TGACv1_557532_AA1782680.1	TaCslF2_7AL	CslF1/2/4
74	TRIAE_CS42_7DL_TGACv1_602590_AA1961740.1	TaCslF2_7DL	CslF1/2/4
75	TRIAE_CS42_2AS_TGACv1_113659_AA0359050.1	TaCslF3_2AS	CslF3
76	TRIAE_CS42_2DS_TGACv1_177641_AA0581710.1	TaCslF3_2DS	CslF3
77	TRIAE_CS42_2BS_TGACv1_148608_AA0494060.1	TaCslF3_2BS	CslF3
78	TRIAE_CS42_2BS_TGACv1_146146_AA0456710.1	TaCslF4_2BS	CslF1/2/4
79	TRIAE_CS42_2DS_TGACv1_179076_AA0604160.1	TaCslF4_2DS	CslF1/2/4
80	TRIAE_CS42_2DS_TGACv1_178985_AA0603230.1	TaCslF5_2DS	CslF3
81	TRIAE_CS42_2AS_TGACv1_112790_AA0345230.1	TaCslF5_2AS	CslF3
82	TRIAE_CS42_2BS_TGACv1_148027_AA0489970.1	TaCslF5_2BS	CslF3
83	TRIAE_CS42_7BL_TGACv1_577473_AA1876170.1	TaCslF6_7BL	CslF6
84	TRIAE_CS42_7AL_TGACv1_555973_AA1751470.1	TaCslF6_7AL	CslF6
85	TRIAE_CS42_7DL_TGACv1_607937_AA2011180.1	TaCslF6_7DL	CslF6
86	TRIAE_CS42_5BL_TGACv1_409916_AA1366600.1	TaCslF7_5BL	CslF7

87	TRIAE_CS42_5DL_TGACv1_433902_AA1424880.1	TaCslF7_5DL	CslF7	
88	TRIAE_CS42_5AL_TGACv1_374191_AA1193100.1	TaCslF7_5AL	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_U_TGACv1_641498_AA2096480.1	TaCslF10	CslF9	
96	TRIAE_CS42_1BS_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_3B_TGACv1_221049_AA0728260.1	TaCslH2_3B	Csl	
101	TRIAE_CS42_3DS_TGACv1_273502_AA0931770.1	TaCslH2_3DS	Csl	
102	TRIAE_CS42_3DS_TGACv1_271739_AA0907200.1	TaCslH3_3DS	Csl	
103	TRIAE_CS42_3AS_TGACv1_212952_AA0704280.1	TaCslH3_3AS	CslH3	
104	TRIAE_CS42_3B_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	
Ensembl Gene ID	Gene name	Predicted amino acids	Splice site	Status
--	--------------	-----------------------	-----------------	---------------------
TRIAE_CS42_6BS_TGACv1_513375_AA1639370.1	TaCslA1_6BS	581	-	Wild type
TRIAE_CS42_6BS_TGACv1_513375_AA1639370.2		390	Exon 1 and 2	Exon skipping
TRIAE_CS42_6BS_TGACv1_513376_AA1639390.2	TaCslA4_6BS	528	-	Wild type
TRIAE_CS42_6BS_TGACv1_513376_AA1639390.1		393	Exon 1 and 2	Exon skipping
TRIAE_CS42_7AS_TGACv1_569190_AA1809650.1	TaCslA3_7AS	551	-	Wild type
TRIAE_CS42_7AS_TGACv1_569190_AA1809650.2		380	Exon 7, 8 and 9	Exon skipping
TRIAE_CS42_7AS_TGACv1_569190_AA1809650.3		503	Exon 9	Exon skipping
TRIAE_CS42_7DL_TGACv1_602617_AA1962870.2	TaCslA10_7DL	515	-	Wild type
TRIAE_CS42_7DL_TGACv1_602617_AA1962870.1		555	Intron 8	Intron retention
TRIAE_CS42_3DL_TGACv1_249033_AA0835410.2	TaCslA6_3DL	524	-	Wild type
TRIAE_CS42_3DL_TGACv1_249033_AA0835410.1		572	Intron 1	Intron retention
TRIAE_CS42_3B_TGACv1_221079_AA0729630.1	TaCslA6_3B	571	-	Wild type
TRIAE_CS42_3B_TGACv1_221079_AA0729630.2		538	Exon 2	Exon skipping
TRIAE_CS42_5BL_TGACv1_404820_AA1311790.1	TaCslC10_5BL	712	-	Wild type
TRIAE_CS42_5BL_TGACv1_404820_AA1311790.2		468	Exon 5	Alternative 5' site
TRIAE_CS42_5BL_TGACv1_404820_AA1311790.3		504	Exon 1	Exon skipping
TRIAE_CS42_5DL_TGACv1_435778_AA1454840.1	TaCslC10_5DL	708	-	Wild type
TRIAE_CS42_5DL_TGACv1_435778_AA1454840.2		502	Exon1	Exon skipping
TRIAE_CS42_5AL_TGACv1_374268_AA1195590.3	TaCslC10_5AL	703	-	Wild type
TRIAE_CS42_5AL_TGACv1_374268_AA1195590.2		496	Exon 5	Alternative 5' site
TRIAE_CS42_5AL_TGACv1_374268_AA1195590.1		501	Exon 5	Exon skipping
TRIAE_CS42_3DL_TGACv1_251593_AA0882850.1	TaCslC1_3DL	704	-	Wild type
TRIAE_CS42_3DL_TGACv1_251593_AA0882850.2		493	Exon 5	Exon skipping
TRIAE_CS42_3DL_TGACv1_251593_AA0882850.3		679	Exon 1	Alternative 3' site
TRIAE_CS42_3AL_TGACv1_197197_AA0665370.1	TaCslC1_3AL	704	-	Wild type

Table 6.2 Status of splice variants of *Csl* genes in wheat genome.

TRIAE_CS42_3AL_TGACv1_197197_AA0665370.2		560	Exon 5	Alternative 3' site
TRIAE_CS42_3AL_TGACv1_197197_AA0665370.3		679	Exon 5	Alternative 5' site
TRIAE_CS42_6AL_TGACv1_471004_AA1500600.1	TaCslE2_6AL	667	-	Wild type
TRIAE_CS42_6AL_TGACv1_471004_AA1500600.2		737	Intron 8	Intron retention
TRIAE_CS42_6AL_TGACv1_471004_AA1500600.3		635	Exon 4	Alternative 5' site
TRIAE_CS42_5DL_TGACv1_433536_AA1415830.1	TaCslE1_5DL	728	-	Wild type
TRIAE_CS42_5DL_TGACv1_433536_AA1415830.2		684	Exon 4	Exon skipping
TRIAE_CS42_5BL_TGACv1_406235_AA1342600.1	TaCslE1_5BL	734	-	Wild type
TRIAE_CS42_5BL_TGACv1_406235_AA1342600.2		728	Exon 1	Exon skipping
TRIAE_CS42_2DS_TGACv1_177641_AA0581710.1	TaCslF3_2DS	847	-	Wild type
TRIAE_CS42_2DS_TGACv1_177641_AA0581710.2		735	Exon 2	Alternative 3' site
TRIAE_CS42_2DS_TGACv1_179076_AA0604160.1	TaCslF4_2DS	783	-	Wild type
TRIAE_CS42_2DS_TGACv1_179076_AA0604160.2		700	Exon 1	Exon skipping
TRIAE_CS42_2BS_TGACv1_147667_AA0486240.1	TaCslF9_2BS	877	-	Wild type
TRIAE_CS42_2BS_TGACv1_147667_AA0486240.2		796	Exon 1	Exon skipping
TRIAE_CS42_5BL_TGACv1_409916_AA1366600.1	TaCslF7_5BL	745	-	Wild type
TRIAE_CS42_5BL_TGACv1_409916_AA1366600.2		815	Intron 2	Intron retention
TRIAE_CS42_5AL_TGACv1_374191_AA1193100.1	TaCslF7_5AL	792	-	Wild type
TRIAE_CS42_5AL_TGACv1_374191_AA1193100.2		807	Intron 1	Intron retention
TRIAE_CS42_2AL_TGACv1_094351_AA0296300.1	TaCslH1_2AL	737	-	Wild type
TRIAE_CS42_2AL_TGACv1_094351_AA0296300.2		660	Exon 9	Exon skipping
TRIAE_CS42_2AL_TGACv1_094351_AA0296300.3		480	Exon 6,7,8 and 9	Exon skipping
TRIAE_CS42_3AS_TGACv1_210908_AA0681280.1	TaCslJ1_3AS	738	-	Wild type
TRIAE_CS42_3AS_TGACv1_210908_AA0681280.2		766	Intron 4	Intron retention
TRIAE_CS42_3DS_TGACv1_272756_AA0924850.2	TaCslJ2_3DS	609	-	Wild type
TRIAE_CS42_3DS_TGACv1_272756_AA0924850.1		734	Intron 1	Intron retention

CHAPTER VII. GENERAL DISCUSSION AND FUTURE STUDIES

7.1 General discussion

Plant cells exhibit special characteristics known as cell walls that provide basic infrastructure, mechanical support and a barrier against pathogen invasion throughout plant's lifecycle. Cell walls being the most abundant renewable biomass are getting attention for their use as dietary fibres, food additives, a raw material for biofuels, and fodder for livestock (Taylor-Teeples et al. 2015). These are the dynamic structures composed of complex polysaccharides such as celluloses, hemicelluloses, pectins and lignins along with highly glycosylated proteins (Doblin et al. 2010). These components vary greatly in their relative proportion and fine structure with the developmental stages and between different species (Fincher 2009; Hatfield et al. 2009).

Primary cell walls usually composed of cellulose, hemicellulose and pectins and provide shape and flexibility to young plant cells. Whereas secondary cell walls are composed of cellulose, hemicellulose and lignin and provide thickness and rigidity to mature plant cells. Secondary cell walls contribute more towards the total biomass production owing to their relatively higher thickness (Keegstra 2010). Considering their vital functions in plants, various medicinal and industrial uses, cell walls are getting much attention for research.

Biofuels from lignocellulosic biomass represent a potential source of energy with low carbon emissions. On the global scale, 3.7×10^{15} g of lignocellulosic biomass is produced per year from the residues of barley, maize, rice, soybean, sugar cane and wheat crops (Bentsen et al. 2014). This enormously abundant biomass can generate the energy equivalent to the 66 % of the energy required for transport worldwide (Baldwin et al. 2017). Among the various crop residues, wheat straw is one of the most practical biomass feedstocks used for the production of commercial

. ۲.

biofuels (Baldwin et al. 2017). Current varieties of wheat have not been designed for cellulosic biofuel production, however, great potential exists at the genetic level to alter lignocellulose composition of wheat and other grasses (Ong et al. 2014). Therefore, an efficient utilisation of lignocellulosic biomass as raw materials for biofuels or other bioproducts requires a thorough investigation of cell wall genetic architecture.

Celluloses and hemicelluloses present in the lignocellulose account for the bulk of renewable biomass. Cellulose is the major structural component of plants and composed of linear chains of β -1, 4-glucan units, synthesised at plasma membranes. A number of genes have been reported to be associated with cellulose synthesis in different plant species. A major class of these genes is known as *Cellulose synthase A* (*CesA*) (Suzuki et al. 2006). On the other hand, hemicelluloses are the group of heterogeneous polysaccharides synthesised on the Golgi membranes. These includes xyloglucans, xylans, mannans and glucomannans, and β -(1-3, 1-4)-glucans in the walls of terrestrial plants. However, *Cellulose synthase-like* genes (*Csl*) genes account for the synthesis of various hemicellulose components in diverse tissues at different developmental stages of plants. Typically, structural and functional characterization of *Csl* genes (Liepman et al. 2005; Burton et al. 2006a; Cocuron et al. 2007; Doblin et al. 2009; Goubet et al. 2009; Yin et al. 2009; Wang et al. 2010b) and *CesA* genes have been performed in Arabidopsis (Arioli et al. 1998; Richmond and Somerville 2000; Taylor et al. 2003), maize (Holland et al. 2000; Appenzeller et al. 2004), and rice (Tanaka et al. 2003; Wang et al. 2012a).

However, due to complex nature of wheat genome, *CesA/Csl* gene families have not been well defined in wheat. Furthermore, the full genome sequence of bread wheat has not been available until recently, which posed a major challenge in exploring these complex gene families. Being hexaploid wheat possess three genomes and corresponding homoeologous copies of each

gene are expected. Genes of *CesA/Csl* belongs to a highly conserved superfamily of genes known as *Glycosyltransferase 2*. These genes share a large sequence similarity among each other or within the subgroup, which makes it a difficult task to identify and characterise these genes in hexaploid wheat.

The results generated in chapter III, are the first report of identification and comprehensive structural analysis of CesA genes in bread wheat. Total 22 CesA genes including their paralogs from the homoeologous bread wheat genomes A, B and D, were identified using a comparative genomics approach. These genes were analysed for specific structural features such as domains, motifs and phases of intron evolution. Previous studies have shown the involvement of distinct CSCs for the synthesis of primary and secondary cell walls in plants (Arioli et al. 1998; Tanaka et al. 2003; Taylor et al. 2003). Following that, a novel motif "CQIC" was identified in the present study that structurally differentiates PCW and SCW CESAs from both the monocots and dicots. Additionally, several other motifs were identified that were highly conserved among the CESA orthologs from different species (Arabidopsis, barley, maize, rice and wheat). The newly identified motifs will enable researchers to easily extricate PCW or SCW related CesAs and to identify one to one orthologs of different CesA genes in various plant species. Comparable to the distribution patterns of CesA genes in Arabidopsis, barley and maize (Holland et al. 2000; Burton et al. 2004), TaCesAs were also found to be scattered all over the wheat genome, which reflect their significance in plants. In vitro expression analysis showed higher transcript abundance of three SCW TaCesAs (TaCesA4, TaCesA7, and TaCesA8) in mature stem tissues. Among these three essential components of SCW (Tanaka et al. 2003; Taylor et al. 2003; Kotake et al. 2011; Wang et al. 2012a), *TaCesA4* showed relatively higher expression in mature stem tissues.

Being an essential component of synthesis of cellulose in SCW, *TaCesA4* gene identified in chapter III was selected to further validate its function in bread wheat. Chapter IV explains the functional characterization of *TaCesA4* gene using BSMV-based VIGS, which has recently emerged as a rapid functional genomics tool in cereals (Bennypaul et al. 2012a). A significantly lower transcript abundance and cellulose content in the *TaCesA4* silenced plants correlates with the previous finding suggesting its role in the cellulose synthesis (Tanaka et al. 2003; Taylor et al. 2003; Kotake et al. 2011; Wang et al. 2012a). Conversely, the histological analysis revealed that the silencing of SCW *CesA* at booting stage does not pose much effect on the shape and arrangement of xylem, phloem and mesophyll cells in the stem tissues.

In addition to *CesA* genes, some other classes of genes including *Glycosyl Hydrolase 9* (*GH9*) and *Sucrose synthase* (*SuSy*) have been reported to affect the cellulose synthesis in plants (Fujii et al. 2010). The involvement of several genes in this process explains the complexity of underlying mechanism (Kotake et al. 2011). Chapter V was planned to explore the novel genomic regions affecting the variability of cellulose content among diverse spring wheat genotypes. The stem internodes of 265 spring wheat varieties were analysed in triplicate for cellulose content variation. The percentage cellulose data was associated with GBS generated 21073 SNP markers using genome-wide association studies (GWAS) using fixed and random model circulating probability unification (FarmCPU) (Liu et al. 2016). Novel genes (β -tubulin, and Auxin-induced protein, 5NG4 and UDP-glycosyl transferase 85A2) were discovered, that are linked to the differences in cellulose content among different wheat genotypes. The genes identified in this study were previously known for their association with cellulose microfibril deposition (Paredez et al. 2006; Chan et al. 2007; Wightman and Turner 2008; Crowell et al. 2009; Gutierrez et al. 2009; Chan et al. 2016), cell division and expansion (Qiu et al. 2013), transfer of UDP-

glucose to the catalytic sites (Lairson et al. 2008), but not for cellulose content variation. Further characterization of these genes can help us to better understand the genetic architecture of cellulose biosynthesis. Moreover, the analysis of cellulose content variability could be an important screening tool for selection of genotypes tolerant to crop lodging, which is a common problem in most cereal crops (Ching et al. 2006).

Chapter VI represents the first report of comprehensive and large-scale data mining for the identification of Csl genes in bread wheat. A total of 108 TaCsl genes were retrieved from available sequence databases using two conserved domains: PF00535, and PF03552 (Yin et al. 2014). The newly identified genes were categorized into different subfamilies (CslA, CslC, CslD, CslE, CslF, CslH, CslJ) based on the phylogenetic analysis (Yin et al. 2009; Yin et al. 2014). As expected, none of the wheat genes were clustered with so-called dicot specific CslB and CslG subfamilies (Schwerdt et al. 2015). A detailed analysis of gene structure and intron evolution was performed for *TaCslD* sub-family, as this family palys a major role in mannan synthesis (Verhertbruggen et al. 2011; Wang et al. 2011), tip growth, development of root hairs (Kim et al. 2007a; Yuo et al. 2011), normal plant growth (Li et al. 2009; Hunter et al. 2012), pollen tube growth, and meristem architecture (Bernal et al. 2007; Li et al. 2009), and resistance to biotic stresses (Douchkov et al. 2016). Tissue or developmental stage specific in silico expression of different TaCsl genes concurred with the variability of cell wall composition among different cells and tissues (Lin et al. 2016). In-depth analysis of gene structure, evolution, and expression of this family offers a valuable resource for breeding and genetic modifications to improve wheat varieties for desirable biomass with appropriate resistance against various stresses.

7.2 Future studies

- Functional characterization of novel motif (CQIC) using CRISPR-Cas9 will generate information for better understanding of cell wall structure and functions
- New molecular markers can be devised from functionally validated *TaCesA4* for markerassisted breeding of wheat for the selection of lodging tolerance and culm strength
- Over expression of *TaCesA4*, β-tubulin, UDP-glycosyltransferase 85A2 genes may allow researchers to increase the cellulose content further
- Upon functional validation, SNPs associated with cellulose content could be used as molecular markers to identify and design appropriate bioenergy crops
- 265 diverse wheat varieties analysed for cellulose content could probably be used as a training set for genomic selection project to predict the breeding values of wheat genotypes
- *Csl* gene enrichment sequencing of EMS mutants could be performed to further validate the physiological roles of these genes.

VIII. APPENDIX

Appendix 5.1 Table showing the percent variation of cellulose content among 288 diverse wheat lines along with their countries of origin.

Line No.	Name	Country of origin	% Cellulose
KSG001	GABO 60	Mexico	46.70472619
KSG002	NACOZARI F 76	Mexico	43.06074918
KSG003	YECORA ROJO 76	Mexico	41.57629579
KSG004	ANNAPURNA 1	Nepal, India	47.03962704
KSG005	KLEIN DRAGON	Argentina	45.1394718
KSG006	MEXIPAK65	Pakistan	43.47821804
KSG007	BLUEBIRD 15	Mexico	43.01749685
KSG008	ABU GHRAIB#3	Iraq	45.2932595
KSG009	FAISLABAD 83	Pakistan	48.25707821
KSG010	PUNJAB 88	Pakistan	47.77705132
KSG011	SAN CAYETANO S 97	Mexico	40.35340966
KSG012	BR 18	Brazil	43.04271265
KSG013	KENYA KWALE	Kenya	44.91836948
KSG014	TEMPORALERA M87	Kansas	42.11726904
KSG015	ESTANZUELA PELON 90	Uruguay	43.80735931
KSG016	CHAM 6	Syria	44.95349446
KSG017	TINAMOUII	Mexico	42.12532419
KSG018	ARIVECHI M 92	Mexico	46.22849525
KSG019	YAQUI 50	Mexico	45.86505317
KSG020	NARINO 59	Colombia	47.53416518
KSG021	PENJAMO T 62	Mexico	44.08791132
KSG022	PITIC62	Mexico	43.39717696
KSG023	CRESPO	Colombia	43.44941278
KSG024	NADADORES M 63	Mexico	43.27110991
KSG025	SONORA 64	Mexico	43.45442644
KSG026	INIA F66	Mexico	42.62369488
KSG027	BAJIO	Mexico	47.57101559
KSG028	KALYANSONA	India	44.8328799
KSG029	SAFED LERMA	India	43.50507742

KSG030	SONALIKA	India	43.66591928
KSG031	CALIDAD	Argentina	43.06248997
KSG032	UP301	India	35.06884153
KSG033	POTAM S 70	Mexico	41.05781368
KSG034	MARCOS JUAREZ INTA	Argentina	46.21094668
KSG035	TANORI F 71	Mexico	45.44585477
KSG036	ARZ	lebanon	44.90792988
KSG037	JUPATECO F 73	Mexico	43.09056956
KSG038	MAYA 74	guatamala	44.13141554
KSG039	SALAMANCA 75	Spain	44.62962963
KSG040	LIESBECK	South Africa	43.34082318
KSG041	PAVON F 76	Mexico	42.54707117
KSG042	SAKHA 8	Egypt	43.99996933
KSG043	CHIVITO	Australia	39.8763407
KSG044	HERMOSILLO M77	Mexico	36.1732
KSG045	SERI M 82	Mexico	44.1478376
KSG046	UP262	Nepal, India	43.9158003
KSG047	BAHAWALPUR 79	Pakistan	43.7416718
KSG048	SAKHA 69	Egypt	44.4301849
KSG049	HARTOG	Australia	43.740219
KSG050	PIRSABAK 85	Pakistan	44.7998216
KSG051	GONEN	Turkey	46.3563940
KSG052	RAYON F 89	Mexico	41.4933914
KSG053	NESSER	Jordan	40.5087252
KSG054	ICA YACUANQUER	Colambia	43.7154211
KSG055	TIA.1	Mexico	43.2240131
KSG056	BORLAUG M 95	Mexico	43.1436521
KSG057	PBW343	India	42.6542093
KSG058	INIFAP M 97	Chile	42.2221491
KSG059	TOBARITO M 97	Mexico	42.8289069
KSG060	GRANERO INTA	Argentina	42.1996641
KSG061	PROINTA OASIS	Argentina	45.8322652
KSG062	ITAPUA 40-OBLIGADO	Paraguay	44.8677377
KSG063	KLEIN DRAGON	Argentina	43.4954673
KSG064	BAW898	Bangladesh	41.623960
KSG065	CUMHURIYET 75	Turkey	39.2344456
KSG066	MILLALEAU INIA	Chile	42.1570745

KSG067	IAN 8-PIRAPO	Turkey	42.46548654
KSG068	PAVON	Mexico	44.79717813
KSG069	POINTA FEDERAL	Argentina	44.1163193
KSG070	SONALIKA	Punjab, India	42.04477453
KSG071	ANDES-56	Colombia	44.15091988
KSG072	SARIAB-92	Pakistan	42.16081471
KSG073	OROFEN 60	Chile	40.16797882
KSG074	LERMA ROJO 64	Mexico	41.695595
KSG075	V-17	Mexico	45.88738332
KSG076	PJ62/GB55	Mexico	46.46232439
KSG077	ZAMINDAR 80	Pakistan	36.96560847
KSG078	PAKISTAN 81	Pakistan	42.62663038
KSG079	CORDILLERA 3	Paraguay	45.06010228
KSG080	IDAHO 61M3404	Idaho	46.96628522
KSG081	IDAHO 62M9-224	Idaho	43.44005421
KSG082	LEMHI 66	Idaho	47.83103307
KSG083	64AB9405	ID	43.04125263
KSG084	TWIN	Idaho	41.84246834
KSG085	OWENS	Idaho	44.5684991
KSG086	IDO190	Idaho	44.83899583
KSG087	IDO232	Idaho	42.35110827
KSG088	COPPER	Idaho	42.15872689
KSG089	VANDAL	Idaho	45.67624932
KSG090	IDAHO 266	Idaho	41.42568531
KSG091	WHITEBIRD	Idaho	44.69516279
KSG092	FREX	Indiana	41.80197902
KSG093	II-53-521	Minnesota	47.64675168
KSG096	II-55-1	Minnesota	39.91464209
KSG097	II-58-60	Minnesota	45.50838985
KSG098	II-62-78	Minnesota	40.84739058
KSG099	MN 6616M	Minnesota	41.96671847
KSG100	WHEATON	Minnesota	43.63580016
KSG101	II-64-20	Minnesota	40.51274456
KSG102	MN 6898	Minnesota	42.2826087
KSG103	VANCE	Minnesota	43.06664091
KSG104	NORM	Minnesota	47.16934327
KSG105	VERDE	Minnesota	38.19458938

KSG106	MCVEY	Nebraska	47.62559438
KSG107	JUSTIN	North Dakota	49.31623442
KSG108	ND 202-2	North Dakota	48.38509648
KSG109	ND 271	North Dakota	49.81917336
KSG110	ND 229-1	North Dakota	49.85835538
KSG111	ND 287	North Dakota	46.91647733
KSG112	FORTUNA	North Dakota	49.50539882
KSG113	LEEDS	North Dakota	45.79200901
KSG114	ND 59-120A	North Dakota	44.30834075
KSG115	ND 407	North Dakota	45.67017079
KSG116	WALDRON	North Dakota	40.83967449
KSG117	ND 22	North Dakota	47.78958387
KSG118	ND 66	North Dakota	42.4103521
KSG119	CI014952	North Dakota	47.55488531
KSG120	CI014953	North Dakota	44.38873116
KSG121	ROLETTE	North Dakota	45.65575577
KSG122	D 6647	North Dakota	45.28231895
KSG123	ND 467	North Dakota	48.8061043
KSG124	ND 476	North Dakota	39.67881485
KSG125	ELLAR	North Dakota	47.47339873
KSG126	EDMORE	North Dakota	47.66839378
KSG127	COTEAU	North Dakota	46.89117454
KSG128	D804	North Dakota	41.7127634
KSG129	MONROE	North Dakota	44.96333195
KSG130	D7925	North Dakota	46.3238966
KSG131	ND 13-137	North Dakota	46.88940781
KSG132	AMIDON	North Dakota	47.91221172
KSG133	MUNICH	North Dakota	43.57684523
KSG134	PIERCE	North Dakota	44.6623158
KSG135	ND 2710	North Dakota	43.36096219
KSG136	STW 598874	Oklahoma	46.21235205
KSG137	YSCA-1	Oklahoma	42.4899502
KSG139	SEL. 90	Washington	46.36374266
KSG140	WA 6101	Washington	40.0853117
KSG141	WA 7175	Washington	46.06972355
KSG142	SPILLMAN	Washington	43.58764394
KSG143	ARS95 451	Washington	45.09472781

KSG144	ARS95 457	Washington	48.02306331
KSG145	EDEN	washington	43.60548617
KSG146	ALPOWA	Washington	43.43949184
KSG147	ALTURAS	Idaho	43.75510767
KSG148	CHALLIS	Montana	44.410047
KSG149	EDWALL	washington	43.36811475
KSG151	JUBILEE	Idaho	47.37820634
KSG152	VANNA	ARIZONA	47.1130596
KSG153	TARA 2002 AKA TARA	Washington	41.09099118
KSG154	SCARLET	Washington	46.83861316
KSG155	JEFFERSON	Idaho	43.34845811
KSG156	HOLLIS	Washington	46.53028667
KSG157	CALORWA	Washington	45.42869581
KSG158	ZAK	washington	50.19169639
KSG159	WAWAWAI	Washington	46.65080457
KSG160	CENTENNIAL	idaho	45.10790766
KSG161	MACON	washington	49.06372049
KSG162	LOLO	Idaho	45.97618203
KSG163	KLASIC	Nebraska	48.77983321
KSG164	IDO377S	Washington	45.83206825
KSG165	YECORA ROJO	Mexico	47.67711192
KSG166	SAXON	Colorado	43.86641714
KSG167	NEWANA	Montana	45.92721642
KSG168	URQUIE	Washington	45.29367021
KSG169	RUSHMORE	South Dakota	52.13290804
KSG170	RAMONA	California	47.25687962
KSG171	HARD FEDERATION AKA PI041079	Australia	46.42912562
KSG172	REDCHAFF	Washington	48.07375876
KSG173	SELKIRK	Canada	47.15380762
KSG176	SAUNDERS	Canada	47.2226853
KSG177	LEE	Minnesota	50.1167154
KSG178	PEAK	Idaho	46.80380726
KSG179	AKA PROBRAND 751	Nebraska	46.80622613
KSG180	WADUAL	Washington	46.88229358
KSG181	WAKANZ	Washington	50.50736472
KSG182	CANTHATCH	Canada	47.40863478
KSG183	CONLEY	North Dakota	50.24719581

KSG184	PEAK 72	Idaho	50.30414443
KSG185	PROSPUR	Minnesota	49.3063974
KSG186	KITT AKA PI518818	Minnesota	40.42727527
KSG187	WAMPUM	Washington	46.32785815
KSG188	WALLADAY	Washington	43.49644857
KSG189	PONDERA	Montana	43.94650672
KSG190	STERLING	Idaho	42.85652588
KSG191	MCKAY	Idaho	47.18660804
KSG192	WAID	Washington	45.85480139
KSG194	NORANA	Montana	45.85363639
KSG195	OLAF	North Dakota	43.50033908
KSG196	BORAH	Idaho	47.90570783
KSG197	WAVERLY	Washington	42.24863107
KSG198	TREASURE	Idaho	43.74368596
KSG199	WESTBRED 906R	Arizona	43.59406286
KSG200	WESTBRED 911	Arizona	43.16655132
KSG201	BLISS	idaho	48.88787787
KSG202	WARD	North Dakota	44.00276206
KSG203	BOUNTY 208	Colorado	46.39909736
KSG204	ANZA	California	47.21021021
KSG205	MORAN	Idaho	48.76949155
KSG206	UNION	oregon	45.29289627
KSG207	UTAC	Utah	44.46124083
KSG208	WHITE FIFE AKA PI061345	Japan	46.52450797
KSG209	WHITE MARQUIS	Minnesota	43.85544415
KSG210	SEA ISLAND	Colorado	48.51260963
KSG211	RUBY	Canada	47.76026137
KSG212	RIVAL	North Dakota	50.55001294
KSG213	LEMHI	Idaho	46.40397413
KSG214	LITTLE CLUB	Oregon	48.76414788
KSG215	MARFED	Washington	49.90889593
KSG216	TOUSE	Utah	47.86276959
KSG217	THATCHER	Minnesota	48.18024363
KSG218	SUPREME	Canada	48.16087216
KSG219	SPINKCOTA	South Dakota	45.55200744
KSG220	SONORA	Mexico	42.30245184
KSG221	GALGALOS AKA PI009872	Armenia	50.41821948

KSG222	FEDERATION 67	Idaho	47.80453371
KSG223	FEDERATION AKA PI041080	Australia	43.81425027
KSG224	REWARD	Canada	48.19479594
KSG225	RESCUE	Canada	48.79979483
KSG226	RELIANCE	Oregon	52.01883133
KSG227	REGENT	canada	48.87994127
KSG228	RED BOBS	Canada	50.13675778
KSG229	RAMONA 50	California	42.14782993
KSG230	ORFED	Washington	43.56689822
KSG231	OREGON ZIMMERMAN	Oregon	48.73593185
KSG232	ONAS 53	California	48.07730773
KSG234	MIDA	North Dakota	50.03260797
KSG235	MARQUIS	Canada	47.73903971
KSG236	PACIFIC BLUESTEM	Oregon	48.54956975
KSG237	PACIFIC BLUESTEM 37	California	48.17517535
KSG238	PILOT	North Dakota	48.5406682
KSG239	PREMIER	North Dakota	48.31727123
KSG240	ALLEN	Washington	45.64510296
KSG241	AWNED ONAS	California	46.61107559
KSG242	BAART EARLY SELECTION	California	43.7966177
KSG243	CANADIAN RED	California	44.43374264
KSG244	CADET	North Dakota	43.73795884
KSG245	BLUECHAFF	Oregon	41.85275831
KSG246	BIG CLUB	Oregon, Calafornia	44.45097118
KSG247	HARD FEDERATION (-31)	Oregon	40.96056197
KSG248	HENRY	Wisconsin	45.20997332
KSG249	HOPE	South Dakota	48.83954145
KSG250	HYBRID 63	Washington	35.98484848
KSG252	KINNEY	Oregon	41.69922384
KSG253	KENHI	Canada	47.60807328
KSG254	CERES	North Dakota	48.79524715
KSG255	WESTBRED EXPRESS	Arizona	37.88456853
KSG256	LAGODA	Russian	45.88576706
KSG257	FLOMAR	Washington	49.08835286
KSG258	HYBRID 123	Washington	36.31063321
KSG259	DICKLOW	Utah	47.27318508
KSG260	GYPSUM	Colorado	46.77553779

KSG261	HYPER	Washington	46.60738832
KSG262	IDAED	Idaho	45.03714753
KSG263	INDIAN	Idaho	45.02423314
KSG264	BAART 46	California	51.13209342
KSG265	NEW ZEALAND	Nevada	41.82661343
KSG267	PILCRAW	California	44.72485318
KSG268	RINK	Oregon	47.30452914
KSG269	SURPRISE	Vermont	46.204743
KSG270	WHITE FEDERATION	Australia	40.98373984
KSG271	BUNYIP	Australia	43.41207034
KSG272	CURRAWA	Australia	44.4541070
KSG273	WILBUR	Oregon	43.52105489
KSG274	EARLY BAART	California	45.3459533
KSG275	MAJOR	Australia	46.1615596
KSG276	LEMHI 53	Idaho	43.7793510
KSG277	SPRINGFIELD	Idaho	46.7687010
KSG278	FIELDER	Idaho	45.7000874
KSG279	FIELDWIN	Idaho	43.4039772
KSG282	SCHLANSTEDT	GermaNew York	44.5545594
KSG283	PRESTON	Canada	46.2632633
KSG284	CHINOOK	North Dakota	42.4816827
KSG285	MANITOU	Canada	44.9263994
KSG286	RED RIVER 68	California	43.170869
KSG287	ERA	Minnesota	40.3722491
KSG288	BOUNTY 309	Colorado	42.1482851
KSG289	WINSOME	Oregon	42.2270726
KSG290	AIM	ARIZONA	44.883437
KSG291	BRONZE CHIEF	USA	41.5813966
KSG292	KODIAK DWARF	USA	48.3356141
KSG293	KUBANKA	USA	43.5847415
KSG294	KAHLA	Algeria	48.8035181
KSG295	SENTRY	North Dakota	47.970231
KSG297	WELLS	North Dakota	50.1856742
KSG298	WANDELL	Washington	49.5998238
KSG299	PRODURA	Minnesota	45.8590594
KSG301	WL 444	-	45.2884188
KSG302	POMERELLE	Idaho	47.034328

Appendix 6.1 List of *CslA* subfamily genes, their protein length (amino acids) and multiple sequence alignment of amino acids showing the conserved motifs (D, D, DXD, QXXRW) specific to *Cellulose synthase like* (*Csl*) family (highlighted with red boxes).

S.No	Gene	name	with	number	of spl	ice '	variants	(CslA)	No. o	f amino	acids	(aa)
1	TRIAE	CS42	2BS	TGACv1	146583	AA04	68630.1		581	aa		
2	TRIAE			TGACv1	113418		55820.2		580	aa		
3	TRIAE			TGACv1	177473	AA05	78070.1		581	aa		
4	TRTAE			TGACv1	113300	AAOR	54190.1		579	aa		
5	TRIAE	CS42	205	TGACv1	177798	AA0.5	84795.1		881	aa		
6	TRIAE	CS42	 6BS	TGACv1	513375	AA16	39370.1		518	aa		
7	TRIAR	CS42	_6AS	TGACV1	485966		54960 1		518	aa		
8	TRIAR	CS42		-10110 VI FACV1 64	12146 AP	2112	270 1		522	aa		
9	TRIAR	CS42	-0_10 7BL	TGACV1	579090	2212 2210	03960 1		375	aa		
10	TRIAR	CS42		TGACV1	558725	 	95700 1		518	aa		
11	TRIAR	_CS42		TGACV1		 	44360 1		531	aa		
12	TRIAD	_CS42	-625-	TCACV1	487286		44500.1 69690 1		528	22		
13	TRIAD	_CS42	- GRG-	TGACVI_	_407200_ _513376	 	39390.1		528	aa 22		
11	TETAE			_IGACVI_	_JIJJ/0_ 121/6 NA	AA10	200 1		512	aa		
15	IRIAE.		_0_10	JACVI_04	E020C0	12112 7710	.290.1		512	aa		
15	TRIAL	_CS42	_/BS_	TGACVI		AAIS	45380.1		547	aa		
10	TRIAE	_CS42	_/DS_	TGACVI_	_623146_	AAZU	150070.1		545	aa		
1/	TRIAE	_CS42	_/AS_	_TGACVI_	_209190_	AAI8	09650.1		551	aa		
18	TRIAE	-CS42	_/DL_	_TGACVI_	_602617_	AAIS	628/0.1		555	aa		
19	TRIAE	_CS42	_'/AL_	_TGACv1_	_557254_	_AA17	78850.1		515	aa		
20	TRIAE	_CS42	_7BL_	_TGACv1_	_578444_	_AA18	95100.1		515	aa		
21	TRIAE	_CS42	_3dl_	_TGACv1_	_249033_	AA08	35410.1		572	aa		
22	TRIAE	_CS42	_38_1	GACv1_2	221079_A	A072	9630.1		571	aa		
23	TRIAE	_CS42	_3AL_	_TGACv1_	_197519_	AA06	66560.1		573	aa		
24	TRIAE	_CS42	_3B_1	GACv1_2	220828_A	A072	20500.1		570	aa		
25	TRIAE	CS42	_3DS	TGACv1	273022	AA09	27600.1		568	aa		
26	TRIAE	CS42	2AL	TGACv1	093375	AA02	78800.1		527	aa		
27	TRIAE	_ CS42	_2BL		129747	AA03	94630.1		528	aa		
28	TRIAE	_ CS42			160461	_ AA05	50770.1		548	aa		
29	TRIAE	_ CS42			019142	_ AA0C	61550.1		515	aa		
30	TRIAE	_ CS42			210508	_ AA06	574280.1		566	aa		
31	TRIAE	_ CS42			272005	- AA09	12960.1		570	aa		
32	TRIAE			GACv1 2	223332 A	A078	0350.1		925	aa		
TRIAE CS42 TRIAE CS42	JAS_IGACV1 BB_TGACV1 BB_TGACV1 JBS_TGACV JBS_TGACV JBS_TGACV1 JAL_TGACV JAL_TGACV1 JTGACV1 JTGACV1 JTGACV1 JTGACV1 JTGACV1 JTGACV1 JTGACV1 JCACV1 JCACV1 JCACV1 JCACV1 JCACV1 JCACV1 JCACV2 JCACV2	MLLLKIT	I AKAFD1	VVSWEYILELL	QRMNFPAHWRE	RIALLI	SSVSSAYLLKGD	PGPAILHQRGLRQG	DPLSAILF	0 		
TRIAE CS42 TRIAE CS42 TRIAE CS42 TRIAE CS42 TRIAE CS42 TRIAE CS42 TRIAE CS42 TRIAE CS42 TRIAE CS42 TRIAE CS42	LAS_TGACV 7DS_TGACV 7AS_TGACV 7BS_TGACV 5DS_TGACV 5BS_TGACV 5AS_TGACV 2BS_TGACV 2DS_TGACV									0 0 0 0 0 0 0 0 0 0 0 0		

		0
TRIAE CS42 2AS TGACV		0
TRINE_CO42_2NG_TONCV		0
IRIAE_C542_2D5_IGACV		0
TRIAE_CS42_3AS_TGACv		0
TRIAE_CS42_3B_TGACv1	PLHRMLEAAQQAGTIAPLPAGAARLRVTLYADDAIFFANPVRQEIDTIMQLLQGFGEAAGLRGNPQKSSAATLNYGSIDL	160
TRIAE CS42 3B TGACv1		0
TRIAE CS42 3DS TGACV		0
TRIAE CS42 3DS TGACV		0
TRITIL_COIL_ODD_TOHOV		0
TRIAL_CS42_SDL_IGACV		0
TRIAE_CS42_3B_TGACVI		0
TRIAE_CS42_3AL_TGACv		0
TRIAE_CS42_6BS_TGACv		0
TRIAE CS42 6AS TGACv		0
TRIAE CS42 7AL TGACV		0
TRIAE CS42 U TGACV1		0
TRIAE CS42 IL TGACV1		0
TRIAR COA2 7DI TCACH		0
TRIAL_CO42_7DL_TGACV		0
TRIAE_CS42_/AL_TGACV		0
TRIAE_CS42_/BL_TGACV		0
TRIAE_CS42_7DL_TGACv		0
TRIAE_CS42_2AL_TGACv		0
TRIAE CS42 2DL TGACv		0
TRIAE CS42 2BL TGACV		0
TRIAE CS42 1AS TGACV		0
TRIAF CS42 7DS TGACT		0
TRIAE CS42 7AS TOACV		õ
TRIAD_CORE_/AD_IGACV		0
TRIAL CO42 /BS TGACV		0
TRIAE_CS42_6DS_TGACv		U
TRIAE_CS42_6BS_TGACv		0
TRIAE_CS42_6AS_TGACv		0
TRIAE_CS42_2BS_TGACv		0
TRIAE CS42 2DS TGACV		0
TRIAE CS42 2AS TGACU		0
TRIAE CS42 2AS TOACT		0
TRIM_CO12_2A0_IGACV		0
IRIAE_C542_2D5_IGACV		0
TRIAE_CS42_3AS_TGACV		0
TRIAE_CS42_3B_TGACv1	IDVLKNFSGTRVGFPIRYLGLPLCIGRLPLCTRVGFPIRYLGWLLGKANSCIAPPLAVASHVLVRCVLSALPAFAMAVLR	240
TRIAE_CS42_3B_TGACv1		0
TRIAE CS42 3DS TGACV		0
TRIAE CS42 3DS TGACV		0
TRIAE CS42 3DL TGACV		0
TRIAE CS42 3B TGACV1		ñ
TRIAL_CO42_DD_IGACVI		0
TRIAE_CS42_SAL_TGACV		0
TRIAE_CS42_6BS_TGACV		0
TRIAE_CS42_6AS_TGACv		0
TRIAE_CS42_7AL_TGACv		0
TRIAE CS42 U TGACv1		0
TRIAE CS42 U TGACv1		0
TRIAE CS42 7BL TGACV		0
TRIAE CS42 7AL TGACV		0
TRIAE_CS42_7AL_TGACV		0
TRIAE_CS42_7AL_TGACV TRIAE_CS42_7BL_TGACV		0 0
TRIAE CS42 7AL TGACV TRIAE CS42 7BL TGACV TRIAE CS42 7BL TGACV		0 0 0
TRIAE_CS42_7AL_TGACv TRIAE_CS42_7BL_TGACv TRIAE_CS42_7DL_TGACv TRIAE_CS42_2AL_TGACv TRIAE_CS42_2AL_TGACv		0 0 0 0
TRIAE_CS42_7AL_TGACv TRIAE_CS42_7BL_TGACv TRIAE_CS42_7DL_TGACv TRIAE_CS42_7DL_TGACv TRIAE_CS42_2AL_TGACv TRIAE_CS42_2DL_TGACv		0 0 0 0
TRIAE_CS42_7AL_TGACv TRIAE_CS42_7BL_TGACv TRIAE_CS42_7DL_TGACv TRIAE_CS42_2AL_TGACv TRIAE_CS42_2DL_TGACv TRIAE_CS42_2BL_TGACv		0 0 0 0 0
TRIAE CS42 7AL TGACV TRIAE CS42 7BL TGACV TRIAE CS42 7DL TGACV TRIAE CS42 2AL TGACV TRIAE CS42 2DL TGACV TRIAE CS42 2BL TGACV TRIAE CS42 LAS TGACV		0 0 0 0 0 0
TRIAE CS42_7AL_TGACV TRIAE CS42_7BL_TGACV TRIAE CS42_7DL_TGACV TRIAE CS42_2DL_TGACV TRIAE CS42_2DL_TGACV TRIAE CS42_2BL_TGACV TRIAE CS42_1AS_TGACV TRIAE CS42_7DS_TGACV		0 0 0 0 0 0 0
TRIAE CS42 7AL TGACW TRIAE CS42 7DL TGACW TRIAE CS42 7DL TGACW TRIAE CS42 2DL TGACW TRIAE CS42 2DL TGACW TRIAE CS42 2DL TGACW TRIAE CS42 1AS TGACW TRIAE CS42 7DS TGACW TRIAE CS42 7AS TGACW		0 0 0 0 0 0 0 0 0 0
TRIAE CS42 7AL TGACV TRIAE CS42 7BL TGACV TRIAE CS42 7DL TGACV TRIAE CS42 7DL TGACV TRIAE CS42 2DL TGACV TRIAE CS42 2DL TGACV TRIAE CS42 2BL TGACV TRIAE CS42 7DS TGACV TRIAE CS42 7AS TGACV TRIAE CS42 7AS TGACV		0 0 0 0 0 0 0 0 0 0 0
TRIAE CS42_7AL_TGACV TRIAE CS42_7BL_TGACV TRIAE CS42_7DL_TGACV TRIAE CS42_2DL_TGACV TRIAE CS42_2DL_TGACV TRIAE CS42_2BL_TGACV TRIAE CS42_1AS_TGACV TRIAE CS42_7AS_TGACV TRIAE CS42_7BS_TGACV TRIAE CS42_7BS_TGACV TRIAE CS42_6DS_TGACV		0 0 0 0 0 0 0 0 0 0 0 0
TRIAE CS42 7AL TGACW TRIAE CS42 7BL TGACW TRIAE CS42 7DL TGACW TRIAE CS42 2DL TGACW TRIAE CS42 2DL TGACW TRIAE CS42 2DL TGACW TRIAE CS42 2DL TGACW TRIAE CS42 1AS TGACW TRIAE CS42 7AS TGACW TRIAE CS42 7BS TGACW TRIAE CS42 6DS TGACW TRIAE CS42 6DS TGACW		0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE CS42 7AL TGACV TRIAE CS42 7BL TGACV TRIAE CS42 7DL TGACV TRIAE CS42 7DL TGACV TRIAE CS42 2DL TGACV TRIAE CS42 2DL TGACV TRIAE CS42 2BL TGACV TRIAE CS42 7DS TGACV TRIAE CS42 7DS TGACV TRIAE CS42 7DS TGACV TRIAE CS42 6DS TGACV TRIAE CS42 6DS TGACV TRIAE CS42 6DS TGACV TRIAE CS42 6DS TGACV		0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE CS42 7AL TGACV TRIAE CS42 7BL TGACV TRIAE CS42 7DL TGACV TRIAE CS42 2DL TGACV TRIAE CS42 2DL TGACV TRIAE CS42 2DL TGACV TRIAE CS42 1AS TGACV TRIAE CS42 7DS TGACV TRIAE CS42 7AS TGACV TRIAE CS42 7AS TGACV TRIAE CS42 6DS TGACV TRIAE CS42 6DS TGACV TRIAE CS42 6AS TGACV		
TRIAE CS42 7AL TGACW TRIAE CS42 7BL TGACW TRIAE CS42 7DL TGACW TRIAE CS42 7DL TGACW TRIAE CS42 2AL TGACW TRIAE CS42 2AL TGACW TRIAE CS42 2AL TGACW TRIAE CS42 1AS TGACW TRIAE CS42 7AS TGACW TRIAE CS42 7AS TGACW TRIAE CS42 7AS TGACW TRIAE CS42 6AS TGACW TRIAE CS42 6AS TGACW TRIAE CS42 2AS TGACW		
TRIAE_CS42_7AL_TGACv TRIAE_CS42_7AL_TGACv TRIAE_CS42_7AL_TGACv TRIAE_CS42_7AL_TGACv TRIAE_CS42_2AL_TGACv TRIAE_CS42_2AL_TGACv TRIAE_CS42_1AS_TGACv TRIAE_CS42_7AS_TGACv TRIAE_CS42_7AS_TGACv TRIAE_CS42_6AS_TGACv TRIAE_CS42_6BS_TGACv TRIAE_CS42_6BS_TGACv TRIAE_CS42_2BS_TGACv TRIAE_CS42_2BS_TGACv		
TRIAE CS42 7AL TGACV TRIAE CS42 7AL TGACV TRIAE CS42 7DL TGACV TRIAE CS42 7DL TGACV TRIAE CS42 2AL TGACV TRIAE CS42 2AL TGACV TRIAE CS42 1AS TGACV TRIAE CS42 1AS TGACV TRIAE CS42 7BS TGACV TRIAE CS42 7BS TGACV TRIAE CS42 6DS TGACV TRIAE CS42 6AS TGACV TRIAE CS42 6AS TGACV TRIAE CS42 2BS TGACV TRIAE CS42 2BS TGACV TRIAE CS42 2BS TGACV TRIAE CS42 2BS TGACV		
TRIAE CS42 7AL TGACV TRIAE CS42 7BL TGACV TRIAE CS42 7DL TGACV TRIAE CS42 7DL TGACV TRIAE CS42 2AL TGACV TRIAE CS42 2AL TGACV TRIAE CS42 2AL TGACV TRIAE CS42 1AS TGACV TRIAE CS42 7DS TGACV TRIAE CS42 7DS TGACV TRIAE CS42 6DS TGACV TRIAE CS42 6DS TGACV TRIAE CS42 6DS TGACV TRIAE CS42 2DS TGACV TRIAE CS42 2DS TGACV TRIAE CS42 2DS TGACV		
TRIAE_CS42_7AL_TGACW TRIAE_CS42_7BL_TGACW TRIAE_CS42_7DL_TGACW TRIAE_CS42_2DL_TGACW TRIAE_CS42_2DL_TGACW TRIAE_CS42_2DL_TGACW TRIAE_CS42_1AS_TGACW TRIAE_CS42_7AS_TGACW TRIAE_CS42_7AS_TGACW TRIAE_CS42_6AS_TGACW TRIAE_CS42_6AS_TGACW TRIAE_CS42_2BS_TGACW TRIAE_CS42_2DS_TGACW TRIAE_CS42_2DS_TGACW TRIAE_CS42_2DS_TGACW TRIAE_CS42_2DS_TGACW		0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE_CS42_7AL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_6DS_TGACV TRIAE_CS42_6AS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2DS_TGACV	MAAWNPETHGSGAIIVGADDCETTVEDEMAAGRDANTKLFHRVANGRK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_3AS_TGACV		0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE CS42 7AL TGACV TRIAE CS42 7AL TGACV TRIAE CS42 7DL TGACV TRIAE CS42 7DL TGACV TRIAE CS42 7DL TGACV TRIAE CS42 2DL TGACV TRIAE CS42 2DL TGACV TRIAE CS42 7DS TGACV TRIAE CS42 7DS TGACV TRIAE CS42 7DS TGACV TRIAE CS42 6DS TGACV TRIAE CS42 6DS TGACV TRIAE CS42 6DS TGACV TRIAE CS42 2DS TGACV	IPKRFYKDVDKARWRFLWVHDHEVTGGRCKVNWRLVTSPVDHGGLGIPSMERFARALCLKWLWLAWTDPARPMARMGTFC	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE CS42 7AL TGACV TRIAE CS42 7BL TGACV TRIAE CS42 7DL TGACV TRIAE CS42 7DL TGACV TRIAE CS42 7DL TGACV TRIAE CS42 2AL TGACV TRIAE CS42 2AL TGACV TRIAE CS42 2AL TGACV TRIAE CS42 1AS TGACV TRIAE CS42 7AS TGACV TRIAE CS42 7BS TGACV TRIAE CS42 6AS TGACV TRIAE CS42 6AS TGACV TRIAE CS42 2AS TGACV TRIAE CS42 2AS TGACV TRIAE CS42 2AS TGACV TRIAE CS42 2AS TGACV TRIAE CS42 3AS TGACV TRIAE CS42 3B TGACV	MAAWNPETHGSGAIIVGADDCETTVEDEMAAGRDANTKLFHRVANGRK IPKRFYKDVDKARWRFLWVHDHEVTGGRCKVNWRLVTSPVDHGGLGIPSMERFARALCLRWLWLAWTDPARPWARMGTPC	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE_CS42_7AL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_6AS_TGACV TRIAE_CS42_6AS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_3AS_TGACV TRIAE_CS42_3B_TGACV1	MAAWNPETHGSGAIIVGADDCETTVEDEMAAGRDANTKLFHRVANGRK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE_CS42_7AL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2LL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_6AS_TGACV TRIAE_CS42_6AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_3AS_TGACV TRIAE_CS42_3B_TGACV1	MAAWNPETHGSGAIIVGADDCETTVEDEMAAGRDANTKLFHRVANGRK IPKRFYKDVDKARWRFLWVHDHEVTGGRCKVNWRLVTSPVDHGGLGIPSMERFARALCLRWLWLAWTDPARPWARMGTPC	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE_CS42_7AL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2LL_TGACV TRIAE_CS42_2LL_TGACV TRIAE_CS42_2BL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3DS_TGACV T	MAAWNPETHGSGAIIVGADCETTVEDEMAAGRDANTKLFHRVANGRK IPKRFYKDVDKARWRFLWVHDHEVTGGRCKVNWRLVTSPVDHGGLGIPSMERFARALCLRWLWLAWTDPARPWARMGTPC	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_7DS_TGACV TRIAE_CS42_7DS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_6DS_TGACV TRIAE_CS42_6SS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_3DS_TGACV	MAAWNPETHGSGAIIVGADCETTVEDEMAAGRDANTKLFHRVANGRK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE_CS42_7AL_TGACW TRIAE_CS42_7BL_TGACW TRIAE_CS42_7DL_TGACW TRIAE_CS42_7DL_TGACW TRIAE_CS42_2AL_TGACW TRIAE_CS42_2AL_TGACW TRIAE_CS42_2AL_TGACW TRIAE_CS42_1AS_TGACW TRIAE_CS42_1AS_TGACW TRIAE_CS42_7BS_TGACW TRIAE_CS42_7BS_TGACW TRIAE_CS42_6BS_TGACW TRIAE_CS42_2BS_TGACW TRIAE_CS42_2DS_TGACW TRIAE_CS42_2DS_TGACW TRIAE_CS42_2DS_TGACW TRIAE_CS42_2DS_TGACW TRIAE_CS42_3B_TGACW TRIAE_CS42_3B_TGACW TRIAE_CS42_3DS_TGACW TRIAE_CS42_3DS_TGACW TRIAE_CS42_3DS_TGACW TRIAE_CS42_3DS_TGACW TRIAE_CS42_3DS_TGACW TRIAE_CS42_3DS_TGACW TRIAE_CS42_3DS_TGACW TRIAE_CS42_3DS_TGACW TRIAE_CS42_3DS_TGACW	MAAWNPETHGSGAIIVGADDCETTVEDEMAAGRDANTKLFHRVANGRK IPKRFYKDVDKARWRFLWVHDHEVTGGRCKVNWRLVTSPVDHGGLGIPSMERFARALCLRWLWLAWTDPARPWARMGTPC	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE_CS42_7AL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_6AS_TGACV TRIAE_CS42_6AS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_3AS_TGACV TRIAE_CS42_3BS_TGACV TRIAE_CS42_3BS_TGACV TRIAE_CS42_3BS_TGACV TRIAE_CS42_3BS_TGACV TRIAE_CS42_3DS_TGACV	MAAWNPETHGSGAIIVGADDCETTVEDEMAAGRDANTKLFHRVANGRK	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE_CS42_7AL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2LL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_6AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DT_TGACV TRIAE_CS42_3DT_TGACV TRIAE_CS42_3DT_TGACV TRIAE_CS42_3DT_TGACV TRIAE_CS42_3DT_TGACV TRIAE_CS42_3DT_TGACV TRIAE_CS42_3DT_TGACV	IPKRFYKDVDKARWRFLWVHDHEVTGGRCKVNWRLVTSPVDHGGLGIPSMERFARALCLRWLWLAWTDPARPWARMGTPC	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE_CS42_7AL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2LL_TGACV TRIAE_CS42_2LL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DL_TGACV1 TRIAE_CS42_3DL_TGACV1 TRIAE_CS42_3AL_TGACV1 TRIAE_CS42_3AL_TGACV1 TRIAE_CS42_3AL_TGACV1 TRIAE_CS42_3AL_TGACV1 TRIAE_CS42_3AL_TGACV1 TRIAE_CS42_3AL_TGACV1 TRIAE_CS42_3AL_TGACV1 TRIAE_CS42_3AL_TGACV1 TRIAE_CS42_3AL_TGACV1 TRIAE_CS42_3AL_TGACV1 TRIAE_CS42_3AL_TGACV1 TRIAE_CS42_3AL_TGACV1 TRIAE_CS42_3AL_TGACV1 TRIAE_CS42_6AS_TGACV1	IPKRFYKDVDKARWRFLWVHDHEVTGGRCKVNWRLVTSPVDHGGLGIPSMERFARALCLRWLWLAWTDPARPWARMGTPC	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_7DS_TGACV TRIAE_CS42_7DS_TGACV TRIAE_CS42_7DS_TGACV TRIAE_CS42_6DS_TGACV TRIAE_CS42_6DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_3DC_TCACV TRIAE_CS42_3DC_TCACV	IPKRFYKDVDKARWRFLWVHDHEVTGGRCKVNWRLVTSPVDHGGLGIPSMERFARALCLRWLWLAWTDPARPWARMGTPC	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE_CS42_7AL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_6DS_TGACV TRIAE_CS42_6AS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_4C_4DCV TRIAE_CS42	IPKRFYKDVDKARWRFLWVHDHEVTGGRCKVNWRLVTSPVDHGGLGIPSMERFARALCLRWLWLAWTDPARPWARMGTPC	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE_CS42_7AL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_GBS_TGACV TRIAE_CS42_GSS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DL_TGACV TRIAE_CS42_3AL_TGACV TRIAE_CS42_3AL_TGACV TRIAE_CS42_7AL_TGACV1 TRIAE_CS42_7AL_TGACV1 TRIAE_CS42_0TGACV1 TRIAE_CS42_0TGACV1 TRIAE_CS42_0TGACV1 TRIAE_CS42_7AL_TGACV1 TRIAE_CS42_0TGACV1 TGA	MAAWNPETHGSGAIIVGADDCETTVEDEMAAGRDANTKLFHRVANGRK IPKRFYKDVDKARWRFLWVHDHEVTGGRCKVNWRLVTSPVDHGGLGIPSMERFARALCLRWLWLAWTDPARPWARMGTPC	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE_CS42_7AL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2LL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_6DS_TGACV TRIAE_CS42_6AS_TGACV TRIAE_CS42_6AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3DS_TGACV TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV1 TRIAE_CS42_3DS_TGACV1 TRIAE_CS42_3DS_TGACV1 TRIAE_CS42_3DS_TGACV1 TRIAE_CS42_3DL_TGACV1 TRIAE_CS42_3DL_TGACV1 TRIAE_CS42_4S_TGACV1 TRIAE_CS42_7EACV	Image: State of the state	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2LL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DL_TGACV TRIAE_CS42_3AL_TGACV TRIAE_CS42_GBS_TGACV TRIAE_CS42_GSTGACV TRIAE_CS42_3AL_TGACV TRIAE_CS42_0CTACV TRIAE_CS42_0TACV1 TRIAE_CS42_0TACV1 TRIAE_CS42_0TACV1 TRIAE_CS42_0TACV1 TRIAE_CS42_7AL_TGACV1 TRIAE_CS42_7A	MAAWNPETHGSGAIIVGADDCETTVEDEMAAGRDANTKLFHRVANGRK IPKRFYKDVDKARWRFLWVHDHEVTGGRCKVNWRLVTSPVDHGGLGIPSMERFARALCLRWLWLAWTDPARPWARMGTPC	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_7DS_TGACV TRIAE_CS42_7DS_TGACV TRIAE_CS42_7DS_TGACV TRIAE_CS42_6DS_TGACV TRIAE_CS42_6DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV	IPKRFYKDVDKARWRFLWVHDHEVTGGRCKVNWRLVTSPVDHGGLGIPSMERFARALCLRWLWLAWTDPARPWARMGTPC	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE_CS42_7AL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2LL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_6AS_TGACV TRIAE_CS42_6AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_3AS_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_4BS_TGACV TRIAE_CS42_4BS_TGACV1 TRIAE_CS42_4CN_TGACV1 TRIAE_CS42_4CN_TGACV1 TRIAE_CS42_7AL_TGACV1 TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV	IPKRFYKDVDKARWRFLWVHDHEVTGGRCKVNWRLVTSPVDHGGLGIPSMERFARALCLRWLWLAWTDPARPWARMGTPC	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7DS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_GBS_TGACV TRIAE_CS42_GSS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_3AS_TGACV TRIAE_CS42_3AS_TGACV TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV	MAAWNPETHGSGAIIVGADDCETTVEDEMAAGRDANTKLFHRVANGRK IPKRFYKDVDKARWRFLWVHDHEVTGGRCKVNWRLVTSPVDHGGLGIPSMERFARALCLRWLWLAWTDPARPWARMGTPC	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_7DS_TGACV TRIAE_CS42_7DS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_6AS_TGACV TRIAE_CS42_6AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_4D_TGACV TRIAE_CS42_4D_TGACV TRIAE_CS42_7AL_TGACV TRIAE	IPKRFYKDVDKARWRFLWVHDHEVTGGRCKVNWRLVTSPVDHGGLGI PSMERFARALCLRWLWLAWTDPARPWARMGTPC	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2BL_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3D_TGACV TRIAE_CS42_3D_TGACV TRIAE_CS42_3D_TGACV TRIAE_CS42_3D_TGACV TRIAE_CS42_3D_TGACV TRIAE_CS42_3D_TGACV TRIAE_CS42_3D_TGACV TRIAE_CS42_4D_TGACV1 TRIAE_CS42_0B_TGACV1 TRIAE_CS42_0B_TGACV1 TRIAE_CS42_0D_TGACV1 TRIAE_CS42_0D_TGACV1 TRIAE_CS42_0D_TGACV1 TRIAE_CS42_7AL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV	MAAWNPETHGSGAIIVGADDCETTVEDEMAAGRDANTKLFHRVANGRK IPKRFYKDVDKARWRFLWVHDHEVTGGRCKVNWRLVTSPVDHGGLGIPSMERFARALCLRWLWLAWTDPARPWARMGTPC	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_7DS_TGACV TRIAE_CS42_7DS_TGACV TRIAE_CS42_7DS_TGACV TRIAE_CS42_7DS_TGACV TRIAE_CS42_GS_TGACV TRIAE_CS42_GS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DL_TGACV TRIAE_CS42_3DL_TGACV TRIAE_CS42_7AL_TGACV	IPKRFYKDVDKARWRFLWVHDHEVTGGRCKVNWRLVTSPVDHGGLGIPSMERFARALCLRWLWLAWTDPARPWARMGTPC	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE_CS42_7AL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2LL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_6DS_TGACV TRIAE_CS42_6AS_TGACV TRIAE_CS42_6AS_TGACV TRIAE_CS42_6AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3D_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3DL_TGACV TRIAE_CS42_3DL_TGACV TRIAE_CS42_4D_TGACV1 TRIAE_CS42_4D_TGACV1 TRIAE_CS42_4D_TGACV1 TRIAE_CS42_4D_TGACV1 TRIAE_CS42_4D_TGACV1 TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV	IPKRFYKDVDKARWRFLWVHDHEVTGGRCKVNWRLVTSPVDHGGLGIPSMERFARALCLRWLWLAWTDPARFWARMGTPC	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_GBS_TGACV TRIAE_CS42_GSS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3B_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DL_TGACV TRIAE_CS42_3DL_TGACV TRIAE_CS42_3DL_TGACV TRIAE_CS42_7AL_TGACV	Image: Sector of the sector	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_7DS_TGACV TRIAE_CS42_7DS_TGACV TRIAE_CS42_7DS_TGACV TRIAE_CS42_6DS_TGACV TRIAE_CS42_6DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV	I PKRFYKDVDKARWRFLWVHDHEVTGGRCKVNWRLVTS PVDHGGLG I PSMERFARALCLRWLWLAWTDPARPWARMGTPC	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

TRIAE CS42 6DS TGACV	C	С
TRIAE CS42 6BS TGACV	C	С
TRIAE CS42 6AS TGACV	0	С
TRIAE CS42 2BS TGACV	МЕА 3	3
TRIAE CS42 2DS TGACV	МЕА 3	3
TRIAE CS42 2AS TGACV	меа 3	3
TRIAE_CS42_2AS_TGACv	МЕА З	3
TRIAE_CS42_2DS_TGACv	LKNFIPAISVEGITITDQAAKEEAFFEAYSELLGRCGSREHTLDLDYLGIESINLEDQDLVFQEEEVWKVVRDMPSDRAL 1	L28
TRIAE_CS42_3AS_TGACv	МАМАА 5	ō
TRIAE_CS42_3B_TGACv1	DDKDRALFASATTVTVGDGNRVLFWHCSWLGEQPVRQDYPNLFRRSTRKNRMADAIRDDRWIMDLRRSGAGEEVMAMAA 4	100
TRIAE_CS42_3B_TGACv1	МАМАА 5	5
TRIAE CS42 3DS TGACV	MAA 3	3
TRIAE CS42 3DS TGACV	MAATA 5	5
TRIAE CS42 3DL TGACV	MAGAGEEFMA 1	10
TRIAE CS42 3B TGACv1	MAGAGEEFMA- 1	10
TRIAE CS42 3AL TGACV	MAGAGEEFMAS 1	11
TRIAF CS42 6BS TGACT		1
TRIME_COA2_OBD_TOMEV		ń
TRIAL_CON2_OAD_IGACV	0 0	, n
TRIAE_CS42_/AL_IGACV) n
TRIAE_C342_0_IGACVI_		2
TRIAE_CS42_0_IGACVI_) n
TRIAE_CS42_/BL_TGACV		2
TRIAL_CS42_/AL_TGACV)
TRIAE_CS42_/BL_TGACV		J
TRIAE_CS42_/DL_TGACV)
TRIAE_CS42_2AL_TGACV)
TRIAE_CS42_2DL_TGACv	MEKKKRRSSIS 1	11
TRIAE_CS42_2BL_TGACV	0	J
TRIAE_CS42_1AS_TGACv	0	J
TRIAE_CS42_7DS_TGACv	Magdgegaaafaaakaew 1	18
TRIAE_CS42_7AS_TGACv	22222	24
TRIAE_CS42_7BS_TGACv	MAGDGEGAAAFAVAKAEW 1	18
TRIAE CS42 6DS TGACV	0	C
TRIAE CS42 6BS TGACV	0	C
TRIAE CS42 6AS TGACV	0	C
TRIAE CS42 2BS TGACV	AEIGGALLFALAAAAALFSAVSTGAVDFSHPLAVGGRVDFQETISWFIG	52
TRIAE CS42 2DS TGACV	AEIGGALLFALAAAAALFAAVSTGAVDFSHPPAVGGRVDF0EAISWFIG	52
TRIAE CS42 2AS TGACV	AEIGGALLFALAAAAALFAAVSTGAIDFSRPLAVGGRVDFOEAISWFIG	52
TRIAE CS42 2AS TGACV	GEIGGALLFVLAAAAAVLAAVSTGAVDFSHPPAVGGOLDFOETISWFTG	52
TRIAE CS42 2DS TGACV	GPNGFIGVFFOKAWATVKRDVMAAINKLFINNGRGFGRINOALITI.TKNHEACOTKDFRPICI.VHSTPKLASKILATRL	20.8
TRIAR CS42 3AS TOACT		R.4
TRIAE CS42 3B TGACV1		479
TRIAL CO42 3D TOACVI		27 J
TRIAE_CS42_SB_IGACVI)4 00
TRIAE_C342_3D3_IGACV		22 04
TRIAE_CS42_3DS_TGACV	WLWAEVPYVVUUWAAVAAQCAWAGQQAAALVVPTVRLUUTSLAMTMILLUKFVAAV-CIAAAAFGARPESRIRWR O	34
TRIAE_CS42_SDL_TGACV		37
TRIAE_CS42_3B_TGACVI	AVWAGLPVRVDWAAVAAQCAWAGMQARAFUVVPAIRLLVLUSLAMTVMILLEKVFVAAV-CFAAKAFGRPPERRYQWR 8	3/
TRIAE_CS42_3AL_TGACV	AAGVWAELPVRVDWAAVAAQCAWAGAQARAFIIVVPAIRLLLVISLTMTVMILLBKHFVAAV-CFAAKAFGHKPERRYQWR 9	30
TRIAE_CS42_6BS_TGACv	MDAAVGLPDAWSQVBAPWIVPLLKLAVAWCLLMSWLLFLBRMYMAVW-IVGVKLLGRRPERRYKCD 6	55
TRIAE_CS42_6AS_TGACv	MDAAVGLPDAWSQVBAPWIVPLLKLAVAWCLLMSWLLFLBRWYMAVW-IVGVKLLGRRPERRYKCD 6	55
TRIAE_CS42_7AL_TGACv	MSTLPGVWQIAAAWEQVEGPVIVPLLRASVLECLAMSAMLFABKVYMAVV-VLAVRLLGRRPERQWKWE 6	58
TRIAE_CS42_U_TGACv1_	MSTLPGAWHVAAAWEQVEGPUIVPLLRASVLUCLAMSEMLFAEKVYMAVU-VLAVRLLGRRPERQYQWE 6	58
TRIAE_CS42_U_TGACv1_	MAAALLPGTRITFSGAWQQV <mark>B</mark> GPVIVPLLRASVL <mark>B</mark> CVAMS <mark>A</mark> MLLA <mark>BKV</mark> YMAVV-VLALRLLGRRPELQYRWE 7	71
TRIAE_CS42_7BL_TGACv	MSTLPRVWQIAAAWEQVEGPUIVPLLRVSVLCLAMSEMLFAEKVYMAVV-VLAVRLLGRRPEQQYRWE 6	58
TRIAE_CS42_7AL_TGACv	MEAAEQIAVVWKQV <mark>R</mark> GPVIAPLLRASVM <mark>W</mark> CLAMCVILFV <mark>E</mark> KWYMAVV-IVAMRLIGRHPERQWRWE 6	55
TRIAE_CS42_7BL_TGACv	MEAAEQIAVVWKQV <mark>R</mark> GP <mark>W</mark> IVPLLRASVM <mark>W</mark> CLAMCWILFVEKWYMAVW-IVAMRLIGRRPERQWRWE 6	55
TRIAE_CS42_7DL_TGACv	MEAAEQIAVVWKQV <mark>R</mark> GP <mark>W</mark> IVPLLRASVM <mark>W</mark> CLAMCWILFVEKWYMAVW-IVAMRLIGRRPERQWRWE 6	55
TRIAE_CS42_2AL_TGACv	MKGVSMLTMARAAWAAVRHAWVVPLLQLAVYCAAMSIMLFABRIYMGLV-VAALWLRRRRQRRNPGR 6	58
TRIAE_CS42_2DL_TGACv	FLLSFGGGRRRMKGVSMLTMARAAWAVVRYAWVVPLLQLAVYCAAMSIMLFAERIYMGLV-VAALWLRRRRQRRSPSR 9	90
TRIAE_CS42_2BL_TGACv	mrgvsmltmaraawaav r ya w vvpllqlavy m caams m mlfa bri ymgl w -vaalwlrrrrrqrrnpsr 6	58
TRIAE_CS42_1AS_TGACv	MSMLPMARAAWLVLAYAWVVPLLQLAIYHCVVMSHMLFABRHYMGLW-VAVLWLYRRCRNRNQRNK 6	<u> 5</u> 5
TRIAE_CS42 7DS TGACv	LDGSGGLPLLRWWRASGGGELLGRWDAVRAGVAPALAAVSGCLAMSEMLLABAVFMAAB-SLVRRRPERRYSAG 9	эз
TRIAE CS42 7AS TGACV	LGGSGGLPLLRWWRASGGGELLRGWDAVRAGWAPALAAVSGCLAMSMILLARAWFMAAR-SLVRRRPERRYSAG 9	39
TRIAE CS42 7BS TGACV	LAGSGGLPLLRWWRASGGGELLRGWDAVRAGWAPALAAVSGCLAMSEMLLABAWFMAAB-SLVRRPERRYSAG 9	эз
TRIAE CS42 6DS TGACV	MAPLGADAAAAAWAAVAARAAWAAAWAAAAWAAAAWAAAAWAAAA	58
TRIAE CS42 6BS TGACV	MAPLGADAAAAAWAAVARARUSPALTAAVWACLAMSAMLLLAAGCMSLASIVAVRLLRLPORRFKWE	58
TRIAE CS42 6AS TGACV	MAPLSAGAAAAAWAAVRARAVAAAAWACLAMSMULLMAACMSLVAVRLLRLRPERRFKWE	68
TRIAE CS42 2BS TGACV	- IFDGSSSSSAAGGVSLAEVYELWVRVRCRUIAPALOVAVWCMVMSWMLVVRAUVNCVG-SLGVKAVCWRPEWRFKWE 1	130
TRIAE CS42 2DS TGACV	-VEDGSSSSSSAAGGVSLAEVYELWVRVEGRWTAPALOVAVWECMVMSWMLVVEAWYNCVW-SLGVKAVGWRPEWRFKWE 1	130
TRIAE CS42 2AS TGACV	-VEDGSSSSS-AAGGVSLAEVYELWVRVEGRWTAPALOVAVWRCMVMSWMLVVPAL YNCVV-SLGVKAVGWRPEWRFKWE 1	129
TRIAE CS42 2AS TGACY		129
TRIME_CO42_2ND_TOMOV		287
	or coordinate the second of th	-07
TRIAE CS42 3AS TOACT		160
TRIAL CO42 SAS IGACV		526
TRIAL CON2_30 TOACVI	TITASACKTCCDDEEDCTVWVCCCCCCBAIEDVMTVTIMTVEDVVKVCTCAACAIEWDCDDWVTOVTDPPPDUVVDT	161
TRIAL CONT DE TORCVI	PIASSOKTCODEFICITIVI/CCCCCC2 EPUMI VOIDMVNEDEVIVVUCICAACAL FWEDEDMVIAU DOGEDMIANT	162
TRIAL COM2_SUS_TGACV	I IMMONORACGODEEDCIVVGGGGGGGGGFFVWUVQCFWINEREVIKVGTGAAGALEWFSDRWVLQVLDDSTDPVVKELVI	161
TRIAL COM2 DDS TGACV		165
TRIAL CS42_3DL TGACV	riaggaaaaagdee - aglvggggsaaf rvilvgirminekeviklsigaacalewrsdrvvigvlddstdpavkblu i	160
TRIAE_CS42_3B_TGACV1	PIAGGAAAAAAGUEE-AGVGGGG-SAAFPVULVQIPMINEREVYKLSIGAACALEWPAERVVIQVLDDSTDPVVKDLU 1	103 105
TRIAE_CS42_3AL_TGACV	PIABACAIGGVDEE-ASVGGGSSAFYWLVQIPMYNEKEVYKLSIGAACALEWPSDRVVIQVLDDSTDPAVKDLW 1	100
TRIAE_CS42_6BS_TGACV	PICEDDDPELGSAAFPIMLVQ1PMFNEREVYQLSIGAVCGLSWPSDRLVVQVLDDSTDPLIKEM 1	130
TRIAE_CS42_6AS_TGACV	PICEDDDPELGSAAFPVMLVQIPMFNEREVYQLSIGAVCGLSWPSDRLVVQVLDDSTDPLVKEMM 1	130
TRIAE_CS42_7AL_TGACv	PVGE-DDPELGSAAYPMULVQIPMYNEREVYQLSIGAACGLSWPSDRIVVQVLDDSTDPVIKELU 1	132
TRIAE_CS42_U_TGACv1_	PMGD-DDPELGSAAYPMULVQIPMYNEREVYQLSIGAACGLSWPSDRIVVQVLDDSTDPVIKELU 1	132
TRIAE_CS42_U_TGACv1_	PMRDGDDPELGSAAYPMULVQIPMYNEREVYQLSIGAACGLSWPSDRIIVQVLDDSTDPVVKELU 1	136
TRIAE_CS42_7BL_TGACV	PVGDGNDPELGSAAYPMULVQIPMYNEREVYQLSIGAACGLSWPSDRIIVQVLDDSTDPVIKELU 1	133
TRIAE_CS42_7AL_TGACv	Plrd-ddpelgnaaypmulvqipmynerevykksigaacglswpsdriviqvlddstdpaikelu 1	129
TRIAE_CS42_7BL_TGACv	PLRD-DDPELGNAAYPMULVQIPMYNEREVYKKSIGAVCGLSWPSDRIVIQVLDDSTEPAIKELU 1	129
TRIAE_CS42_7DL_TGACv	Plrd-ddpelgnaaypmulvqipmynerevykksigaacglswpsdriviqvlddstdpaikelu 1	129
TRIAE_CS42_2AL_TGACv	NKGGDDDVGDLESGAAEDLPVWLVQIPMFNEKQVYRLSIGAACGLWWPADKLVIQVLDDSTDAGIRAMW 1	137
TRIAE CS42 2DL TGACV	NKGGDDDDLESGAAEDLPLVLVQIPMFNEKQVYRLSIGAACGLWWPADKLVIQVLDDSTDAGIRAMM 1	157

TRIAR CS42 2BL TGACT	NKGDDDGGAGDLF	
TRIME_CO42_IDE_TONCV	CDDDNI ESD	DADD DAME I VOT DMENEKOVEDI STCAACCI NWDADKI VIOVI DDSTDACIDSI 125
TRIAL_CONZ_IAS_IGACV	BLCAODCEDE	EDCLICYDMULUOIDMYNEDEUVYKICICAACCICWDCDDUUUOUDDCEDDEUVDIW 160
TRIAE_CS42_7DS_IGACV	FLGAQDGEDE	ERGLIGIERWILVQIERINEREVIKISIGAACGLSWESDRVIVQVLDDSIDFIIKDLV 100
TRIAE_CS42_7AS_IGACV	PLGAQDGEDED	ERGLIGIEMVLVQIEMINEREVIKLSIGAACGLSWESDRVIVQVLDDSIDFIIKDLV 100
TRIAE_C342_7B3_IGACV	PLGAQUGEDEDE	ERGELGIPHVLVQIPHINEREVIKLGIGNGSIBWPSDRVIVQVLDDSIDFIIKDLV 102
TRIAE_CS42_6DS_TGACV	PMAGALEGGEADVEDPPA	.SAGRREFPMVLVQIPMINEREVIRLSIGAVCALTWPPDRIIIQVLDDSTDPIIRELV 143
TRIAE_CS42_6BS_TGACV	PMPGALPGAEADAEDPPG	RREFPMVLVQIPMINEKEVIKLSIGAVCALTWPPDRIIIQVLDDSTDPIIKELV 140
TRIAE_CS42_6AS_TGACV	PMTGALEGGEADVEDPAG	RREFPMWLVQIPMYNEKEVYKLSIGAVCALTWPPDRIIIQVLDDSTDPIIKELM 140
TRIAE_CS42_2BS_TGACV	PLAGDDEEKGG	AHYPMVLVQIPMYNELEVYKLSIGAACELQWPKDRIIVQVLDDSTDPFIKNLV 194
TRIAE_CS42_2DS_TGACv	PLAGDDEEKGG	AHYPMWLVQIPMYNELEVYKLSIGAACELQWPKDRIIVQVLDDSTDPFIKNLW 194
TRIAE_CS42_2AS_TGACv	PLAGDDEEKGG	AHYPVWLVQIPMYNELEVYKLSIGAACELQWPKDRIIVQVLDDSTDPFIKNLW 193
TRIAE_CS42_2AS_TGACV	PLAGD-EEKGS	AHYPMVLVQ1PMYNELEVYKLSIGAACELKWPKDRMIVQVLDNSTDPLIKNLV 192
TRIAE_CS42_2DS_TGACv	TLLTTASSRVV	VNGCVGKKFMHACGLRQGDSISPLLFVIAMDVLSAMILKARETNAVSKIPGCA 351
TRIAE_CS42_3AS_TGACV	KINCQRWKSKGVNIRYEVRQNRF	.GMKAGALMEGLMRDYMRB 201
TRIAE_CS42_3B_TGACVI	KINCQRWKSKGVNIRYEVRQNRF	.GWKAGMLWEGLIRDYWRB 56/
TRIAE_CS42_3B_TGACVI	KTECQRWKGKGVNIRYEVRGNRF	.GMKAGALWQGLMRDYWRB 205
TRIAE_CS42_3DS_TGACv	KT CQRWKGKGVNIRYEVRGNRF	.GWKAGMLWQGLMRDYWR© 203
TRIAE_CS42_3DS_TGACv	KI CQRWKSKGVNIRYEVRENRF	.GWKAGMLWQGLMRDYWR© 205
TRIAE_CS42_3DL_TGACv	EICQRWKGKGVNIKYEVRGNRF	.GMKAGALMEGLKHDYMQE 206
TRIAE_CS42_3B_TGACv1	EIECQRWKGKGVNIKYEVRGNRF	.GMKAGALMEGLKHDYVQE 204
TRIAE_CS42_3AL_TGACv	EIBCQRWKGKGVNIKYEVRGNRF	.GWKAGMLWEGLKHDYWQE 206
TRIAE_CS42_6BS_TGACv	RMBCERWAHKGINITYQIREDRF	.GWKAGMLWAGMKHGYWRD 171
TRIAE_CS42_6AS_TGACv	RMBCERWAHKGINITYQIREDRF	.GWKAGMLWAGMKHGYWRD 171
TRIAE_CS42_7AL_TGACv	QV CRRWARKGVNIKYEIRDNRF	.GWKAGMLWEGMKHGYWKD 173
TRIAE_CS42_U_TGACv1_	RVECRRWARKGVNIKYEIRDNRF	.GMKAGALMEGMKHGYMKD 173
TRIAE_CS42_U_TGACv1_	QVECQRWARKGVNIKYETRNNRF	.GMKAGALMEAMKHGYVKD 177
TRIAE_CS42_7BL_TGACv	QV CRRWARKGVNIKYEIRDNRF	GMKAGALKEGMKHGYVKD 174
TRIAE_CS42_7AL_TGACv	QA CHRWANKGVNIKYEIRDNRF	.GMKAGALMEGMKHGYMKD 170
TRIAE_CS42_7BL_TGACv	QVECQRWANKGVNIKYEIRDNRF	GYKAGALKEGMKHGYVKD 170
TRIAE_CS42_7DL_TGACv	QV CQRWANKGVNIKYEIRDNRF	.GMKAG <mark>A</mark> lkegmkhgYVKD 170
TRIAE_CS42_2AL_TGACV	EASCRRWAGKGVHIRYENRSNRS	.GWKAGAMREGLKKGYAKD 178
TRIAE_CS42_2DL_TGACv	EABCRRWAGKGVQIRYENRSNRS	GWKAGMMREGLKKGYAKD 198
TRIAE_CS42_2BL_TGACv	EASCRRWAGKGVQIRYENRSNRS	GYKAGAMREGLKKGYARD 179
TRIAE_CS42_1AS_TGACv	EABCRRWAGKGVHIRYENRSNRS	GWKAGMMRDGLKKQYVKD 169
TRIAE_CS42_7DS_TGACv	ELECKIWAKKGKNVKYEVRNNRE	.G w kag w lkegmlhaywoo 201
TRIAE_CS42_7AS_TGACv	ELECKIWAKKGKNVKYEVRNNRE	.GYKAGALKEGMLHAYVQQ 207
TRIAE_CS42_7BS_TGACv	ELECKIWAKKGKNVKYEVRNNRE	.GMKAGALMEGMLHAYNQQ 203
TRIAE_CS42_6DS_TGACv	ELCQEWASKKIDIKYEVRNNRF	.GMKAGALMKGMEHVYAQQ 184
TRIAE_CS42_6BS_TGACv	ELCQEWASKKIDIKYEVRNNRF	.GMKAGALMKGMEHVYAQO 181
TRIAE_CS42_6AS_TGACv	ELCQEWASKKIDIKYEVRNNRF	.GMKAGMLKKGMEHVYAQQ 181
TRIAE_CS42_2BS_TGACv	ELECESWAVKGLNIKYATRSSRF	.G_KAGAL&KGMECDYAKQ 235
TRIAE_CS42_2DS_TGACv	ELECESWAVKGLNIKYATRSSRF	.G . KAG A L K KGMECDY A K Q 235
TRIAE_CS42_2AS_TGACv	ELECESWSVKGLNIKYATRSSRF	.G.KAGALMKGMEYDYAKO 234
TRIAE_CS42_2AS_TGACv	ELECETWVTKGLNIKYAPRSGQF	.GEKAGALMKGMECDYARO 233
TRIAE_CS42_2DS_TGACv	PIORLSLYVDDVVMFIKPSWTDI	WEVQEALEVFGEASGLKVNFSKSSAVMIRSEEEEEVLVRKAMPWKMETFPIKYLCLO 431
TRIAE_CS42_3AS_TGACv	CEFIAMFD7DDQPESDFILRTV	FLVHN 229
TRIAE_CS42_3B_TGACv1	CEFIAMFD7DDQPESDFILRTV	FLVHN 595
TRIAE_CS42_3B_TGACv1	CEFIAMFD7DDQPESDFILRTV	FLVHN 233
TRIAE_CS42_3DS_TGACV	CKF1AMFDAD OPESDFLLRTV	FLVHN 231
TRIAE_CS42_3DS_TGACV	CEFIAMEDAD OPESDELLRIV	FLVHN 233
TRIAE_CS42_3DL_TGACV	CEFIAMFDAD OPESDFLLRTV	FLVHN 234
TRIAE_CS42_3B_TGACVI	CEFIAMFDAD OPESDFLLRTV	PLVHN 232
TRIAE_CS42_3AL_TGACV	CEFIAMFDAD OPESDFLLRTV	FLVHN 234
TRIAE_CS42_6BS_TGACV	CEYMVIFDAD QPDPDFTHRTI	YLHHN 195
TRIAE_CS42_6AS_TGACV	CEYMVIFDAD OPDPDF HRTI	YLHHN 195
TRIAE_CS42_/AL_TGACV	CDLVAIFDAD OPEPDFI WRAV	FLVHN 201
TRIAE_CS42_U_TGACVI_	CDLVAIFDAD WRAV	FLVHN 195
TRIAE_CS42_0_TGACVI_	CDLVAIFDFD, QPEPDFLSRSV	FLVHN203
TRIAE_CS42_7BL_TGACV	CDLVAIFDFD OPEPDFI SRSV	F LVHN 202
TRIAD_COM2_/AL_TGACV	CDEWUIED BLODEDDY CEAM	16° I. V H N = = = = = = = = = = = = = = = = = =
TRIAL COM2 /BL TGAUV		FLVHN 195
INIAL COME IDE IGACV	CDEWLEDZD ODEDDY CRAM	FLUHN 196 FLIHN 196
TOTAE COA2 2NT TOTAC-	CDFVVIFDADIQPEPDYLSRAM	FLVHN
TRIAE_CS42_2AL_TGACV	CDFVVIFD7 D OPEPDYLSRAM CELVAVFD7 D OPDADFLRRTV CELVAVFD7 D OPDADFLRRTV	FLVHN
TRIAE CS42 2AL TGACV TRIAE CS42 2DL TGACV TRIAE CS42 2DL TGACV	CDFVVIFDA D QPEPDY SRAM CELVAVFDA D QPDADF RRTV CELVAVFDA D QPDADF RRTV CELVAVFDA D QPDADF RRTV	FLVHN
TRIAE_CS42_2AL_TGACv TRIAE_CS42_2DL_TGACv TRIAE_CS42_2BL_TGACv TRIAE_CS42_1AS_TCACv	CDFVVIFD2 D QPEPDYLSRAM CELVAVFD2 D QPDADFLRTV CELVAVFD2 D QPDADFLRTV CELVAVFD2 D QPDADFLRTV CELVAVFD2 D QPDADFLRTV	FLVHN
TRIAE CS42 2AL TGACV TRIAE CS42 2DL TGACV TRIAE CS42 2BL TGACV TRIAE CS42 1AS TGACV TRIAE CS42 1AS TGACV	CDFVVIFD2 D OPEPDY SRAM CELVAVFD2 D OPDADF RRTV CELVAVFD2 D OPDADF RRTV CELVAVFD2 D OPDADF RRTV CEFVAVFD2 D OPDADF RRTV CEFVAVFD2 D OPDADF RRTV	FLVHN 196 FLIHN 196 FLIHN 196 VLQAD 206 VLQAD 207 VLQAD 207 VLQAD 207 VLQAD 207 VLQAD 207 VLBAD 197 VLBAD 207
TRIAE CS42 2AL TGACV TRIAE CS42 2DL TGACV TRIAE CS42 2BL TGACV TRIAE CS42 1AS TGACV TRIAE CS42 1AS TGACV TRIAE CS42 7DS TGACV TRIAE CS42 7DS TGACV	CDFVVIFDA D OPEPDY SRAM CELVAVFDA D OPDADF BRTV CELVAVFDA D OPDADF BRTV CELVAVFDA D OPDADF BRTV CEFVAVFDA D OPDADF BRTV CDFLAVFDA D OPEPDF WATT CDFLAVFDA D OPEPDF WATT	FLVHN 196 FLIHN 196 FLIHN 196 VLQAD 206 VLQAD 206 VLQAD 207 VLQAD 202 VLQAD 202 VLAN 225
TRIAE_CS42_2AL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2BL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_7AS_TGACV	CDFVVIFDA D QPEPDY SRAM CELVAVFDA D QPDADF RRTV CELVAVFDA D QPDADF RRTV CELVAVFDA D QPDADF RRTV CELVAVFDA D QPDADF RRTV CDFLAVFDA D QPEDF MRTI CDFLAVFDA D QPEPDF MRTI CDFLAVFDA D QPEPDF MRTI	FLVHN 19 FLTHN 19 FLTHN 19 VLQAD 20 VLQAD 20 VLQAD 19 VLQAD 20 VLQAD 19 VLQAD 19 YLARN 22 YLARN 23 YLARN 23
TRIAE_CS42_2AL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2BL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_7BS_TGACV	CDFVVIFD2 D OPEPDY SRAM CELVAVFD2 D OPDADF DRTV CELVAVFD2 D OPDADF DRTV CELVAVFD2 D OPDADF DRTV CEFVAVFD2 D OPDADF DRTV CDFLAVFD2 D OPDADF DRTV CDFLAVFD2 D OPEPDF MRTI CDFLAVFD2 D OPEPDF MRTI CDFLAVFD2 D OPEPDF MRTI	FLVHN 19 FLIHN 19 FLIHN 19 VLQAD 20 VLQAD 20 VLQAD 207 VLQAD 207 VLQAD 207 VLRN 207 VLARN 225 VLSRN 231 VLVN 213
TRIAE CS42_2AL_TGACW TRIAE CS42_2DL_TGACW TRIAE CS42_2DL_TGACW TRIAE CS42_1AS_TGACW TRIAE CS42_1AS_TGACW TRIAE CS42_7AS_TGACW TRIAE CS42_7BS_TGACW TRIAE CS42_6DS_TGACW TRIAE CS42_6DS_TGACW	CDFVVIFD2 D OPEPDY SRAM CELVAVFD2 D OPDADF BRTV CELVAVFD2 D OPDADF BRTV CELVAVFD2 D OPDADF BRTV CEFVAVFD2 D OPDADF BRTV CDFLAVFD2 D OPEPDF MRTI CDFLAVFD2 D OPEPDF MRTI CDFLAVFD2 D OPEPDF MRTI CEFVAIFD2 D OPEDF MRTI CEFVAIFD2 D OPEDF MRTI	FLVHN 196 FLIHN 196 FLIHN 196 VLQAD 206 VLQAD 207 YLQAN 207 YLQAN 207 YLARN 207 YLARN 231 FLVHN 202
TRIAE_CS42_2AL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_TAS_TGACV TRIAE_CS42_TAS_TGACV TRIAE_CS42_6DS_TGACV TRIAE_CS42_6DS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_6BS_TGACV	CDFVVIFDA D OPEPDY SRAM CELVAVFDA D OPDADF BRTV CELVAVFDA D OPDADF BRTV CELVAVFDA D OPDADF BRTV CEFVAVFDA D OPDADF BRTV CDFLAVFDA D OPEDF MRTI CDFLAVFDA D OPEDF MRTI CDFLAVFDA D OPEDF MRTI CEFVAIFDA D OPESDF LKTI CEFVAIFDA D OPESDF LKTI	F LVHN 19 FL HN 19 FL HN 19 FL HN 19 VLQAD 206 VLQAD 207 VLQAD 207 VLAAD 197 YLARN 225 YLARN 235 FLVHN 212 FLVHN 212 FLVHN 209
TRIAE_CS42_2AL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2BL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_6AS_TGACV	CDFVVIFD2 0 OPEPDY SRAM CELVAVFD2 0 OPDADF BRTV CELVAVFD2 0 OPDADF BRTV CELVAVFD2 0 OPDADF BRTV CEFVAVFD2 0 OPDADF BRTV CDFLAVFD2 0 OPEPDF MRTI CDFLAVFD2 0 OPEPDF MRTI CDFLAVFD2 0 OPEPDF MRTI CEFVAIFD2 0 OPEDF LKTI CEFVAIFD2 0 OPEDF LKTI CEFVAIFD2 0 OPEDF LKTI	FLVHN 19 FLIHN 19 FLIHN 19 VLQAD 206 VLQAD 207 VLQAD 207 VLQAD 207 VLQAD 207 VLARN 225 YLARN 225 YLSRN 231 FLVHN 212 FLVHN 212 FLVHN 206 FLVHN 207 FLVHN 208 FLVHN 209 FLVHN 209 FLVHN 209
TRIAE CS42_2AL TGACW TRIAE CS42_2DL TGACW TRIAE CS42_2DL TGACW TRIAE CS42_1AS TGACW TRIAE CS42_1AS TGACW TRIAE CS42_7AS TGACW TRIAE CS42_7BS TGACW TRIAE CS42_6DS TGACW TRIAE CS42_6DS TGACW TRIAE CS42_6AS TGACW TRIAE CS42_2BS TGACW TRIAE CS42_2DS TGACW	CDFVVIFD2 D QPEPDY SRAM CELVAVFD2 D QPDADF PRTV CELVAVFD2 D QPDADF PRTV CELVAVFD2 D QPDADF PRTV CEFVAVFD2 D QPDADF PRTV CDFLAVFD2 D QPEPDF PRTTI CDFLAVFD2 D QPEPDF PRTTI CDFLAVFD2 D QPEPDF PRTTI CEFVAIFD2 D QPESDF LKTI CEFVAIFD2 D QPESDF LKTI CEFVAIFD2 D QPESDF LKTI CEFVAIFD2 D QPEPDF LKTI	FLVHN 196 FLIHN 197 FLIHN 196 VLQAD 206 VLQAD 226 VLQAD 207 VLQAD 226 VLQAD 207 VLQAN 207 YLARN 225 YLARN 225 YLARN 231 FLVHN 201 FLVHN 201 FLVHN 209 FLVHN 203 FLVHN 203 FLVHN 203 FLVHN 203 FLVHN 204 FEVHN 263 FEVHN 263
TRIAE_CS42_2AL_TGACW TRIAE_CS42_2DL_TGACW TRIAE_CS42_2DL_TGACW TRIAE_CS42_1AS_TGACW TRIAE_CS42_1AS_TGACW TRIAE_CS42_7AS_TGACW TRIAE_CS42_7AS_TGACW TRIAE_CS42_6DS_TGACW TRIAE_CS42_6BS_TGACW TRIAE_CS42_2DS_TGACW TRIAE_CS42_2DS_TGACW TRIAE_CS42_2DS_TGACW	CDFVVIFD2 D QPEPDY SRAM CELVAVFD2 D QPDADF RETY CELVAVFD2 D QPDADF RETY CELVAVFD2 D QPDADF RETY CEFVAVFD2 D QPDADF RETY CDFLAVFD2 D QPEDF NETI CDFLAVFD2 D QPEDF NETI CDFLAVFD2 D QPEDF NETI CEFVAIFD2 D QPEDF NETI CEFVAIFD2 D QPESDF LETI CEFVAIFD2 D QPESDF LETI CEFVAIFD2 D QPESDF LETI CEYVAIFD2 D QPEDF LETY CEYVAIFD2 D QPEDF LETY	FU-NN 198 FL LHN 199 FL HN 199 FL HN 199 VLQAD 206 VLQAD 207 VLQAD 207 VLQAD 207 VLAAD 197 YLARN 235 YLARN 235 YLSRN 231 FLVHN 212 FLVHN 209 FLVHN 209 FFVHN 203 FFVHN 203 FFVHN 203 FFVHN 203
TRIAE CS42_2AL TGACW TRIAE CS42_2DL TGACW TRIAE CS42_2DL TGACW TRIAE CS42_2AL TGACW TRIAE CS42_1AS TGACW TRIAE CS42_7AS TGACW TRIAE CS42_7AS TGACW TRIAE CS42_6AS TGACW TRIAE CS42_6AS TGACW TRIAE CS42_2BS TGACW TRIAE CS42_2DS TGACW TRIAE CS42_2AS TGACW TRIAE CS42_2AS TGACW	CDFVVIFD2 0 OPEPDY SRAM CELVAVFD2 0 OPDADF BRTV CELVAVFD2 0 OPDADF BRTV CELVAVFD2 0 OPDADF BRTV CEFVAVFD2 0 OPDADF BRTV CDFLAVFD2 0 OPEPDF MRTI CDFLAVFD2 0 OPEPDF MRTI CDFLAVFD2 0 OPEDF MRTI CEFVAIFD2 0 OPEDF LRTI CEFVAIFD2 0 OPEDF LRTI CEYVAIFD2 0 OPEDF LRTV CEYVAIFD2 0 OPEDF LRTV CEYVAIFD2 0 OPEDF LRTV CEYVAIFD2 0 OPEDF LRTV	FLVHN 19 FLIHN 19 FLIHN 19 FLIHN 19 VLQAD 20 VLQAD 20 VLQAD 20 VLQAD 207 VLARN 20 YLARN 23 YLARN 23 FLVHN 21 FLVHN 20 FLVHN 21 FLVHN 20 FVHN 20
TRIAE CS42_2AL TGACW TRIAE CS42_2DL TGACW TRIAE CS42_2DL TGACW TRIAE CS42_1AS TGACW TRIAE CS42_1AS TGACW TRIAE CS42_7AS TGACW TRIAE CS42_7AS TGACW TRIAE CS42_7AS TGACW TRIAE CS42_6AS TGACW TRIAE CS42_6AS TGACW TRIAE CS42_2AS TGACW TRIAE CS42_2AS TGACW TRIAE CS42_2AS TGACW TRIAE CS42_2AS TGACW TRIAE CS42_2AS TGACW	CDFVVIFD2 D QPEPDY SRAM CELVAVFD2 D QPDADF PRTV CELVAVFD2 D QPDADF PRTV CELVAVFD2 D QPDADF PRTV CEFVAVFD2 D QPDADF PRTV CDFLAVFD2 D QPEPDF WRTI CDFLAVFD2 D QPEPDF WRTI CDFLAVFD2 D QPEPDF WRTI CEFVAIFD2 D QPESDF LKTI CEFVAIFD2 D QPESDF LKTI CEFVAIFD2 D QPESDF LKTI CEYVAIFD2 D QPEPDF LRTV CEYVAIFD2 D QPEPDF LRTV	FLVHN 196 FLIHN 196 FLIHN 196 VLQAD 206 VLQAD 207 VLARN 228 YLARN 223 FLVHN 212 FLVHN 205 FLVHN 205 FLVHN 205 FLVHN 205 FLVHN 205 FFVHN 263 FFVHN 263 FFVHN 262 FFVHN 262 FFVHN 263 FFVHN 261 GWORGPVTRPGRLPLVNOVVRARPIHHLIVARAPKRALDRVDKGCRAFFWAGSEEI0 261
TRIAE CS42_2AL_TGACW TRIAE CS42_2DL_TGACW TRIAE CS42_2DL_TGACW TRIAE CS42_2AL_TGACW TRIAE CS42_1AS_TGACW TRIAE CS42_7AS_TGACW TRIAE CS42_7BS_TGACW TRIAE CS42_6BS_TGACW TRIAE CS42_6BS_TGACW TRIAE CS42_2DS_TGACW TRIAE CS42_2DS_TGACW TRIAE CS42_2AS_TGACW TRIAE CS42_2AS_TGACW TRIAE_CS42_2DS_TGACW	CDFVVIFD2 D QPEPDY SRAM CELVAVFD2 D QPDADF PRTV CELVAVFD2 D QPDADF PRTV CELVAVFD2 D QPDADF PRTV CEVAVFD2 D QPDADF PRTV CDFLAVFD2 D QPEPDF WRTI CDFLAVFD2 D QPEPDF WRTI CDFLAVFD2 D QPEPDF WRTI CEFVAIFD2 D QPEDF LRTI CEFVAIFD2 D QPEDF LRTI CEFVAIFD2 D QPEDF LRTV CEYVAIFD2 D QPEPDF LRTV	FLVHN 19 FLTHN 19 FLTHN 19 FLTHN 19 VLQAD 20 YLARN 22 YLARN 23 YLSRN 212 FLVHN 212 FLVHN 20 FLVHN 212 FLVHN 20 FVHN 263 FFVHN 261 GWQRGPVTRFGRLPLVNQVVRARPIHHLIVAEAPKRALDRVDKGCRAFFWAGSEEI 511
TRIAE_CS42_2AL_TGACW TRIAE_CS42_2DL_TGACW TRIAE_CS42_2DL_TGACW TRIAE_CS42_1AS_TGACW TRIAE_CS42_1AS_TGACW TRIAE_CS42_7AS_TGACW TRIAE_CS42_7AS_TGACW TRIAE_CS42_6DS_TGACW TRIAE_CS42_6BS_TGACW TRIAE_CS42_2DS_TGACW TRIAE_CS42_2DS_TGACW TRIAE_CS42_2AS_TGACW TRIAE_CS42_2DS_TGACW TRIAE_CS42_2DS_TGACW TRIAE_CS42_2DS_TGACW TRIAE_CS42_2DS_TGACW	CDFVVIFD2 0 OPEPDY SRAM CELVAVFD2 0 OPDAF BRTY CELVAVFD2 0 OPDAF BRTY CELVAVFD2 0 OPDAF BRTY CEFVAVFD2 0 OPDAF BRTY CDFLAVFD2 0 OPEPDF MRTI CDFLAVFD2 0 OPEPDF MRTI CDFLAVFD2 0 OPEDF MRTI CEFVAIFD2 0 OPESDF LKTI CEFVAIFD2 0 OPESDF LKTI CEFVAIFD2 0 OPESDF LKTI CEYVAIFD2 0 OPEDF LRTY CEYVAIFD2 0 OPEDF LRTY CEYVAIFD2 0 OPEPDF LRTY	FLVHN 198 FLIHN 199 FLIHN 199 FLIHN 199 FLIHN 199 FLIHN 199 FLIHN 199 FLIHN 190 YLQAD 200 VLQAD 201 VLQAD 201 YLARN 225 YLARN 225 YLARN 231 FLVHN 212 FLVHN 212 FLVHN 212 FLVHN 212 FLVHN 216 FLVHN 202 FLVHN 203 FFVHN 216 FFVHN 263 FUMALVQTRWKFINSDEC 111
TRIAE_CS42_2AL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_6DS_TGACV TRIAE_CS42_6AS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_3AS_TGACV	CDFVVIFD2 0 QPEPDY SRAM CELVAVFD2 0 QPDADF RETV CELVAVFD2 0 QPDADF RETV CELVAVFD2 0 QPDADF RETV CEFVAVFD2 0 QPDADF RETV CDFLAVFD2 0 QPEPDF MRTI CDFLAVFD2 0 QPEPDF MRTI CEFVAIFD2 0 QPEDF MRTI CEFVAIFD2 0 QPESDF LKTI CEFVAIFD2 0 QPESDF LKTI CEFVAIFD2 0 QPESDF LKTI CEYVAIFD2 0 QPEPDF LETV CEYVAIFD2 0 QPEPDF LETV	FLVHN 198 FLIHN 198 FLIHN 198 YLQAD 206 VLQAD 207 VLQAD 207 VLQAD 207 VLQAD 207 VLARN 207 YLARN 208 YLARN 232 YLSRN 231 FLVHN 209 FLVHN 202 FLVHN 203 FLVHN 205 FLVHN 206 FLVHN 206 FLVHN 262 FLVHN 262 FLVHN 263 FLVHN 264 FLVHN 264 FLVHN 265 FLVHN 264 GWQRGPVTR
TRIAE CS42_2AL_TGACW TRIAE CS42_2DL_TGACW TRIAE CS42_2DL_TGACW TRIAE CS42_2AL_TGACW TRIAE CS42_1AS_TGACW TRIAE CS42_7BS_TGACW TRIAE CS42_7BS_TGACW TRIAE CS42_6DS_TGACW TRIAE CS42_6AS_TGACW TRIAE CS42_2BS_TGACW TRIAE CS42_2DS_TGACW TRIAE CS42_2DS_TGACW TRIAE CS42_2DS_TGACW TRIAE CS42_2DS_TGACW TRIAE CS42_3AS_TGACW TRIAE CS42_3B_TGACW TRIAE CS42_3B_TGACV1	CDFVVIFD2 0 QPEPDY SRAM CELVAVFD2 0 QPDADF PRTV CELVAVFD2 0 QPDADF PRTV CELVAVFD2 0 QPDADF PRTV CEVVAVFD2 0 QPDADF PRTV CDFLAVFD2 0 QPEPDF WRTI CDFLAVFD2 0 QPEPDF WRTI CDFLAVFD2 0 QPEPDF WRTI CEFVAIFD2 0 QPESDF LKTI CEFVAIFD2 0 QPESDF LKTI CEFVAIFD2 0 QPESDF LKTI CEYVAIFD2 0 QPEPDF LRTV CEYVAIFD2 0 QPEPDF LRTV	FLVHN 198 FLTHN 199 FLTHN 199 FLTHN 199 VLQAD 206 VLQAD 207 VLQAD 207 VLQAD 207 VLARN 207 YLARN 212 YLARN 213 YLVNN 212 FLVHN 213 GQQRGPVTRYGRLPLVNQVVRARPIHHLIVAEAPKRALDRVDKGCRAFFWAGSEEIQ 511
TRIAE_CS42_2AL_TGACW TRIAE_CS42_2DL_TGACW TRIAE_CS42_2DL_TGACW TRIAE_CS42_1AS_TGACW TRIAE_CS42_1AS_TGACW TRIAE_CS42_7AS_TGACW TRIAE_CS42_7AS_TGACW TRIAE_CS42_6DS_TGACW TRIAE_CS42_6BS_TGACW TRIAE_CS42_2DS_TGACW TRIAE_CS42_2DS_TGACW TRIAE_CS42_2AS_TGACW TRIAE_CS42_2AS_TGACW TRIAE_CS42_3AS_TGACW TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3D_TGACV1	CDFVVIFD2 0 QPEPDY SRAM CELVAVFD2 0 QPADF RETY CELVAVFD2 0 QPADF RETY CELVAVFD2 0 QPADF RETY CEFVAVFD2 0 QPADF RETY CDFLAVFD2 0 QPEPDF METT CDFLAVFD2 0 QPEPDF METT CDFLAVFD2 0 QPEPDF METT CEFVAIFD2 0 QPEPDF LETT CEFVAIFD2 0 QPEDF LETT CEYVAIFD2 0 QPEDF LETT CEYVAIFD2 0 QPEPDF LETT LGIKQLTR5 0 QPEVD2 LETM	FLVHN 198 FLIHN 199 FLIHN 199 FLIHN 199 FLIHN 199 FLIHN 199 FLIHN 199 FLIHN 190 YLQAD 200 VLQAD 201 VLQAD 201 VLARN 202 YLARN 223 YLARN 231 FLVHN 212 FLVHN 212 FLVHN 212 FLVHN 202 FLVHN 212 FVHN 213 FFVHN 216 FFVHN 263 FFVHN 263 FFVHN 263 FFVHN 263 SQRGPVTRPGRLPLVNQVVRARPIHHLIVAEAPKRALDRVDKGCRAFFWAGSEEIQ 511
TRIAE_CS42_2AL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_6DS_TGACV TRIAE_CS42_6AS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV	CDFVVIFD2 0 QPEPDY SRAM CELVAVFD2 0 QPDADF RETV CELVAVFD2 0 QPDADF RETV CELVAVFD2 0 QPDADF RETV CEFVAVFD2 0 QPDADF RETV CDFLAVFD2 0 QPEPDF M&TI CDFLAVFD2 0 QPEPDF M&TI CEFVAIFD2 0 QPESDF L&TI CEFVAIFD2 0 QPESDF L&TI CEFVAIFD2 0 QPESDF L&TI CEFVAIFD2 0 QPESDF L&TI CEYVAIFD2 0 QPEPDF LETV CEYVAIFD2 0 QPEPDF LETV	FLVHN 198 FLTHN 198 FLTHN 198 FLTHN 198 VLQAD 206 VLQAD 207 VLQAD 207 VLQAD 207 VLQAD 207 VLQAD 207 VLQAD 207 VLARN 225 YLARN 232 FLVHN 201 FLVHN 202 FLVHN 203 FLVHN 203 FLVHN 204 FLVHN 205 FLVHN 205 FFVHN 263 FFVHN 263 FFVHN 263 FFVHN 263 FFVHN 263 FFVHN 264 FFVHN 265 FFVHN 265 FFVHN 264 FFVHN 265 PDFALVQTRWKFINSDEC 117 PDFALVQTRWKFINSDEC 117 PDFALVQTRWKFINSDEC 117 PDFALVQTRWKFINSDEC
TRIAE CS42_2AL_TGACW TRIAE CS42_2DL_TGACW TRIAE CS42_2DL_TGACW TRIAE CS42_2DL_TGACW TRIAE CS42_1AS_TGACW TRIAE CS42_7AS_TGACW TRIAE CS42_7AS_TGACW TRIAE CS42_7BS_TGACW TRIAE CS42_6AS_TGACW TRIAE CS42_6AS_TGACW TRIAE CS42_2AS_TGACW TRIAE CS42_2AS_TGACW TRIAE CS42_2DS_TGACW TRIAE CS42_3AS_TGACW TRIAE CS42_3B_TGACVI TRIAE CS42_3B_TGACVI TRIAE CS42_3B_TGACVI TRIAE CS42_3DS_TGACVI TRIAE CS42_3DS_T	CDFVVIFD2 0 QPEPDY SRAM CELVAVFD2 0 QPDADF PRTV CELVAVFD2 0 QPDADF PRTV CELVAVFD2 0 QPDADF PRTV CEVAVFD2 0 QPDADF PRTV CDFLAVFD2 0 QPEPDF WRTI CDFLAVFD2 0 QPEPDF WRTI CDFLAVFD2 0 QPEPDF WRTI CEFVAIFD2 0 QPESDF LKTI CEFVAIFD2 0 QPESDF LKTI CEFVAIFD2 0 QPESDF LKTI CEYVAIFD2 0 QPEPDF LRTV CEYVAIFD2 0 QPEPDF LRTV	FUNN 198 FLINN 206 VLQAD 207 VLQAD 207 VLQAD 207 VLARN 202 YLARN 225 YLARN 212 FLVHN 209 FFVHN 263 FFVHN 263 FFVHN 263 GWQRGPVTRPGRLPLVNQVVRARPIHHLIVAEAPKRALDRVDKGCRAFFWAGSEEIQ 511
TRIAE_CS42_2AL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_TAS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_3B_TGACVI TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV	CDFVVIFD2 0 QPEPDY SRAM CELVAVFD2 0 QPDADF RETV CELVAVFD2 0 QPDADF RETV CELVAVFD2 0 QPDADF RETV CEFVAVFD2 0 QPDADF RETV CDFLAVFD2 0 QPEPDF WETTI CDFLAVFD2 0 QPEPDF WETTI CDFLAVFD2 0 QPEDF WETTI CEFVAIFD2 0 QPEDF LETTI CEFVAIFD2 0 QPEDF LETTI CEYVAIFD2 0 QPEDF LETTI CEYVAIFD2 0 QPEDF LETTI CEYVAIFD2 0 QPEDF LETTI LGIKQLTR5 0 QPEPDF LETV LGIKQLTR5 0 QPEPDF LETV	FLVHN 198 FLIHN 199 FLIHN 190 YLQAD 200 VLQAD 201 VLQAD 201 VLARN 202 YLARN 233 YLSRN 231 FLVHN 212 FLVHN 202 FLVHN 202 FLVHN 202 FLVHN 202 FFVHN 203 FFVHN 205 FFVHN 263 FFVHN 263 FFVHN 263 FFVHN 263 FFVHN 263 Sequer Berger Be
TRIAE_CS42_2AL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_TAS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV1 TRIAE_CS42_3DT_TGACV2 TRIAE_CS42_3DT_TGACV2 TRIAE_CS42_3DT_TGACV2 TRIAE_CS42_3DT_TGACV2 TR	CDFVVIFD2 0 QPEPDY SRAM CELVAVFD2 0 QPDADF RETV CELVAVFD2 0 QPDADF RETV CELVAVFD2 0 QPDADF RETV CELVAVFD2 0 QPDADF RETV CDFLAVFD2 0 QPEPDF MRTII CDFLAVFD2 0 QPEPDF MRTII CDFLAVFD2 0 QPEPDF MRTII CEFVAIFD2 0 QPEDF LETVI CEFVAIFD2 0 QPEDF LETVI CEVVAIFD2 0 QPEDF LETVI CEVVAF D2 0 QPEDF LETVI CEVAF	FLVHN 198 FLTHN 198 FLTHN 198 YLQAD 206 VLQAD 207 VLARN 225 YLARN 235 FLVHN 209 FLVHN 201 FLVHN 202 FLVHN 203 FLVHN 205 FFVHN 205 FFVHN 206 FFVHN 263 FFVHN 263 FFVHN 263 FFVHN 263 Segog VTRPGRLPLVNQVVRARPIHHLIVAEAPKRALDRVDKGCRAFFWAGSEEIQ 511
TRIAE_CS42_2AL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_3B_TGACV1 TRIAE_CS42_5B_TGACV1 TRIAE_CS42_5B_TGACV1 TRIAE_CS42_5B_TGACV1 TRIAE_CS42_5B_TGACV1 TRIAE_CS42_5B_TGACV1 TRIAE_CS42_5B_TGACV1 TRIAE_CS42_5B_TGACV1 TRIAE_CS42_5B_TGACV1 TRIAE_CS42_5B_TGACV1 TRIAE_CS42_5B_TGACV1 TRIAE_CS42_5B_TGACV1 TRIAE_CS42_5B_TGACV1 TRIAE_CS42_5B_TGACV1 TRIAE_CS42_5B_TGACV1 TRIAE_CS42_5B_TGACV1 TRIAE_CS42_5B_TGACV1 TRIAE_CS42_5B_TGACV2 TRIAE_CS42_5B_TGACV2 TRIAE_CS42_5B_TGACV2 TRIAE_CS42_5B_TGACV2 TRIAE_CS42_5B_TGACV2 TRIAE_CS42_5B_TGACV2	CDFVVIFD2 0 QPEPDY SRAM CELVAVFD2 0 QPDADF PRTV CELVAVFD2 0 QPDADF PRTV CELVAVFD2 0 QPDADF PRTV CEVAVFD2 0 QPDADF PRTV CDFLAVFD2 0 QPEPDF WRTI CDFLAVFD2 0 QPEPDF WRTI CDFLAVFD2 0 QPESDF LKTI CEFVAIFD2 0 QPESDF LKTI CEFVAIFD2 0 QPESDF LKTI CEFVAIFD2 0 QPESDF LKTI CEYVAIFD2 0 QPEPDF LTTV CEYVAIFD2 0 QPEPDF LTTV	FUNN 198 FULHN 199 FLIHN 199 FULHN 199 FULHN 199 FULAD 200 VLQAD 201 VLQAD 201 VLQAD 201 VLQAD 201 VLQAD 201 VLQAD 201 VLARN 225 YLARN 235 YLSRN 212 FLVHN 212 FLVHN 212 FLVHN 203 FFVHN 263 PDFALVQTRWKFNSDEC 275 SUQRGPVTRFGRLPLVNQVVRARPIHHLIVAEAPKRALDRVDKGCRAFFWAGSEEIQ 511
TRIAE_CS42_2AL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV1 TRIAE_CS42_3DS_	CDFVVIFD2 0 QPEPDY SRAM CELVAVFD2 0 QPADF RETV CELVAVFD2 0 QPADF RETV CELVAVFD2 0 QPADF RETV CEFVAVFD2 0 QPADF RETV CDFLAVFD2 0 QPEPDF METTI CDFLAVFD2 0 QPEPDF METTI CDFLAVFD2 0 QPEPDF METTI CEFVAIFD2 0 QPEPDF LETTI CEFVAIFD2 0 QPEPDF LETTI CEYVAIFD2 0 QPEPDF LETTI CEYVAIFD2 0 QPEPDF LETTI CEYVAIFD2 0 QPEPDF LETTI LGIKQLTR5 0 QPEPDF LETTI LGIKQLTR5 0 QPEPDF LETTI	FLVHN 198 FLTHN 199 FLTHN 200 VLQAD 201 VLQAD 201 VLQAD 201 VLARN 225 YLARN 231 FLVHN 212 FLVHN 212 FLVHN 212 FLVHN 212 FLVHN 216 FVHN 263 FFVHN 263 FFVHN 264 FFVHN 265 FFVHN 265 FFVHN 264 GWQRGPVTRPGRLPLVNQVVRARPIHHLIVAEAPKRALDRVDKGCRAFFWAGSEEIQ 511
TRIAE_CS42_2AL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_6AS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3DS_TGACV	CDFVVIFD2 0 QPEPDY SRAM CELVAVFD2 0 QPDADF RETV CELVAVFD2 0 QPDADF RETV CELVAVFD2 0 QPDADF RETV CEFVAVFD2 0 QPDADF RETV CDFLAVFD2 0 QPEPDF METTI CDFLAVFD2 0 QPEPDF METTI CEFVAIFD2 0 QPEDF LETTI CEFVAIFD2 0 QPEDF LETTI CEFVAIFD2 0 QPEDF LETTI CEYVAIFD2 0 QPEDF LETTI CEYVAIFD2 0 QPEDF LETTI LGIKQLTR: 0 QPEDF LETTI LGIKQLTR: 0 QPUPDF LETTI	FLVHN 198 FLIHN 198 FLIHN 198 YLQAD 206 VLQAD 207 VLARN 228 YLARN 231 FLVHN 212 FLVHN 213 FLVHN 206 FLVHN 205 FLVHN 205 FLVHN 206 FLVHN 205 FLVHN 206 FLVHN 206 FLVHN 263 FVHN 263 FVHN 264 FVHN 265 FVHN 264 FVUN 265 FVHN 265 FVUN 266 FVUN 261 FVUN 261 FVUN 261 FVUN 261 FVUN
TRIAE CS42_2AL_TGACV TRIAE CS42_2DL_TGACV TRIAE CS42_2DL_TGACV TRIAE CS42_2DL_TGACV TRIAE CS42_1AS_TGACV TRIAE CS42_1AS_TGACV TRIAE CS42_7BS_TGACV TRIAE CS42_7BS_TGACV TRIAE CS42_6BS_TGACV TRIAE CS42_6BS_TGACV TRIAE CS42_2BS_TGACV TRIAE CS42_2AS_TGACV TRIAE CS42_2AS_TGACV TRIAE CS42_2DS_TGACV TRIAE CS42_3BS_TGACV TRIAE CS42_3B_TGACV1 TRIAE CS42_3B_TGACV1 TRIAE CS42_3DL_TGACV TRIAE CS42_3DL_TGACV	CDFVVIFD2 0 QPEPDY SRAM CELVAVFD2 0 QPDADF RETV CELVAVFD2 0 QPDADF RETV CELVAVFD2 0 QPDADF RETV CELVAVFD2 0 QPDADF RETV CEFVAVFD2 0 QPEPDF WETT CDFLAVFD2 0 QPEPDF WETT CDFLAVFD2 0 QPEPDF WETT CEFVAIFD2 0 QPESDF LKTI CEFVAIFD2 0 QPESDF LKTI CEFVAIFD2 0 QPESDF LKTI CEYVAIFD2 0 QPEPDF LETV CEYVAIFD2 0 QPEPDF LETV	FLVHN 198 FLIHN 198 FLIHN 198 YLQAD 206 VLQAD 207 VLARN 225 YLARN 223 YLSRN 231 FLVHN 212 FLVHN 209 FLVHN 202 FLVHN 203 FLVHN 205 FLVHN 206 FFVHN 263 FFVHN 264 GWQRGPVTRPGRLPLVNQVVRARPIHHLIVAEAPKRALDRVDKGCRAFFWAGSEEIQ 511
TRIAE_CS42_2AL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_6DS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV1 TRIAE_CS42_3DS_TGACV1 TRIAE_CS42_3DS_TGACV1 TRIAE_CS42_3DS_TGACV1 TRIAE_CS42_3DS_TGACV1 TRIAE_CS42_3DS_TGACV1 TRIAE_CS42_3DS_TGACV1 TRIAE_CS42_3DS_TGACV1 TRIAE_CS42_3DS_TGACV1 TRIAE_CS42_3DS_TGACV1 TRIAE_CS42_3DT_TGACV1 TRIAE_CS42_0AL_TCACV1 TRIAE_CS42_0AL_TCACV1 TRIAE_CS42_0AL_TCACV1 TRIAE_CS42_0AL_TCACV1 TRIAE_CS42_0AL_TCACV1 TRIAE_CS42_0AL_TCACV1 TRIAE_CS42_0AL_TCACV1 TRIAE_CS42_0AL_T	CDFVVIFD2 0 QPEPDY SRAM CELVAVFD2 0 QPDADF RETV CELVAVFD2 0 QPDADF RETV CELVAVFD2 0 QPDADF RETV CEFVAVFD2 0 QPDADF RETV CDFLAVFD2 0 QPEPDF METTI CDFLAVFD2 0 QPEPDF METTI CEFVAIFD2 0 QPESDF LETTI CEFVAIFD2 0 QPESDF LETTI CEFVAIFD2 0 QPESDF LETTI CEYVAIFD2 0 QPEPDF LETTI	FUNN 198 FLINN 198 VLQAD 200 VLQAD 201 VLQAD 201 VLARN 202 YLARN 223 YLARN 224 FLVHN 221 FLVHN 212 FLVHN 212 FLVHN 212 FVHN 212 FVHN 263 FFVHN 263 FFVHN 264 GWQRGPVTRPGRLPLVNQVVRARPIHHLIVAEAPKRALDRVDKGCRAFFWAGSEEIQ 511

TRIAE_CS42_7AL_TGACv TRIAE_CS42_7BL_TGACv	PETALVQARWVFVMANECIM TR MQEMSL PETALVQARWVFVMANECIM TR MQEMSL	226 226
TRIAE_CS42_7DL_TGACV		226
TRIAE_CS42_2AL_IGACV	PS <mark>VALVQARWRFVNADECHLARTQOO</mark> SL PS <mark>VALVQARWRFVN</mark> ADECHL <mark>AR</mark> TQOOSL	254
TRIAE_CS42_2BL_TGACV TRIAE_CS42_1AS_TGACV		235 225
TRIAE_CS42_7DS_TGACv	PQIALVQARWEFVNPNECLMTRIQKWTL	257
TRIAE_CS42_/AS_TGACv TRIAE_CS42_7BS_TGACv	PQIALVQARWEFVNPNECIMURIQKWTL POIALVQARWEFVNPNECIMURIQKWTL	263 259
TRIAE_CS42_6DS_TGACv	PŔTALVQTRWKFVNYDACIMTRIQKUSL	240
TRIAE_CS42_6BS_TGACV TRIAE_CS42_6AS_TGACV	PRIALVQTRWKFVMYDACLYTRIQKWSL PRIALVQTRWKFVMYDACLYTRIQKWSL	237
TRIAE CS42 2BS TGACV	PEVALVQARWSFVNDTASILERVQKWFF	291
TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV	PEVALVQARWSFVNDIASILIEV QAMFF PEVALVQARWSFVNDIASILIEV QAMFF	291
TRIAE_CS42_2AS_TGACV	PKVALVQARWSFVNGTVSILARIQKMFF	289 591
ININE_COAL_200_IONCV		551
TRIAE_CS42_3AS_TGACv TRIAE_CS42_3B_TGACv1	D/HEKFDQDAGSIVYSDGCDNGTAGVWRISAINDAGGWRERTIVIDMDJAVRTALLGLKEVYYGAWKWKSDDCSTFKAMR D/HEKFDQDAGSIVYSFFGENGTAGVWRISAINDAGGWRERTWVFDMDJVVRTALLGLKEVYIGAWKWKSDDCSTFKAWR	. 337
TRIAE_CS42_3B_TGACv1	D /HEKFBQBACSIVYSEFGENGTAGVWRISAINDAGGWEDRTIVIDMD /AVRTSLLGWKEVYVGAVKWKSELPSTFKAMR	341
TRIAE_CS42_3DS_TGACV	D HEARING DAG IVISENGINGINGINGINGING SEING AGGED KIINT DID SVAIALIGWAVIVGANAVGSDESI PATA D HEAFEQEAGSIVISEFGENGTAGVWRISAIND AGGWNDRTIVIDHD JAVRTALLGWAVIVGDVKWRSELPSTFKAVR	341
TRIAE_CS42_3DL_TGACV	D (HEKEBQEAGSIVYSEFGENGTACYWRISHIDDA COMKIRTIWEDMD AVRTALKGWKEVY (GAWKURSELPSTEKAYR	342
TRIAE_CS42_3AL_TGACV	D (HERFEQEAGSIVISTEGENGTAGVWRISAIDDAGGWRDRTTVIDMCARTALKGWREVYVGAVKWRSELPS	342
TRIAE_CS42_6BS_TGACV TRIAE_CS42_6AS_TGACV	D (HEKVEQEVSSSVCAFEGENGTAGVWRIAAVNEAGGIKERTIVEDMD AALRASLKGWKEVYLGDVOVKSELPSTEKAER D (HEKVEOEVSSSVCAFEGENGTAGVWRIABVNEAGGIKERTIVEDMD AALRASLKGWKEVYLGDVOVKSELPSTEKAER	307
TRIAE_CS42_7AL_TGACv	D (HERVEQEVCSSTYAFFGENGTACYWRISAINBACGWKERTIVEDMD JAVRASLKGWKEVYIGDIKUKNELESSTFKAFR	309
TRIAE_CS42_U_TGACv1_ TRIAE_CS42_U_TGACv1	D (HEKVEQEVGSSTYAESGENGTAGVWRISH NEAGGWRIRTV: DMD JAVRASLKGWREVYLED KWKNELPSTFKAER D (HEKVEQEVGSSTYAESGENGTAGVWRISH NEAGGWRIRTTV: DMD JAVRASLKGWREVYLED KWKNELPSTFKAER	303
TRIAE_CS42_7BL_TGACV	D (HEKVEQEVCSSTYAFFGENGTAGVWRISAINDAGGWKIRTTV: DMD JAVRASLKGWKEVYIGDIKWKNELESTFKAFR	310
TRIAE_CS42_/AL_TGACV TRIAE_CS42_7BL_TGACV	D (HEKVEQEVGSSAYAFF GENGTAGVWRISA INDAGGWADRTIVEDMD DAVRASLKGWKEVCIGDI RUKSELESTFKAAR D (HEKVEQEVGSSAYAFF GENGTAGVWRISALNDAGGWADRTIVEDMD DAVRASLKGWKEVVIGDI RUKSELESTFKAAR	306
TRIAE_CS42_7DL_TGACV	D HEKVEQEVCSSAYAFFGENGTACVWRISA IN ACGWK RTTVIDMD AVRASLKGWKEVYLCD I RYKSELPSTFKAR Na herveren von State of the second state of the se	306
TRIAE_CS42_2AL_IGACV	D THE VEODICEACHERS CHROMENOUS FRANK D THE SVEQEVCSACHERS CENERACY WINVER AD A CONKERT IVED DAVIS SHREM VECDVO WANDERS SHREMA D THE SVEQEVCSACHERS CENERACY WANDERS SHREMAN	334
TRIAE_CS42_2BL_TGACV	D (HESVEQEVCSACHGESGENGTACVWRVQALADAGGWKDRTTWEDMD AVRASMRGWREVYEGDVQWRNELEPSSEKAYR	315
TRIAE_CS42_7DS_TGACv	D (HEKVEQEAGSSTFAFEGENGTAGVWRISALKEAGGWDIRTTVEDMDIAVRAGLKGWKEVYVGDVKVKSELPSNLKAVR	337
TRIAE_CS42_7AS_TGACv TRIAE_CS42_7BS_TGACv	D (HEKVEQEACSSTFAEFGENGTAGWIRISH KEAGONDIRTIVEDMD JAVRAGLKGWKEVY (GDWKWKSEDESNIKAMR D (HEKVEOFACSSTFAEFGENGTAGWIRISH KEAGONDIRTIVEDMD JAVRAGLKGWKEVY (GDWKWKSEDESNIKAWR	343
TRIAE_CS42_6DS_TGACv	D (HEKVEQESCSFMHAREGENGTAGVWRVSAUNESGGWKERTTV: DMD JAVRACLKEWERLY (GDURVKSELLSSTFKAVR	320
TRIAE_CS42_6BS_TGACV TRIAE_CS42_6AS_TGACV	D (HEKVEQESGSFMHAFFGENGTAGVWRVSALNESGGWARTIVEDMD JAVRACLKEWEFTY (GDLRWKSELPSTFKAYR D (HEKVEQESGSFMHAFFGENGTAGVWRVSALNESGGWARTIVEDMD JAVRACLKEWEFTY (GDLRWKSELPSTFKAYR	317
TRIAE CS42 2BS TGACV	D /HEKVEQEACSATFSEESENGTACVWRTSATKEACCWRTVI DMD /AVRATLKGWKEVYYCDI RWKSELPSTYKAYC	371
TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV	D HEAVEODAGSATTSEESENGTASVWATAETABAGGMORTIVEDWD AAVAATLAGWAEVIIGDIAWAASEDESTIAATC D HEKVEQEACSATTSEESENGTAGVWAAAIKDAGGMKORTIVEDMD AAVAATLAGWAEVIIGDIAWAASEDESTIAATC	370
TRIAE_CS42_2AS_TGACV	D /HEKVEQEAGSATFAFFSENGTAGVWRTARIKEAGGWK RTTVIDMD AI RATLKGWKFIYYCDI RWKSELPSSYKAYC	369
IRIAE_C542_2D5_IGACV		0/1
TRIAE_CS42_3AS_TGACv TRIAE_CS42_3B_TGACv1	FOORWOOD FOOR FOR THE AND FOUND TO THE AND FOUND FOR THE AND FOUND FOR THE ADDRESS OF THE ADDRES	415 781
TRIAE CS42 3B TGACV1	FORKASCEPANIFKKILLDIIKNKKVSFWSKIHLIYD FFVGKIAAHTVTFIYYCFAIPVSVFFPEIQIPLWGVVYV	419
TRIAE_CS42_3DS_TGACV TRIAE_CS42_3DS_TGACV	FOR THE AND CERTIFIC TO THE ANALYSIS WAS ALLED TO FINGET AND ANY IFTIC FOR THE STAFF - EIGIFLWGVIV FOR THE STAFF AND THE ANALYSIS ANALYSIS ANALYSIS ANALYSIS AND THE ANALYSIS ANALYSIS ANALYSIS A	419
TRIAE CS42 3DL TGACV	FOR THE SECTION OF THE SECTION. SECTION OF THE SECTION OF THE SECT	420
TRIAE_CS42_3AL_TGACV	FOOR WAS COPANDER KNIVET DENKKVSEWSKIHLIYD FFVGKTAAHTVTFIYYCF IPLSVFFPEIQIPLWGVVVV	420
TRIAE_CS42_6BS_TGACV TRIAE_CS42_6AS_TGACV	FYOHRWSCEPANEFREMILEFVTNKKVTIWKKFHVTYNEFLVRKTVEHIVTFTFYCI IPTTIFVPEVHIPKWGCVYI FYOHRWSCEPANEFREMI.VETVTNKKVTIWKKFHVTYNEFINRKTVAHIVTFTFYCI IPTTIFVPEVHIPKWGCVYI	385 385
TRIAE_CS42_7AL_TGACv	Y OHRNSCEPANIFERMVMETIRNKKVILWKKIHVVYSEFLVRKVVHHIVTFVFYCLVIPATVLVPEVEVPKWGCVYI	387
TRIAE_CS42_U_TGACV1_ TRIAE_CS42_U_TGACV1	YOHRWICGPANLFRKMVMBIIRNKKVILWKKIHVVYSFLVRKVVAHIVTFVFYCLVIPATVLVP-EVEVPKWGCVYI YOHRWICGPANLFRKMVMBIIKNKKVILWKKIHVVYN FFIRKVVAHIMTFVFYCLVIPATVLVP-EVEVPKWGCVYI	381 391
TRIAE CS42 7BL TGACV	Y CORRECCEPANE FRAMMET VRNKKVILWKKIHVIYN FLWRWVAHIVTFVFYCVWERIDMVD	375
TRIAE_CS42_7AL_TGACV	YOHMAN COPANIFRAMING HAVAN INANI IVIINA FUNCTIONIISVICONIPATVEVP-EVEIPANGYFYI YOHMAN COPANIFRAMING WANQAWI INANI IVIINA FEWRAIICHIITSVFICINIPATVEVP-EVEIPANGYFYI	384 384
TRIAE_CS42_7DL_TGACV	Y OHRWSCEPANERRAMMET VKNOKVILWKRIYVIYN FFVRKTICHILTSVFYCLUIPATVFVPEIEIPRWGYFYI Y OHRWSCEPANERRAMENER WASBOWSAWKRUUW Y DEFWERWYHLUTFLEVOUWTPAVVLVGCODVRLEKYVAWV	384
TRIAE_CS42_2DL_TGACV	YOHRATCO PANIMERAMENT VASCON AND A CONTRACT AND A C	414
TRIAE_CS42_2BL_TGACV TRIAE_CS42_1AS_TGACV	YOOHNWY COOPANIM MIKMFWDTWASROWSAWKW VHYTYGEFFWRWYDHLVTFLFYCW I PAYVLVGGODVRLPKYVAMYV YOOHNWY COOPANIM MIKMFWDTWANKOWSAWKWLHYTWGEFFWRWYDHLATFLFCCWU PYYVLVGGODVWLPOYVPMYV	395 385
TRIAE_CS42_7DS_TGACv	RVOHRWICCAANIFRKMGAFIULTKEVSLWWKLYLIYSFFLVRKVVAHVVPFVLYCVUIPFSVLIPEIKIPAWGVVYI	415
TRIAE_CS42_/AS_TGACv TRIAE_CS42_7BS_TGACv	ROHRWICCAANUFRKMGAPULITKEVSLWWKLYLYS FLWRKVVAHVLPFVLYCVWIPFSVLIPEIKIPAWGVVYI ROHRWICCAANLFRKMGAPULITKEVSLWWKLYLYS FLWRKVVAHVVPFVLYCVWIPFSVLIPEIKIPAWGVVYI	421 417
TRIAE CS42 6DS TGACV	HOHRNICCAANIFREMGWEIVINKGVSIWKEWHLIYSELEVRRVIAPILTFLFYCVVIPLSAMVPEVHIPVWGLVYI	398
TRIAE_CS42_6AS_TGACV TRIAE_CS42_6AS_TGACV	HOHKNICGAANDFERMIGMEIVINKGVSIWKKWHLISELFVRKVIHPILIFLFICVMIPLSAMVP-EVHIPVWGLVII H <mark>OHKNICG</mark> AANDFERMIGMEIVINKGVSIWKKWHLISELFVRKVIAPILIFLFYCVVIPLSAMVP-EVHIPVWGLVII	395
TRIAE_CS42_2BS_TGACV	ROFEWSCGGARUFREVAKDILTAKDVSLIKEFHNIVSEFLWRVVPPVACILYNILLPISVMIPELFLPVWGIAYI	449 440
TRIAE_CS42_2DS_IGACV TRIAE_CS42_2AS_TGACV	ROFKNOCCARDENNARDIIIRREVOLIKKERMINSEFLURAVNAFIVACILINIEPISVMIP-ELFLEVWGIAII ROFKNOCCARDERKVARDIIIRREVSLIKKERMINSEFLURRVVAPILACILINIELPISVMIP-ELFLEVWGIAII	449
TRIAE CS42 2AS TGACV	ROFEWLOGGANLFRKVALDIJJSKDVSVVKKFYMLVSELFVRRVVEPAVACILSNILVPLSVMIPELYLPVWGVAYI	447 740
	- WAT HAT DE AN THE AND THE AN	
TRIAE_CS42_3AS_TGACv TRIAE_CS42_3B_TGACv1	PTVITLCKALGSPSSFHLVILWVLFDNVMSLHRIKATITGLLDARRVNEWVVTEKLGDANKTEPAVEGLNDVQVIDVELS PTVITLCKALGSPSSFHLVILWVLFDNVMSLHRIKATITGLLDARRVNEWVVTEKLGDTNKTEPAMEGLNDVOVIDVELS	495 861
TRIAE_CS42_3B_TGACv1	PTVITLCKALGSPSSFHLVILWVLFDNVMSLHRIKATITGLLDTRRVNEWVVTEKLGDANKTEPAMERLDDVQVIDVELS	499
TRIAE_CS42_3DS_TGACv TRIAE_CS42_3DS_TGACv	PTVITLCKALGSPSSFHLVILWVLFDNVMSLHRIKATITGLLDTRRVNEWVVTEKLGDANKTEPAMEGLDDVQVIDVELS PTVITLCKSLGSPSSFHLVILWVLFENVMSLHRIKATITGLLDTRRVNEWVVTEKLGDANKTEPAMEGLDDVOVIDVELS	497 499
TRIAE_CS42_3DL_TGACV	PTVITLCKALGSPSSFHLVILWVLFENVMSLHRIRAAITGLLDAGRVNEWVVTEKLGDANKTKPATEVLDAVKVIDVELT	500
TRIAE_CS42_3B_TGACv1 TRIAE_CS42_3AL_TGACv1	PTVITLCKALGSPSSFHLVILWVLFENVMSLHRIRAAVTGLLDAGRVNEWVVTEKLGDANKTKPAMEVLDAVKVIDVELT PTVITLCKALGSPSSFHLVILWVLFENVMSLHRIKAAVTGLLDAGRVNEWVVTEKLGDANKTKPAMEALDAVKVIDVELA	498
TRIAE_CS42_6BS_TGACv	PTIITLLNSVGTPRSFHLLFFWILFENVMSLHRTKATLIGLLEAGRANEWVVTEKLGSAMKMK	448

TRIAE_CS42_6AS_TGACv	PTIITLLNSVGTPRSFHLLFFWILFENVMS:	LHRTKATLIGLLEAGRANEWVVTEKLG	SAMKMK	448
TRIAE_CS42_7AL_TGACv	PTIITLLNAVGTPRSVHLVVFWVLFENVMS:	LHRAKATFIGLLEAGTVNEWVVTEKLG	DTLKAK	450
TRIAE_CS42_U_TGACv1_	PTIITLLNAVGTPRSVHLVVFWVLFENVMS:	LHRAKATFIGLLEVGTVNEWVVTEKLG	DTLKAK	444
TRIAE_CS42_U_TGACv1_	PAIITLLSVVGTPRSVHLVIFWALFENVMS:	LHRTKATFIGLLEAHTVNEWVVTEKLG	DTVKTK	454
TRIAE_CS42_7BL_TGACv				375
TRIAE_CS42_7AL_TGACv	PTVITLLNAVGTPRSFHLVIFWVLFENVMS:	LHRTKATFSGLLELGRVNEWVVTEKLG	DILKMK	447
TRIAE_CS42_7BL_TGACv	PTVITLLNAVGTPRSFHLVIFWVLFENVMS:	LHRTKATFSGLLELGRVNEWVVTEKLG	DVLKMK	447
TRIAE_CS42_7DL_TGACv	PTIITLLNAVGTPRSFHLVIFWVLFENVMS:	LHRTKATFSGLLELGRVNEWVVTEKLG	DVLKMK	447
TRIAE_CS42_2AL_TGACv	PAIITLLNAVCTPRSWHLLVFWILFENVMS	MHRSKATIIGLVEASRANEWVVTEKLGSV	TS-TPAATT	461
TRIAE_CS42_2DL_TGACv	PAIITLLNAVCTPRSWHLLVFWILFENVMS	MHRSKATIIGLVEASRANEWVVTEKLGSV	TSSTPAATT	482
TRIAE_CS42_2BL_TGACv	PAIITLLNAVCTPRSWHLLVFWILFENVMS	MHRSKATIIGLVEASRANEWVVTEKLGSV	TS-TPAAAT	462
TRIAE_CS42_1AS_TGACv	AAVLTLLNAVCTPRSCHLLVFWILFENVMS	IHRCKATIIGLLEASRANEWVVTEKLGGS	TTSTPAAAT	453
TRIAE CS42 7DS TGACV	PTAITVLYAVRNPSSIHFIPFWILFENVMS	FHRTKATFIGLLELGSVNEWVVTEKLG	SASNT	477
TRIAE CS42 7AS TGACV	PTAITILYAVRNPSSIHFIPFWILFENVMS	FHRTKATFIGLLELGSVNEWVVTEKLG	SVSNT	483
TRIAE CS42 7BS TGACV	PTAITILYAVRNPSSIHFIPFWILFENVMS	FHRTKATFIGLLELGSVNEWVVTEKLG	SVSNT	479
TRIAE CS42 6DS TGACV	PTAITVMNAIRNPGSLHLMPFWILFENVMS	MHRMRAALTGLLETAHVNDWVVTEKVG	DLVKDD	461
TRIAE CS42 6BS TGACV	PTAITIMNAIRNPGSLHLMPFWILFENVMS	MHRMRAALTGLLETAHVNDWVVTEKVG	DLVKDD	458
TRIAE CS42 6AS TGACV	PTAITIMNAIRNPGSLHLMPFWILFENVMS	MHRMRAALTGLLETAHVNDWVVTEKVG	DVVKDD	458
TRIAE CS42 2BS TGACV	PTVLLVVTAIRHPKNLHILPFWILFESVMT	MHRMRAALSGLFELSEFNEWVVTKKTG	NNFE	510
TRIAE CS42 2DS TGACV	PTVLLVVTAIRHPKNLHILPFWILFESVMT	MHRMRAALSGLFELSEFNEWVVTKKTG	NNFE	510
TRIAE CS42 2AS TGACV	PTVLLVVTAIRHPKNLHILPFWILFESVMT	MHRMRAALSGLFELSEFNEWVVTKKTG	NNFE	509
TRIAE CS42 2AS TGACV	PAVLLVVTAIRNPKNIHLLPFWILFESVMT	IHRTRAALVGLFEFSEFNEWVVTKKTG	NNFE	508
TRIAE CS42 2DS TGACV	PTVI.I.VVTATRNPKNTHLI.PFWTLFESVMT	THRTRAALVGLEELTEEDEWLVTKKTG	NNFE	810
				010
TRIAR CS42 3AS TCACT	TPL VPKLEKERTELWOKYNOSELEVOTOTT	CCCVDVLVA-NKCVVTVLFTOCLAFLVTCFFVTCTRPPNT		566
TRIME_CS42_SHO_IGHEV	THE VIRGENERATION DRINGSET FUCTOR	LCCCADAIN MAGILLELIQUEN EVICEENICABBECO		925
TRIAD_CO42_3D_TGACV1	TIDALUPEUN INCOPILAGICII	LOCOADATY MUGCITITEIOCANETAICEEAICEEDE		520
TRIAL COME DE TGACVI	TIT A SUPER CONTRACT AND A CONTRACT	ICCCADALAY - KKCAALAL ETOCLYELALCEEATCEEDED		560
TRIAE_C342_3D3_IGACV	TPLVPRLERRRIRLWDRINCSEIFVGICII.		.E	500
TRIAL COME SUS TRACV	TI DV ENDERNALINDA INCOLTRUCTULI.	ICCEVDIEVA-NECOVIVIEIOCIAEIVOCEEVICEDDODE	27	570
TRIAE_CS42_3DL_TGACV	TPLVPKLKKRRIRLWDKINCSEIFVGTCIV.		SA	572
TRIAL CS42_3B_TGACVI	IFLVPKLKKKRIKLWDKINCSEIFVGTCII.	LOGFIDLFIA-NKGIIIILFIQGLAFLVVGFEYIGTRPPTF	SAG	3/1 572
TRIAE_CS42_3AL_TGACV	TPLVPKLKKRRIKLWDKINCSEIFVGTSII.	ICGFYDLFYA-NKGYYIYLFIQGLAFLVVGFEYIGTRPPTF	SAE	5/3
TRIAE_CS42_6BS_TGACv	SANKASARKSFMRMWERLNVPELGVGAFLF	SCGWYDVAFG-KDNFFIYLFFQSMAFFVVGVGYVGTIVPPS	}	518
TRIAE_CS42_6AS_TGACv	SANKASARKSFMRMWERLNVPELGVGAFLF:	SCGWYDVAFG-KDNFFIYLFFQSMAFFVVGVGYVGTIVPQS		518
TRIAE_CS42_7AL_TGACv	MPSKALK-KLRMRIGERLHLWELGVAAYLF	LCGCYDISFG-NNRYFIFLFMQSIAFFIVGVGVVGTFVAQ-		518
TRIAE_CS42_U_TGACv1_	MPSKALR-KLRMRIGERLHLWELGVAAYLF	LCGCYDISFG-NNRYFIFLFMQSIAFFIVGVGYVGTFVAQ-		512
TRIAE_CS42_U_TGACv1_	MPSKALK-KLRIGIGERLHLWELGVAAYLF	ICGCYSISFG-NNHYFIFLLMQSIAFFIVGVGYVGTFVTQ-		522
TRIAE_CS42_7BL_TGACv				375
TRIAE_CS42_7AL_TGACv	VQSKVTK-KLRMRIRERLQLLELGVAAYIF	FCGSYDLLFG-KRYYYVFLFMQSIAFFVVGVGFVGTLVPN-		515
TRIAE_CS42_7BL_TGACv	VQSKVTK-KLRMRIRERLQLLELGVAAYIF	FCGSYDLLFG-KRYYYIFLFMQSIAFFVVGVGFVGTLVPN-		515
TRIAE CS42 7DL TGACV	VQSKVTK-KLRMRIRERYIIIIGLLMCISQ	LSYLNFNMES-GCSFWSLVLQPISSFVEVTTFCLAKDITIS	FSSCNPSLS	525
TRIAE CS42 2AL TGACV	TMATNKGAMKKKKSQSSILAPEIVMGLCLL	YCAVYDIFFG-HDHFYVYLLMQSAAAFVIGFGYVGSQ		527
TRIAE CS42 2DL TGACV	TMATNKGATKKKKSQSSILAPEIVMGLCLL	YCAVYDIVFG-HDHFYVYLLMQSAAAFVIGFGYVGSQ		548
TRIAE CS42 2BL TGACV	TMAANKGAMKKKKSQSSILAPEIVMGLCLL	YCAVYDIVFG-HDHFYVYLLMQSAAAFVIGFGYVGSQ		528
TRIAE CS42 1AS TGACV	TTMVAKKKKSSSSFLAPEIVMGLFLL	YCALYDIVFG-HDHFYVYLLMOSAAAFVIGFGYVGSO		515
TRIAE CS42 7DS TGACV	KPVPQILERPRCRFWDRWTVSELLFAVFLF"	VCATYNLVYG-SDFYFIYIYLQAITFIIVGTGFCGTSNS		545
TRIAE CS42 7AS TGACV	KPVPOILERPRCRFWDRWTVSELLFAVFLF	VCATYNLVYG-SDFYFIYIYLOAITFIIVGTGFCGTSNS		551
TRIAE CS42 7BS TGACV	KPVPOILEKPRCRFWDRWTVSELLFAVELF	VCATYNLVYG-SDFYFIYIYLOAITFIIVGTGFCGTSNS		547
TRIAE CS42 6DS TGACV	FDVPLLEPLKPTECVERTYTPELLLALYLL	CASYDYVIG-SOTYFMYTYLOALAFIVIGEGEVGMKTPCS		531
TRIAE CS42 6BS TGACV	FDVPLIEPIKPTECVERIVIPELLALVIL	CASYDYVIC-SOTYFMYTYLOALAFIVLGEGEVGPKTPCS		528
TRIME_CS42_OBD_TOMEV	FEVELLEDI KOTECVERTVIDELLIAIVII	CASYDYVIC-SOTYFMYIYIOAIAEIVICECEVCMKTPCS		528
TRIAE_CS42_0AS_IGACV	DEVIDER DRI DECVERTITEEDERDIDE.	ECY SANI ABACKASAASINI AI UCI YEACI CI MEACACSCCC	,)	581
TRIAE_CS42_2BS_IGACV	DIEVELLOKTRIKERDI DODINEDETVEGA ELE	ECY SAMI AEDCKASAAEMI AI UCI YEACI CI MEACACSCCC	·)	581
TRIAE_CS42_2DS_IGACV	DNEVFLLQAIRARLADAVNEREIVESAFLE	FCASINLVFFGRISIIFNLILQGLAFVCLGLNFIGICSCCQ		201
TRIAE_C342_2AS_IGACV	DNEVFLLQKIKKERDKVNFKEIVFSAFLF	FCASINLVFFGRIRIFNLILQGLAFVCLGLNFIGICSCCQ		500
TRIAE CS42 2AS TGACV	DNKVPLLQKIKKRLRDKVNFPEILFSAFLF	FCASINLVFPGATSIIFNLILQGLAFAFLGLNFSGTCTCFQ		001
IRIAE_C342_2D3_IGACV	DNRVFLLQRIKKKLKDRVNFFLILFSAFLF	CASINDVFFGRISIIFNDIFQGLAFAFLGLNFIGICICFÇ	2	001
		EGG		
TRIAL COME SAS TGACV		025		
TRIAL COME OF TRACVI		520		
TRIAL CS42_3B_TGACVI		570		
INIAE CS42 3DS TGACV		500		
TRIAL CS42 3DS TGACV		570		
TRIAL COM2 DL TGACV		571		
TRIAL COM2_OB_TGACVI		571		
TRIAE_CS42_JAL_TGACV		575		
TRIAE_CS42_6BS_TGACV		810		
TRIAE_CS42_6AS_TGACV		510		
TRIAE_CS42_/AL_TGACV		810		
TRIAE_CS42_U_TGACv1_		512		
TRIAE_CS42_U_TGACv1_		522		
TRIAE_CS42_/BL_TGACV		3 / 3		
TRIAE_CS42_7AL_TGACv		575		
TRIAE_CS42_7BL_TGACv		515		
		515 515		
TRIAE_CS42_7DL_TGACv	LSSVWDLLGHLSPTDNALRSWCNVVRKDSV	515 515 555		
TRIAE_CS42_7DL_TGACv TRIAE_CS42_2AL_TGACv	LSSVWDLLGHLSPTDNALRSWCNVVRKDSV	515 515 515 555 527		
TRIAE_CS42_7DL_TGACv TRIAE_CS42_2AL_TGACv TRIAE_CS42_2DL_TGACv	LSSVWDLLGHLSPTDNALRSWCNVVRKDSV	515 515 515 555 527 548		
TRIAE_CS42_7DL_TGACv TRIAE_CS42_2AL_TGACv TRIAE_CS42_2DL_TGACv TRIAE_CS42_2DL_TGACv TRIAE_CS42_2BL_TGACv	LSSVWDLLGHLSPTDNALRSWCNVVRKDSV	515 515 555 527 548 528		
TRIAE_CS42_7DL_TGACv TRIAE_CS42_2AL_TGACv TRIAE_CS42_2DL_TGACv TRIAE_CS42_2BL_TGACv TRIAE_CS42_2BL_TGACv TRIAE_CS42_1AS_TGACv	LSSVWDLLGHLSPTDNALRSWCNVVRKDSV	515 515 555 527 548 528 515		
TRIAE CS42 7DL TGACV TRIAE CS42 2AL TGACV TRIAE CS42 2DL TGACV TRIAE CS42 2DL TGACV TRIAE CS42 2BL TGACV TRIAE CS42 1AS TGACV TRIAE CS42 7DS TGACV	LSSVWDLLGHLSPTDNALRSWCNVVRKDSV	515 515 555 527 548 528 515 515 545		
TRIAE_CS42_7DL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2BL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7DS_TGACV TRIAE_CS42_7AS_TGACV	LSSVWDLLGHLSPTDNALRSWCNVVRKDSV	515 515 555 527 548 528 515 545 545 551		
TRIAE CS42 7DL TGACW TRIAE CS42 2AL TGACW TRIAE CS42 2DL TGACW TRIAE CS42 2DL TGACW TRIAE CS42 1AS TGACW TRIAE CS42 1AS TGACW TRIAE CS42 7DS TGACW TRIAE CS42 7BS TGACW	LSSVWDLLGHLSPTDNALRSWCNVVRKDSV	515 515 555 527 548 528 515 545 545 547		
TRIAE CS42 7DL TGACW TRIAE CS42 2AL TGACW TRIAE CS42 2DL TGACW TRIAE CS42 2DL TGACW TRIAE CS42 1AS TGACW TRIAE CS42 7AS TGACW TRIAE CS42 7AS TGACW TRIAE CS42 7BS TGACW TRIAE CS42 6DS TGACW	LSSVWDLLGHLSPTDNALRSWCNVVRKDSV	515 515 555 527 548 528 515 545 551 545 551 547 531		
TRIAE CS42 7DL TGACW TRIAE CS42 2AL TGACW TRIAE CS42 2DL TGACW TRIAE CS42 2BL TGACW TRIAE CS42 2BL TGACW TRIAE CS42 7DS TGACW TRIAE CS42 7DS TGACW TRIAE CS42 7DS TGACW TRIAE CS42 6DS TGACW TRIAE CS42 6DS TGACW	LSSVWDLLGHLSPTDNALRSWCNVVRKDSV	515 515 555 527 548 528 515 545 545 551 547 531 528		
TRIAE_CS42_7DL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2BL_TGACV TRIAE_CS42_2BL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7DS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_6DS_TGACV TRIAE_CS42_6BS_TGACV	LSSVWDLLGHLSPTDNALRSWCNVVRKDSV	515 515 555 527 548 528 515 545 551 547 531 528 528		
TRIAE_CS42_7DL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2BL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7DS_TGACV TRIAE_CS42_7DS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_6AS_TGACV TRIAE_CS42_2BS_TGACV	LSSVWDLLGHLSPTDNALRSWCNVVRKDSV	515 515 555 527 548 528 551 545 551 545 551 547 531 528 528 528 528		
TRIAE CS42 7DL TGACV TRIAE CS42 2AL TGACV TRIAE CS42 2DL TGACV TRIAE CS42 2DL TGACV TRIAE CS42 2BL TGACV TRIAE CS42 7AS TGACV TRIAE CS42 7AS TGACV TRIAE CS42 7BS TGACV TRIAE CS42 6BS TGACV TRIAE CS42 6BS TGACV TRIAE CS42 6BS TGACV TRIAE CS42 2BS TGACV TRIAE CS42 2DS TGACV	LSSVWDLLGHLSPTDNALRSWCNVVRKDSV	515 515 527 548 528 515 545 545 545 547 531 528 528 528 528 528 528 528 528		
TRIAE CS42 7DL TGACW TRIAE CS42 2DL TGACW TRIAE CS42 2DL TGACW TRIAE CS42 2DL TGACW TRIAE CS42 1AS TGACW TRIAE CS42 7DS TGACW TRIAE CS42 7DS TGACW TRIAE CS42 7DS TGACW TRIAE CS42 6DS TGACW TRIAE CS42 6DS TGACW TRIAE CS42 6DS TGACW TRIAE CS42 2DS TGACW TRIAE CS42 2DS TGACW TRIAE CS42 2DS TGACW	LSSVWDLLGHLSPTDNALRSWCNVVRKDSV	515 515 555 527 548 528 515 545 545 547 531 528 528 541 528 528 528 541 528 528 528 521 545 545 545 545 545 545 545 54		
TRIAE CS42 7DL TGACV TRIAE CS42 2AL TGACV TRIAE CS42 2DL TGACV TRIAE CS42 2DL TGACV TRIAE CS42 1AS TGACV TRIAE CS42 7DS TGACV TRIAE CS42 7DS TGACV TRIAE CS42 7DS TGACV TRIAE CS42 6DS TGACV TRIAE CS42 6DS TGACV TRIAE CS42 6DS TGACV TRIAE CS42 2DS TGACV	LSSVWDLLGHLSPTDNALRSWCNVVRKDSV	515 515 555 527 548 528 515 545 545 551 547 531 528 528 528 528 528 528 528 528		
TRIAE CS42_7DL TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_7AS_TGACV TRIAE_CS42_7BS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_6BS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV	LSSVWDLLGHLSPTDNALRSWCNVVRKDSV	515 515 527 548 528 515 545 545 547 531 547 531 528 528 528 528 528 528 528 528		

Appendix 6.2 List of *CslC* subfamily genes, their protein length (amino acids) and multiple sequence alignment of amino acids showing the conserved motifs (D, D, DXD, QXXRW) specific to *Cellulose synthase like* (*Csl*) family (highlighted with red boxes).

S.No	Gene name with number of splice variants (CslC)	No. of amino acids (aa)
1	TRIAE CS42 1DL TGACv1 062162 AA0209740.1	690 aa
2	TRIAE_CS42_1BL_TGACv1_030501_AA0092480.1	656 aa
3	TRIAE_CS42_5BL_TGACv1_404820_AA1311790.1	712 aa
4	TRIAE CS42 5DL TGACv1 435778 AA1454840.1	708 aa
5	TRIAE CS42 5AL TGACv1 374268 AA1195590.3	703 aa
6	TRIAE_CS42_1DL_TGACv1_061928_AA0205730.1	702 aa
7	TRIAE_CS42_1BL_TGACv1_030750_AA0099830.1	702 aa
8	TRIAE_CS42_1AL_TGACv1_001272_AA0028090.1	702 aa
9	TRIAE CS42 3DL TGACv1 251593 AA0882850.1	704 aa
10	TRIAE CS42 3AL TGACv1 197197 AA0665370.1	704 aa
11	TRIAE CS42 3DS TGACv1 271926 AA0910940.1	758 aa
12	TRIAE CS42 3B TGACv1 220758 AA0718310.2	751 aa
13	TRIAE CS42 3AS TGACv1 211225 AA0686890.2	750 aa

TRAF COAS 1DT TCACT		0
TRIAL_CS42_IDL_IGACV		20
TRIAE_CS42_IBL_IGACV		20
TRIAE_C342_TAL_IGACV		10
TRIAE_CS42_3DL_TGACV	MAPWWGQEARGGVSGGVTGTPVVVVKMOTTDWALSEVPPPGSPAAGGKDGRGKNARQITWVLLLK 6	,4
TRIAE_CS42_3AL_TGACv	MAPWWGQEARGGVSGGVTGTPVVVKMOTPDWAHSEVPPPGSPAAGGKDGRGKNAKOITWULLIK 6	,4
TRIAE_CS42_1DL_TGACv	MAPWNGLWGGRAAIAGGN-AYRDMPVIVKWMNPNWSMSEINGGGDNGEDFLARVGGQRRRVKNTKQITWVFRLK 7	3
TRIAE_CS42_1BL_TGACv	QRR <mark>VKN</mark> TKQITKVFI 4	15
TRIAE_CS42_5BL_TGACv	MAPWTGLWGARAGAGAGAGAGAYRGTPVVVKM2NBNWSISEISPEDAEDEDFLVSGAGAARRSRKGGRGKNAKQITMVLLLK 8	10
TRIAE_CS42_5DL_TGACv	MAPWTGLWGARAGAGAGAYRGTPVVVKMENBNWSISEISPEDAEDEDFLVSGAGAARR-RKGGRGKNAKQITWVLLLK 7	17
TRIAE CS42 5AL TGACV	MAPWTGLWGARAGAGAYRGTPVVVKMENBNWSISEISPEDAEDEDFLVSGAARRKGGRGKNAKQITWVLLIK 7	12
TRIAE CS42 3DS TGACv	masswwgdkeehgtpvvvkMonPyslyeidgpgmdssekarrsknakofkwvllIIR 5	i6
TRIAE CS42 3B TGACv1	MASSWWGDKEEHGTPVVVKMONEYSLWEIDGPGMDSSEKARRSKNAKOFKWVLLLR 5	i6
TRIAE CS42 3AS TGACV	MASSWWGDKEEHGTPVVVKWONEYSLWEIDGPGMDSSEKARRSKNAKOFKWVLLLR 5	<i>i</i> 6
TRIAE CS42 1DL TGACV	AHRAAGRITGAASAALAWAAAARREVAAGRIDGDAAPGESTALRARFYGCIRUFWULSMLLIAWVAAYLOG 1	40
TRIAE CS42 1BL TGACV	AHRAACRITCAASAALAVAAAARREVAACETDGDAAPGESTALBARFYGCLRIFYVLSMLLLAVEVAAYLOG 1	40
TRIAF CS42 1AL TGACT		10
TRIAL_CO42_IAL_IGACV		. 11
TRIAE_CS42_SDL_IGACV		. 4 4
TRIAE_CS42_SAL_IGACV		.44
TRIAE_CS42_IDL_TGACV		.52
TRIAE_CS42_IBL_TGACV		.24
TRIAE_CS42_5BL_TGACV	AHRAAGCIASIASIASIASIAVIT GAMARIKKADGATDADAGAPG-SAGESPVLRSRFYAFTKAFTLLSLLLDAVDIAARIHGII	.56
TRIAE_CS42_5DL_TGACV	AHRAAGCIASIASIASIASIAVILGAMARRIVADGUTDADAGATPGSAGESPVLRSRFYAFIRAFILLSLLLDAVELAARFHG I	.54
TRIAE_CS42_5AL_TGACv	AHRAAGCIASIASIAVTIGAMAR <mark>RRV</mark> ADG <mark>R</mark> TDADAGAPG-PARESPVLRSRFYAFIRAFILLSLLLLAVELAARFHR 1	.48
TRIAE_CS42_3DS_TGACv	AHRAVGCVAWLAGGFWGLLGAVN <mark>RRV</mark> RRS <mark>R</mark> DADAEPDAEASGRGRHMLGFLRAFTLLSLAMLAFBTAAYLKG 1	.28
TRIAE_CS42_3B_TGACv1	AHRAVGCVAWLAGGFWGLLGAVN <mark>RRV</mark> RRS <mark>R</mark> DADAEPDAEASGRGRHMLGFLRAFULLSLAMLAFBTAAYLKG 1	.28
TRIAE_CS42_3AS_TGACv	AHRAVGCVAWLAGCFWGLLGAVNRRVRRSRDADAEPDAEASGRGRHMLGFLRAFILLSLAMLAFETAAYLKG 1	.28
TRIAE CS42 1DL TGACV	WHLQMPEMPEMPEQLAMDGLLAVDGLAASAYAGMMRVRLQVLAPPLQ 1	87
TRIAE CS42 1BL TGACV	WHLOMPEMPEMPEMPEGOLANDGLLAVDGLAAAAYAGWMRVRLOYTAPPLO 1	87
TRIAE CS42 1AL TGACV	WHLOMPOMPEMPGOLAMDGLLAVDGLAAAAYAGWMRVRLOYTAPPLO 1	87
TRIAE CS42 3DL TGACV	WI.PDI.EAVECI.FAAGYAAMMRABAAYIGPALO 1	77
TRIAE CS42 JAL TGACV		80
TRIAE CS42 1DL TGACV	GRVNLATEINSENTSOIREBATYVAPPIO 1	81
TRIAE CS42 1BL TGACV		53
TRIAE CS42 5BL TGACV		91
TRIAL_CO42_DDL_IGACV		90
TRIAE_CS42_SDL_IGACV		.03
TRIAE_CS42_SAL_IGACV		.03
TRIAE_CS42_3DS_TGACV	WHIF PROLPEHILKQLEEHLQNLEEHLKHLPENLKHLPEOLKHLPDGLEMPEQQEIQGWLHKAIVAWLAFKIDIIAWA E	108
TRIAE_CS42_3B_TGACVI	WHYFPRDLPEHILRQLEEHLQNLPENLRHLPENLRHLPDGLRMPEQQEIQGWLHRAYVAWLAFRIDYIAWAIE 2	TOT
TRIAE_CS42_3AS_TGACv	WHYFPRDLPEHYLRQLPEHLQNLPEHLRHLPENLRHLPDGLRMPEQQEIQGWLHRAYVAMLAFRUDYIAWAIE 2	:01
TRIAE CS42 1DL TGACV	FITNSCVVLEMIQSVDRIVLCLCCLWIKIRGIKPVPIAADKDDVBAGDEDBPMVLVQMPMCNE 2	250
TRIAE CS42 1BL TGACV	FTINSCVVLEMIQSVDRIILCLGCLWIKLRGIKPVPIAADKDDVBAGEEDAPMVLVOMPMCNE 2	250
TRIAE CS42 1AL TGACV	FUTNSCVVLEMIQSVDRIVLCLGCLWIKLRGIKPVPIAADKDDVBAGDEDPPWVLVOMPMCNE 2	250
TRIAE CS42 3DL TGACV	FTTNACVVLFMIQSADRIILCLGCFWIKLRGIRPVPNAAAAAGNGNGKGSDDVEAGAQDEGDPPMVLVQIPMCNE 2	252

TRIAE_CS42_3AL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_5BL_TGACv TRIAE_CS42_5DL_TGACv	FITNACVVLFMIQSADRIILCL ILADACVVLFIVQSADRIFQSL ILANACVVLFIVQSADRIFQSL FITDACVVLFIVQSADRIIQCL FITDACVVLFIQSADRIIQCL	GCFNIKLRGIREVPNAAT. GCFYILVKRIKEKPLSPALA- GCFYILVKRIKEKPLFLALS- SSFYITVKRIKETLKSPALP- SSFYITVKRIKERLKSPALP-	AGNGKGSDDVEAGAQEEEGE DAEDPDAG DAEDPDAG 	FPMVLVQIPMCNE YPMVLVQIPMCNE YPMVLVQIPMCNE YPMVLVQIPMCNE YPMVLVQIPMCNE	254 245 217 255 253
TRIAE_CS42_5AL_TGACv TRIAE_CS42_3DS_TGACv TRIAE_CS42_3B_TGACv1 TRIAE_CS42_3AS_TGACv	FUTDACVVLELIQSADR.IQCL KLSGFCIVLFMVQSIDRILLCL KLSGFCIVLFMVQSIDRILLCL KLSGFCIVLFMVQSIDRILLCL	GSFYITVKRIKERLRSPALP- GCFWIKLRGIKPGLKAAANKR GCFWIKLRGIKEGLKAAASKR GCFWIKVRGIKEGLVATK-KR	GSKYADDDDLEDGDDLGAY GSKYADDDDLEDGDDLGAY GSKYADDNDLEDGDDLGAY GNKYADDNDLEDGDDLGAY	YPMVLVQIPMCNE FPMVLLQMPMCNE FPMVLLQMPMCNE FPMVLLQMPMCNE	247 283 276 275
TRIAE_CS42_1DL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_3DL_TGACv TRIAE_CS42_3AL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_5BL_TGACv TRIAE_CS42_5DL_TGACv TRIAE_CS42_5AL_TGACv TRIAE_CS42_3DS_TGACv1 TRIAE_CS42_3AS_TGACv	REVYQQSIGAICAL DWPRSNFI REVYQQSIGAICAL DWPRSNFI REVYQQSIGAICAL DWPRSNFI KEVYQQSIGAVCNL DWPRSNFI KEVYQQSIGAVCNL DWPRSNFI KEVYQQSIAAVCNL DWPRSNFI KEVYQQSIAAVCNL DWPRSNFI KEVYQQSIAAVCNL DWPRSNFI KEVYQQSIAAVCNL DWPRSNFI KEVYQQSIAAVCNL DWPRSNFI KEVYQDSIAAVCNL DWPRSNFI KEVYQTSISHVCQL DWPRDMI KEVYETSISHVCQL DWPRDMI	VQVLDDSDDATTSALIKE EVE VQVLDDSDDATTSALIKE EVE VQVLDDSDDATTSALIKE EVE VQVLDDSDDAATSALIRE EVE VQVLDDSDDAATSALIRE EVE VQVLDDSDDVATQALIKE EVE VQVLDDSDDPTTQSLIRE EVA VQVLDDSDDPTTQSLIRE EVA VQVLDDSDDPTTQSLIRE EVA VQVLDDSDDPTTQSLIRE EVA VQVLDDSDDETCQMLIRAEVT VQVLDDSDDETCQMLIRAEVT	KWQREGVRIVYRHRVIRDGYKAGN KWQREGVRIVYRHRVIRDGYKAGN KWQREGVRIVYRHVIRDGYKAGN KWQREGVRIJYRHRVIRDGYKAGN KWRHSGAHIVYRHRVIRDGYKAGN KWRHSGAHIVYRHRVIRDGYKAGN KWQQTGARIIYRHRVIRDGYKAGN KWQQTGARIIYRHRVIRDGYKAGN KWQQTGARIIYRHRVIRDGYKAGN KWQQTGARIIYRHRVIRDGYKAGN KWQQTGARIIYRHRVIRDGYKAGN KWQQTGARIIYRHRVIRDGYKAGN KWQQTGARIIYRHRVIRDGYKAGN	ILKSAMNGSYVKDY ILKSAMNGSYVKDY ILKSAMNGSYVKDY ILKSAMNGSYVKDY ILKSAMSGSYVKDY ILKSAMSGSYVKDY ILKSAMAGSYVKDY ILKSAMAGSYVKDY ILKSAMAGSYVKDY ILKSAMSGPYVKDY ILKSAMSGPYVKDY ILKSAMSGPYVKDY	330 330 332 334 325 297 335 333 327 363 356 355
TRIAE_CS42_1DL_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_3AL_TGACV TRIAE_CS42_3AL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_5BL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3AS_TGACV	EYVVIFDADFQP01D :LKRAMP EYVVIFDADFQP01D :LKRAMP EYVVIFDADFQP01D :LKRAMP EFVVIFDADFQP01D :LKLTVP EFVVIFDADFQP1D :LKLTVP EYVAIFDADFQP1D :LKRTVP EYVAIFDADFQP1D :LKRTVP EFVAIFDADFQP1D :LKRTVP EFVAIFDADFQP1D :LKRTVP EFVAIFDADFQP1D :LKRTVP EFVAIFDADFQP1D :LKLTVP EFVAIFDADFQP1D :LKLTVP EFVAIFDADFQP1D :LKLTVP	HFKGKDDVGLVQARWSFVNID HFKGKDDVGLVQARWSFVNID HFKGKDDVGLVQARWSFVNID HFKGKEDVGLVQARWSFVNID HFKDNEDIGLVQARWSFVNID HFKDNEDIGLVQARWSFVNID HFKDNDDIGLVQARWSFVNID HFKDNDDIGLVQARWSFVNID HFKDNDDIGLVQARWSFVNID HFKGNPDIGLVQARWSFVNID HFKGNPDIGLVQARWSFVNID HFKGNPIGLVQARWSFVNID HFKGNPIGLVQARWSFVNID	NLLTRLQNINLCFHFEVEQQVNG NLLTRLQNINLCFHFEVEQQVNG NLLTRLQNINLCFHFEVEQQVNG NLLTRLQNINLCFHFEVEQQVNG NLLTRLQNINLCFHFEVEQQVNG NLLTRLQNINLCFHFEVEQQVNG NLLTRLQNINLCFHFEVEQQVNG NLLTRLQNINLCFHFEVEQQVNG NLLTRLQNINLCFHFEVEQQVNG NLLTRLQNINLCFHFEVEQQVNG NLLTRLQNINLCFHFEVEQQVNG NLLTRLQNINLCFHFEVEQQVNG	AFLNFFGFNGTAG AFLNFFGFNGTAG AFLNFFGFNGTAG AFLNFFGFNGTAG VFINFFGFNGTAG VFINFFGFNGTAG VFLNFFGFNGTAG VFLNFFGFNGTAG IYLNFFGFNGTAG IYLNFFGFNGTAG	$\begin{array}{c} 410\\ 410\\ 412\\ 414\\ 405\\ 377\\ 415\\ 413\\ 407\\ 443\\ 436\\ 435 \end{array}$
TRIAE_CS42_1DL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_3AL_TGACv TRIAE_CS42_3AL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_5BL_TGACv TRIAE_CS42_5AL_TGACv TRIAE_CS42_3DS_TGACv TRIAE_CS42_3AS_TGACv1 TRIAE_CS42_3AS_TGACv	VWRIKALE DSGGWMERTTVE DM VWRIKALE DSGGWMERTTVE DM VWRIKALE DSGGWMERTTVE DM VWRIKALE DSGGWMERTTVE DM VWRIKALE DSGGWMERTTVE DM VWRIKAVE DSGGWMERTTVE DM VWRIKALE DSGGWMERTTVE DM VWRIKALE DSGGWMERTTVE DM VWRIKALE DSGGWMERTTVE DM VWRIKALE DSGGWMERTTVE DM VWRIEALE DSGGWMERTTVE DM VWRIEALE DSGGWMERTTVE DM VWRIEALE DSGGWMERTTVE DM	D LAVRAHLKGKKFLYLNDVEC D AVRAHLKGKKFLYLNDVEC D AVRAHLKGKKFLYLNDVEC D AVRAHLKGKKFLYLNDVEC D AVRAHLKGKKFLYLNDVEC D AVRAHLKGKFFLYLNDVEC D AVRAHLKGKFFFLNDVEC D AVRAHLHGKKFFFLNDVEC D AVRAHLHGKKFFFLNDVEC D AVRAHLHGKKFFFLNDVEC D AVRAHLHGKKFFFLNDVEV D AVRAHLLGKKFFYLNDVKV D AVRAHLQGKFFYLNDVKV	CCLPESYEAYRKQHRWHSGPMQ CCLPESYEAYRKQHRWHSGPMQ CCLPESYEAYRKQHRWHSGPMQ CCLPESYEAYRKQHRWHSGPMQ CCLPESYEAYRKQHRWHSGPMQ CCLPESYEAYRKQHRWHSGPMQ CCLPESYEAYRKQHRWHSGPMQ CCLPESYEAYRKQHRWHSGPMQ CCLPESYEAYRKQHRWHSGPMQ CCLPESYEAYRKQHRWHSGPMQ LCLPESYEAYRKQHRWHSGPMQ LCLPESYEAYRKQHRWHSGPMQ LCLPESYEAYRKQHRWHSGPMQ LCLPESYEAYRKQHRWHSGPMQ	DIFRLOFVDIIKSK DIFRLOFVDIIKSK DIFRLOFVDIIKSK DIFRLOFVDIIKSK DIFRLOIPDIIRCK DIFRLOIPDIIRCK DIFRLOIPDIIKSK DIFRLOIPDIIKSK DIFRLOIPDIIKSK DIFRLOIPAIIKSK DIFRLOIPAIIKSK	490 490 492 494 485 495 495 495 495 493 487 523 516 515
TRIAE_CS42_1DL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_3DL_TGACv TRIAE_CS42_3AL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_5BL_TGACv TRIAE_CS42_5DL_TGACv TRIAE_CS42_3DS_TGACv TRIAE_CS42_3AS_TGACv1 TRIAE_CS42_3AS_TGACv	IGF WKKONLIFLFFLLRKLILP IGF WKKONLIFLFFLLRKLILP IGF WKKONLIFLFFLLRKLILP IGF WKKONLIFLFFLLRKLILP IGF WKKONLIFLFFLLRKLILP IVF WKKANLIFLFFLLRKLILP IVF WKKONLIFLFFLLRKLILP ISV WKKFNLIFLFFLLRKLILP ISV WKKFNLIFLFFLLRKLILP ISV WKKONLIFFLLRKLILP IPL WKKANLWLFFLLRKLILP IPL WKKANLWLFFLLRKLILP	FYSFTLFCVILPMTMFVPEAE FYSFTLFCVILPMTMFVPEAE FYSFTLFCVILPMTMFVPEAE FYSFTLFCVILPMTMFAPEAE FYSFTLFCVILPMTMFVPEAE FYSFTLFCILLPMTMFVPEAE FYSFTLFCILLPMTMFVPEAE FYSFTLFCILLPMTMFVPEAE FYSFTLFCVILPLTMFVPEAE FYSFTLFCVILPLTMFVPEAE FYSFTLFCVILPLTMFVPEAE	LBAWVCY BATMSIMS LPSPKS LPAWVCY BATMSIMS LPSPKS LPAWVCY BATMSIMS LPSPKS LPAWVCY BATMSLINI LPAPKS LPAWVCY BATMSLINI LPAPKS LPDWVCY BVIMSFINI APAPKS LPDWVCY BVIMSFINI APAPKS LPDWVCY BALMSLINI LPSPKS LPDWVCY BALMSLINI LPSPKS LPDWVCY BALMSLINI LPSPKS LPIWVCY BALMSLINI LPAPKS LPIWVCY PMIMSVINI LPAPKS LPIWICY PMIMSVINI LPAPKS LPIWICY PMIMSVINI LPAPKS	F PF I V PY LLFENT F PF I PY LLFENT F PF I I PY LLFENT F PF I I PY LLFENT F PF V PY LLFENT F PF V PY LLFENT	570 570 572 574 565 573 575 573 567 603 596 595
TRIAE_CS42_1DL_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_3DL_TGACV TRIAE_CS42_3AL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_5BL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_3DS_TGACV TRIAE_CS42_3B_TGACV1	MSVTKFNAMI SGLFQLGSAYEW MSVTKFNAMI SGLFQLGSAYEW MSVTKFNAMI SGLFQLGSAYEW MSVTKFNAMI SGLFQLGSAYEW MSVTKFNAMI SGLFQLGSAYEW MSVTKFNAMI SGLFQLGSAYEW MSVTKFNAMI SGLFQLGSAYEW MSVTKFNAMI SGLFQLGSAYEW MSVTKFNAMI SGLFQLGSAYEW MSVTKFNAMI SGLFQLGSAYEW	VVTKKSGRSSEGDIVALVEKH VVTKKSGRSSEGDIVALVEKH VVTKKSGRSSEGDIVALVEKH VVTKKSGRSSEGDIVALVENE VVTKKSGRSSEGDIVALVENE VVTKKSGRSEGDISLAPKG VVTKKSGRSSEGDISLAAAA VVTKKSGRSSEGDISLAAAA VVTKKSGRSSEGDISLAAAA VVTKKAGRISSESDIFLAAEE VVTKKAGRISSESDIFLAAE	TVQQQQRVG TVQQQQRVG KQSKQLRVG KQSKQQRVG LKQLKYG	SAPDL SAPDL SAPDL SAPNL SAPNL SAPNL SAPNL SAPNL SAPSLEA SAPSLEA SAPSLEA KTHQLDNKDLQLK	627 627 629 631 615 581 637 634 629 683 676
TRIAE_CS42_3AS_TGACV	MSVIKENAMVSGLEQLGSSYEW	VVIARAGRISSESDIFAMAEK	TUTATKPAPKLVRGVSEAGLEAWA	AVI.HÖTDNKDTÖTK	0/5

TRIAE_CS42_1DL_TGACv	AGLAAKDSSLPKKDAPKKKQKH	RIYRKELALSF <mark>L</mark> I	lltaaarsvls	AQGIHFYFLLFQGVSF	IVMGLDLIGEQVE	702
TRIAE_CS42_1BL_TGACv	AGLAAKDSSLPKKDAPKKKQKH	RIYR <mark>KELA</mark> LSF <mark>L</mark> I	ll <mark>taa</mark> arsvls	AQGIHFYFLLFQGVSF	LVM <mark>GLDLIG</mark> EQV <mark>E</mark>	702
TRIAE_CS42_1AL_TGACv	AGLAAKDSSLPKKDAPKKKQKH	RIYR <mark>KELA</mark> LSF <mark>L</mark> I	ll <mark>taa</mark> arsvls	AQGIHFYFLLFQGVSF	LVM <mark>GLDLIG</mark> EQV <mark>E</mark>	702
TRIAE_CS42_3DL_TGACv	DSLAAKEELYPKAEPKPKKKKH	RLYR <mark>KELA</mark> LSFLI	ll <mark>taa</mark> arslls	VQGIHFYFLLFQGVSF	LV <mark>V</mark> GLDLIGEQVE	704
TRIAE_CS42_3AL_TGACv	DSLAAKEELYPKSEPKKKKH	RLYR <mark>KELA</mark> LSF <mark>L</mark> I	ll <mark>taa</mark> arslls	VQGIHFYFLLFQGVSF	LVV <mark>GLDLIG</mark> EQV <mark>E</mark>	704
TRIAE_CS42_1DL_TGACv	SVPAINVAIKEQSKAKKESKKY	RIYK <mark>KELA</mark> MSL <mark>LI</mark>	ll <mark>saa</mark> arslls	KQGIHFYFLLFQGISF	'LL <mark>VGLDLIG</mark> QDI <mark>K</mark>	690
TRIAE_CS42_1BL_TGACv	SVPAINVAIKEKLKAKKESKKY	RIYK <mark>KELA</mark> MSL <mark>LI</mark>	LL <mark>SAAIRSLL</mark> S	KQGIHFYFLLFQGISF	LLV <mark>GLDLIG</mark> QDIK	656
TRIAE_CS42_5BL_TGACv	LMVLKEQQPSPKKEGKKQQKKH	RIYK <mark>KELA</mark> LSL <mark>LI</mark>	ll <mark>taa</mark> arsllt	KQGIHFYFLLFQGISF	LL <mark>VGLDLIG</mark> EQV <mark>E</mark>	712
TRIAE_CS42_5DL_TGACv	LMVLKEQ-PSPKKEGKKQQKKH	RIYK <mark>KELA</mark> LSL <mark>L</mark> I	LLTAA <mark>ARSLL</mark> T	KQGIHFYFLLFQGISF	LL <mark>VGLDLIG</mark> EQV <mark>E</mark>	708
TRIAE_CS42_5AL_TGACv	LMVLKEEQASPRKEGKKQ-KKH	RIYK <mark>KELA</mark> LSL <mark>L</mark> I	LLTAA <mark>ARSLL</mark> T	KQGIHFYFLLFQGISF	LL <mark>VGLDLIG</mark> EQV <mark>E</mark>	703
TRIAE_CS42_3DS_TGACv	AQAEEVTSLAAAIKKTSKAKPP	RIFK <mark>KELA</mark> LAFLI	LLIAATRSLLS	AQGLHFYFLLFQGVTF	LV <mark>V</mark> GLDLIGEQVS	758
TRIAE_CS42_3B_TGACv1	AEAEEVTSLAAAIKKTSKAKPP	RIFK <mark>KELA</mark> LAFLI	LLIAATRSLLS	AQGLHFYFLLFQGVTF	LV <mark>V</mark> GLDLIGEQVS	751
TRIAE_CS42_3AS_TGACv	AEAEEVTSLAAAIKKTSKAKPPN	RIFK <mark>KELA</mark> LAF <mark>LI</mark>	LLIAATRSLLS	AQGLHFYFLLFQGVTF	LVV <mark>GLDLIG</mark> EQVS	750

Appendix 6.3 List of *CslD* subfamily genes, their protein length (amino acids) and multiple sequence alignment of amino acids showing the conserved motifs (D, D, DXD, QXXRW) specific to *Cellulose synthase like* (*Csl*) family (highlighted with red boxes).

S.No	Gene name with number of splice variants (CslD)	No. of amino acids (aa)
1	TRIAE CS42 2BS TGACv1 148683 AA0494520.1	1121 aa
2	TRIAE CS42 2DS TGACv1 177279 AA0572180.1	1120 aa
3	TRIAE CS42 2AS TGACv1 114244 AA0365360.1	1120 aa
4	TRIAE CS42 1BL TGACv1 030586 AA0094860.1	1189 aa
5	TRIAE CS42 1AL TGACv1 001700 AA0034150.2	1146 aa
6	TRIAE CS42 1DL TGACv1 063091 AA0223780.1	1014 aa
7	TRIAE CS42 1BS TGACv1 049706 AA0160220.1	330 aa
8	TRIAE CS42 5BS TGACv1 425241 AA1392650.1	1022 aa
9	TRIAE CS42 5DS TGACv1 457675 AA1488780.1	989 aa
10	TRIAE CS42 7BL TGACv1 577301 AA1871610.1	994 aa
11	TRIAE_CS42_7AL_TGACv1_559436_AA1799630.1	993 aa
12	TRIAE_CS42_7DL_TGACv1_603510_AA1985050.1	994 aa

TRIAE_CS42_1BL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2AS_TGACv TRIAE_CS42_2BS_TGACv TRIAE_CS42_7BL_TGACv	MGSKGILKNSGSSRMPPHGPSKPPTAPTSAPQVVFGRRTESGRFISYSRDDLDS-EISSV MGSKGILKNSGSSRVPPHGPSKPPTAPTSAPQVVFGRRTESGRFISYSRDDLDS-EISSV 	59 59 51 51 51 51 3
TRIAE CS42 7DL TGACV	······ MAS	3
TRIAE CS42 7AL TGACV	MAS	3
TRIAE_CS42_1BS_TGACv		0
TRIAE_CS42_5BS_TGACv	${\tt MSRRLSLPAGSPVTVTVSPTKGKGAGGGSPGDGVVRRGSGLTSPVPRHSIGSSTATLQVSPVRRSGGSRYASRDGADASA}$	80
TRIAE_CS42_5DS_TGACv	${\tt MSRRLSLPASSPVTVTVSPTRGKGAGGGSPGDGVVRRGSGLTSPVPRHSIGSSTATLQVSPVRRSGGSRYASRDGADASA}$	80
TRIAE_CS42_1BL_TGACv	DFQDYHVHIPMTPDNQPMEEDGTKADEQYVSSSLFTGGFNSVTRAHVMDKQGPDSDIGRSGPKGSICMVEGC	131
TRIAE_CS42_1DL_TGACv	DFQDYHVHIPMTPDNQPMEEDGTKADEQYVSSSLFTGGFNSVTRAHVMDKQGPDSDMGRSGPKGSICMVEGC	131
TRIAE_CS42_1AL_TGACV	DFQDYHVHIPMTPDNQPMEEDGTKADEQYVSSSLFTGGFNSVTRAHVMDKQGPDSDMGRSGPKGSICMVEGC	131
TRIAE_CS42_2DS_TGACv	DYANYTYHIPPTPDNQPMKDGAERTAVAMKAEEQYVSNSLFTGGFNSVTRAHLMDRVIDSDVKHPQMAGARPARCAMPAC	131
TRIAE_CS42_2AS_TGACV	DYANYTVHIPPTPDNQPMKDGAEPTAVAMKAEEQYVSNSLFTGGFNSVTRAHLMDRVIDSDVKHPQMAGAKATRCAMPAC	131
TRIAE_CS42_2BS_TGACV	DYANYTVHIPPTPDNQPMKDGSEPTAVAMKAEEQYVSNSLFTGGFNSVTRAHLMDRVIDSDVKHPQMAGAKATRCAMPAC	131
TRIAE_CS42_7BL_TGACv	DHTNYTVFMPPTPDNQPGAAPAPASGGSTKPDNLPLPRYTSGSKLVNRRSGDDGAAGGAKMDRGLS	69
TRIAE_CS42_7DL_TGACv	DHTNYTVFMPPTPDNQPGAAPTPASGGSTKPENLPLP-RYTSGSKLVNRRSGDDGAAGGAKMDRWLS	69
TRIAE_CS42_/AL_TGACV	DHTNYTVFMPPTPDNQPGAASAPASGGPTKPDNLPLPRSS-GSKLVNRRSGDDGAAGGGKMDRRLS	68
TRIAE_CS42_IBS_TGACV	MSCKMRGC	8
TRIAE_CS42_5BS_TGACV	EFVHYTVHIPPTPDRTTASASTDVPAAEEEGEVLPQRSYVSGTIFTGGLNCATRAHVLSNSADGARPAASANMSCKMRGC	160
TRIAE_CS42_5DS_TGACV	EFVHYTVHIPPTPDRNTASASTDAPVAEEEGEVLPQRSYVSGTIFTGGLNCTTRAHVLSNSADGARPAASVNMSCKMRGC	160
TRIAR CS42 1RI TCACT		211
TRIAE_CS42_IDL_IGACV	DSKIMMORGEDII CECDEKICUDOFEDAVKGCBCVCECKEI VKHTEWEFVI SNSSNEI FRAISI DEGGGKMERDI	211
TRIAE_CS42_IDL_IGACV	DSKIMMINGEDTIL DEEDEKTCVDCFTDAVKGGGVCPGCKELYKHTEWEEVI.SNSSNEI.TRALSI.PHGPGGKMERRI.S	211
TRIAE CS42 2DS TGACV		189
TRIAE CS42 2AS TGACV		189
TRIAE CS42 2BS TGACV		189
TRIAE CS42 7BL TGACV		75
TRIAE CS42 7DL TGACV	TEOVAS	75
TRIAE CS42 7AL TGACV		74
TRIAE CS42 1BS TGACV	DMLALAATRPMICEECYMDCVAASGNCPGCKEAYSAGSDTDDSVDEDDDDAISSSEERDOMPMTSMSKRF	78
TRIAE CS42 5BS TGACV	DMPAFLNAGRGGHPPCDCGFMICEECYMDCVAAAGNCPGCKEAYSAGSDTDDSVDEDDDDAISSSEERDOMPMTSMSKRF	240

TRIAE_CS42_1BL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_5DS_TGACV TRIAE_CS42_5DS_TGACV	LVKQGTMNNQSGEPDHNRWLFEIKGTYGYGNAIWEDDNVDDDGRNGVPGHPKELMSKPWRPLT LVKQGTMNNQSGEPDHNRWLFEIKGTYGYGNAIWEDDNVDDDGRNGVPGHPKELMSKPWRPLT LVKQGTMNNQSGEPDHNRWLFESGTYGYGNAIWEDDNVDDDGRNGVPGHPKELMSKPWRPLT AGKSLLARNQNGEPDHNRWLFESSTYGYGNAFMEKGGMYEDDLDEDGAGCDG-GMQDMNQKPFKPLT SGKSLLARNQNGEPDHNRWLFESSTYGYGNAFMEKGGMYEDDLDEDGAGCDGGMPADLSQKPFKPLT PSKSLLVRSQTGEPDHNRWLFESSTYGYGNAFMEKGGMYEDDLDEDGAGCDGGMPADLSQKPFKPLT SGKSLLVRSQT	274 274 256 256 257 139 139 138 145 316 320
TRIAE_CS42_1BL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2AS_TGACv TRIAE_CS42_2BS_TGACv TRIAE_CS42_7BL_TGACv TRIAE_CS42_7AL_TGACv TRIAE_CS42_7AL_TGACv TRIAE_CS42_1BS_TGACv TRIAE_CS42_5BS_TGACv TRIAE_CS42_5DS_TGACv	RKLQIPAAVISPYRLLVLIRLVALAFFLMWRIKHQNDDAIWLWGMSIVCELWFAFSWVLDQLPKLCPINRATDLSVLKEK RKLQIPAAVISPYRLLVLIRLVALAFFLMWRIKHQNDDAIWLWGMSIVCELWFAFSWVLDQLPKLCPINRATDLSVLKEK RKLQIPAAVISPYRLLVLIRLVALAFFLMWRIKHQNDDAIWLWGMSIVCELWFALSWVLDQLPKLCPINRATDLSVLKEK RKIPMPASISPYRIFIVIRFFVLIFYLTWRIRNPNMEALWLWGMSIVCELWFAFSWLLDMLPKVNPINRSTDLAVLKEK RKIPMPTSIISPYRIFIVIRFFVLIFYLTWRIRNPNMEALWLWGMSIVCELWFAFSWLLDMLPKVNPINRSTDLAVLKEK RKIPMPTSISPYRIFIVIRFFVLIFYLTWRIRNPNMEALWLWGMSIVCELWFAFSWLLDMLPKVNPINRSTDLAVLKEK RKIPMPTSISPYRIFIVIRFFVLIFYLTWRIRNPNMEALWLWGMSIVCELWFAFSWLLDMLPKVNPINRSTDLAVLKEK RKVAIPPGILSPYRLLVLVRFVALFLFLIWRATNPNPDAMWLWGISIVCELWFAFSWLLDMLPKVNPINRSTDLAVLKEK RKVAIPPGILSPYRLLVLVRFVALFLFLIWRATNPNPDAMWLWGISIVCEYFALSWLLDQMPKLNPINRAADLAALREK RKVAIPPGILSPYRLLVLVRFVALFLFLVWRATNPNPDAMWLWGISIVCEYFALSWLLDQMPKLNPINRAADLAALREK RKTSVSQAILSPYRMLIAIRLVALGFFLAWRIRHPNPDAMWLWALSVTCEVWFAFSWLLDSLPKLCPVNRSCDLDVLADR RKTSVSQAILSPYRMLIAIRLVALGFFLAWRIRHPNPDAMWLWALSVTCEVWFAFSWLLDSLPKLCPVNRSCDLDVLADR	354 354 336 336 337 219 219 219 218 145 396 400
TRIAE_CS42_1BL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2AS_TGACv TRIAE_CS42_2BS_TGACv TRIAE_CS42_7BL_TGACv TRIAE_CS42_7AL_TGACv TRIAE_CS42_1BS_TGACv TRIAE_CS42_1BS_TGACv TRIAE_CS42_5BS_TGACv TRIAE_CS42_5DS_TGACv	FETPTPSNPTGKSDLPGIDIFVSTADPEKEPVLVTANTILSILAVDYPVDKLACYVSDDGGALLTFEAMAEAASFANFWV FETPTPSNPTGKSDLPGIDIFVSTADPEKEPVLVTANTILSILAVDYPVDKLACYVSDDGGALLTFEAMAEAASFANFWV FETPTPSNPTGKSDLPGLDIFVSTADPEKEPVLVTANTILSILAVDYPVDKLACYVSDDGGALLTFEAMAEAASFANFWV FETPSPSNPHGRSDLPGLDVFVSTADPEKEPVLTTANTILSILAVDYPVEKLACYVSDDGGALLTFEAMAEAASFANIWV FETPSPSNPHGRSDLPGLDIFVSTADPEKEPVLTTANTILSILAVDYPVEKLACYVSDDGGALLTFEAMAEAASFANIWV FETHSPSNPHGRSDLPGLDVFVSTADPEKEPVLTTANTILSILAVDYPVEKLACYVSDDGGALLTFEAMAEAASFANIWV FESKTPSNPTGRSDLPGLDVFVSTADPEKEPVLTTANTILSILAVDYPVEKLACYVSDDGGALLTFEAMAEAASFANIWV FESKTPSNPTGRSDLPGLDVFISTADPYKEPPLVTANTILSILAVDYPVEKLACYVSDDGGALLTFEAMAEAASFANIWV FESKTPSNPTGRSDLPGLDVFISTADPYKEPPLVTANTLLSILATDYPVEKLFVYISDDGGALLTFEAMAEACAYAKVWV FESKTPSNPTGRSDLPGLDVFISTADPYKEPPLVTANTLLSILATDYPVEKLFVYISDDGGALLTFEAMAEACAYAKVWV FESKTPSNPTGRSDLPGLDVFISTADPYKEPPLVTANTILSILATDYPVEKLFVYISDDGGALLTFEAMAEACAYAKVWV FELPTARNPKGRSDLPGIDVFVSTADPEKEPPLVTANTILSILAADYPVEKLACYLSDDGGALLTFEAMAEACAYAKWV	434 434 416 416 417 299 299 298 145 476 480
TRIAE_CS42_1BL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_5BS_TGACV TRIAE_CS42_5DS_TGACV	PFCRKHDIEPRNPDSYFNLKRDPFKNKVKADFVKDRRRIKREYDEFKVRVNGLPDSIRRRSDAYHAREEIQAMNLQREKI PFCRKHDIEPRNPDSYFNLKRDPFKNKVKADFVKDRRRIKREYDEFKVRVNGLPDSIRRRSDAYHAREEIQAMNLQREKI PFCRKHDIEPRNPDSYFNLKRDPFKNKVKADFVKDRRIKREYDEFKVRINGLPDSIRRRSDAYHAREEIQAMNLQREKI PFCKKHDIEPRNPDSYFALKGDPTKGKRRSDFVKDRRKVKREYDEFKVRINGLPDSIRRRSDAFNAREDMKMLKHL PFCKKHDIEPRNPDSYFALKGDPTKGKRRSDFVKDRRKVKREYDEFKVRINGLPDSIRRSDAFNAREDMKMLKHL PFCKKHDIEPRNPDSYFALKGDPTKGKRRSDFVKDRRKVREYDEFKVRINGLPDSIRRSDAFNAREDMKMLKHL PFCKKHDIEPRNPESYFALKGDPTKGKRRSDFVKDRRKVREYDEFKVRINGLPDSIRRSDAFNAREDMKMLKHL PFCKKHSIEPRNPEAYFTQKGDPTKGKRRPDFVKDRRWIKREYDEFKVRINDLPEAIKRRAKAMNAHERKIARETA PFCRKHSIEPRNPEAYFTQKGDPTKGKKRPDFVKDRRWIKREYDEYKVRINDLPEAIKRRAKAMNAHERKIARETA PFCRKHSIEPRNPEAYFTQKGDPTKGKKRPDFVKDRRWIKREYDEYKVRINDLPEAIKRRAKAMNAHERKIARETA PFCRKHSIEPRNPEAYFTQKGDPTKGKKRPDFVKDRRWIKREYDEFKVRINSLTEAIRRRAMAMAHERKIARETA PFCRKHGVEPRCPESYFGQKRDFLKNKVRLDFVRERRKVKREYDEFKVRVNSLTEAIRRRSDAYNAGEELRARRRLQEEA	514 514 492 492 375 375 374 145 556 560
TRIAE_CS42_1BL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2AS_TGACv TRIAE_CS42_2BS_TGACv TRIAE_CS42_7BL_TGACv TRIAE_CS42_7DL_TGACv TRIAE_CS42_7AL_TGACv TRIAE_CS42_1BS_TGACv TRIAE_CS42_5BS_TGACv TRIAE_CS42_5DS_TGACv	KAGGDEQFEPVKIPKATWADSHAPGTWIHSSCD ARGDHAGIIQVMLKPPSDMPMYGNIEK-SPLDFSEVDT KAGGDEQFEPVKIPKATWADSHAPGTWIHSSCD ARGHAGIIQVMLKPPSDMPMYGNIEK-SPLDFSEVDT KAGGDEQFEPVKIPKATWADSHAPGTWIHPSCD ARGHAGIIQVMLKPPSDMPMYGNIEK-SPLDFSEVDT RETGADPSEQPKVKKATWADGHAPGTWAVSSPD AKGHAGIIQVMLRPPSPDPLYGMHDEDQLIDYSDVDT RETGADPSEQPKVKKATWADGHAPGTWAVSSPD AKGHAGIIQVMLRPPSPDPLYGMHDEDQLIDYSDVDT AASSDAAPPVKATWADGHAPGTWAVSSPD AKGHAGIIQVMLRPPSPDPLYGMHDEDQLIDYSDVDT AASSDAAPPVKATWADGHAPGTWAVSSPD AKGHAGIIQVMLRPPSPDPLYGMHDEDQLIDYSDVDT AASSDAAPPVKATWADGHAPGTWADSAFD GKGHASIQVMLRPPSPDPLYG-DADDHAYLDFTNVDV AASSDAAPPVKATWADGHAPGTWADSAFD GKGHASIQVMLKNPHHDVVYG-DADDHAYLDFTNVDV AASSDAAPPVKATWADGHAPGTWADSAFD GKGHASIQVMLKNPHHDVVYG-DADDHAYLDFTNVDV AASSDAAPPVKATWADGHAPGTWADSAFD GKGHASIQVMLKNPHHDVVYG-DADDHAYLDFTNVDV AASSDAAPPVKATWADGHAPGTWADSAFD GKGHASIQVMLKNPHHDVVYG-DADDHAYLDFTNVDV AASSDAAPPVKATWADGHAPGTWADSAFD GKGHASIQVVMLKNPHHDVVYG-DADDHAYLDFTNVDV AASSDAAPPVKATWADGHAPGTWADSAFD GKGHASIQVVMLKNPHHDVVYG-DADDHAYLDFTNVDV AASSDAAPPVKATWADGHAPGTWADSAFD GKGHASIQVVMLKNPHHDVVYG-DADDHAYLDFTNVDV AASSDAAPPVKATWASGSONASTWAACATD ARGHAGIIQ	587 587 566 566 446 446 445 176 636 640
TRIAE_CS42_1BL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_5BS_TGACV TRIAE_CS42_5DS_TGACV	RLPMLVYMSREK RPG D INKKAGAMNAL VRASAL MSNOPE I LNI DOD HYVYNSKAF RECMOEMMDRG DRLC YV OFPORF RLPMLVYMSREK RPG D INKKAGAMNAL VRASAL MSNOPE I LNI DOD HYVYNSKAF RECMOEMMDRG DRLC YV OFPORF RLPMLVYMSREK RPG D INKKAGAMNAL VRASAL MSNOPE I LNI DOD HYVYNSKAF RECMOEMMDRG DRLC YV OFPORF RLPMLVYMSREK RPG D INKKAGAMNAL VRCSAV SNAPE I LNI DOD HY INNNOAVREAMCEMDRG DRLC Y OFPORF RLPMLVYMSREK RPG D INKKAGAMNAL VRCSAV SNAPE I LNI DOD HY INNNOAVREAMCEMDRG DRLC Y OFPORF RLPMLVYMSREK RPG D INKKAGAMNAL VRCSAV SNAPE I LNI DOD HY INNNOAVREAMCEMDRG DRLC Y OFPORF RLPMLVYMSREK RPG D INKKAGAMNAL VRCSAV SNAPE I LNI DOD HY INNNOAVREAMCEMDRG DRLC Y OFPORF RLPMLVYMSREK RPG D INKKAGAMNAL VRCSAV SNAPE I LNI DOD HY INNNOAVREAMCEMDRG DRLC Y I OFPORF RLPMFVYLSREK RPG D INKKAGAMNAL VRCSAV SNAPE I LNI DOD HY INNOAR REAMOEMDRG DRLC Y I OFPORF RIPMFVYLSREK RPG D INKKAGAMNA VRASAVI SNOFFMLNI DOD HY WN COA I REAMOEMDRG DRLC Y I OFPORF RIPMFVYLSREK RPG D INKKAGAMNAL VRCSAV SNAPE I LNI DOD HY WN COA I REAMOEMDRG DRLC Y I OFPORF RIPMFVYLSREK RPG D INKKAGAMNAL VRTSAN I SNOFFMLNI DOD HY WN COA I REAMOEMDRG DRLC Y I OFPORF RIPMFVYLSREK RPG D INKKAGAMNAL VRTSAN SNOFFMLNI DOD HY WN COA I REAMOEMDRG DRLC Y I OFPORF RLPMLVYVSREK RPG D INKKAGAMNAL VRTSAL SNOFFMLNI DOD HY WN COA I REAMOEMDRG DRLC Y I OFPORF RLPMLVYVSREK RPG D INKKAGAMNAL VRTSAL SNOFFMLNI DOD HY WN COA I REAMOEMDRG DRCC V I OFPORF RLPMLVYVSREK RPG D INKKAGAMNAL VRTSAL SNOFFMLNI DOD HY WN SAAL REAMOEMDRG DRCC V OFPORF RLPMLVYVSREK RPG D INKKAGAMNAL VRTSAL SNOFFMLNI DOD HY WN SAAL REAMOEMDRG DRCC V V OFPORF RLPMLVYVSREK RPG D INKKAGAMNAL VRTSAL SNOFFMLNI DOD HY WN SAAL REAMOEMDRG DRCC V V OFPORF RLPMLVYVSREK RPG D INKKAGAMNAL VRTSAL SNOFFMLNI DOD HY WN SAAL RECMOEMDRG DRVC V OFPORF RLPMLVYVSREK RPG D INKKAGAMNAL VRTSAL SNOFFMLNI DOD HY WN SAAL RECMOEMDRG DRVC V OFPORF	667 667 646 646 646 526 525 244 716 720

TRIAE_CS42_5DS_TGACv DMPAFLNAGRGGRPPCDCGFMICEECYMDCVAAAGNCPGCKEAYSAGSDTDDSVDEDDDDAISSSEERDQMPMTSMSKRF 240

TICTUD 0012 TDD 10000	EGIDESDRYANHNTVFFDUNMRALDGLOGEVYVGTGCLERRIAUYGFDPPRSKDHSPGFCGCCLPRRKASASNANPEET	747
TRIAE CS42 1DL TGACV	EGIDPSDRYANHNTVFFDINMRALDGLOGPVYVGTGCLERRIA YGEDPPRSKDHSPGFCGCCLPRRKASASNANPEET	747
TRIAE CS42 1AL TGACV	EGIDPSDRYANHNTVFFDINMRALDGLOGPVYVGTGCLERRIALYGEDPPRSKDHSPGFCGCCLPRRKASASNANPEET	747
TRIAF CS42 2DS TGACT		720
TRIAE_0042_200_TGACV		720
TRIAE_CS42_2AS_IGACV		720
TRIAE_CS42_2BS_TGACV	EGIDPSDRYANHNI VFFDGNRRALDGLOGPMI VGIGGNFRRFALIGEDPPRIAEITGWLFRRRKVINFRDPESD	121
TRIAE_CS42_/BL_TGACV	EGIDPSDRYANHNTVFFDGNMRALDGLQGPMYVGTGCLERRYALYGENPPRAVEYHGLVG-QTRVPIDPHARSG	599
TRIAE_CS42_7DL_TGACv	EGIDESDRYANHNTVFFDCNMRALDELQGPMYVGTGCLERRYAIYGENPERAVEYHGLVG-QTRVPIDPHARSG	599
TRIAE_CS42_7AL_TGACv	EGIDESDRYANHNTVFFDCNMRALDGLQGPMYVGTGCLFRRYAIYGENPPRAVEYHGLVG-QTRVPIDPNARSG	598
TRIAE_CS42_1BS_TGACv	EGIDPNDRYANHNLVFFDVAMRAMDGLQGPMYVGG-CIFRRIALYGFSPPRATKHHGWLG-RKIKLFLRKPTMGKKTDRE	322
TRIAE CS42 5BS TGACv	EGIDPNDRYANHNLVFFDVAMRAMDGLQGPMYVGTGCIFRRTALYGFSPPRATEHHGWLGRKKIKLFLRKPTMGKKTDRE	796
TRIAE CS42 5DS TGACV	EGIDPNDRYANHNLVFFDVAMRAMDGLQGPMYVGTGCIFRRTALYGFSPPRATEHHGWLGRKKIKLFLRKPTTGKKTDRE	800
TRIAE CS42 1BL TGACV	MALENGDEDGDSMNLATEPKKEGNSSELIDSIPVAEFOGRPLADHPSVKNGRPPGALTIPREILDASIVAE	818
TRIAF CS42 1DL TGACT		818
TRIAL_CO42_IDL_IGACV		010
TRIAE_CS42_TAL_IGACV	MALKNIGDE DEDSMINLATE FARE GISSE LIDSTE VALE VGER LADIESV NIGREPEGALI TEKET LDAST VALE	701
TRIAE_CS42_2DS_TGACV	TQLLKAEDFDAELTAQLVPKKFGNSSAMLASIPIAEFQAKPIADHPAVLHGKPPGTLTVPKPPLDPPTVAE	/91
TRIAE_CS42_2AS_TGACV	TQQLKAEDFDAELTAQLVPRRFGNSSAMLASIPIAEFQARPIADHPAVLHGRPPGTLTVPRPPLDPPTVAE	/91
TRIAE_CS42_2BS_TGACv	TQQLKAEDFDAELTAQLVPRRFGNSSAMLASIPIAEFQARPIADHPAVLHGRPPGTLTVPRPPLDPPTVAE	792
TRIAE_CS42_7BL_TGACv	DGVADELRPLSDHPDHEAPQRFGKSKMFIESIAVAEYQGRPLADHPSVRNGRPAGALLMPRPPLDAATVAE	670
TRIAE_CS42_7DL_TGACv	DGIADELRPLSDHPDHEAPQRFGKSKMFIESIAVAEYQGRPLADHPSVRNGRPAGALLMPRPPLDAATVAE	670
TRIAE CS42 7AL TGACv	DGVADELRPLSDHPDHEAPQRFGKSKMFIESIAVAEYQGRPLADHPSVRNGRPPGALLMPRPPLDAATVAE	669
TRIAE CS42 1BS TGACV	LVMAILQK	330
TRIAE CS42 5BS TGACV	SEHESMLPPIEDDDHNOLGDGVRGDLLLPOORTVRHPADEAPAARGLLORGHVPVHLHVPHRLLRAPGRLPLHROVHRPA	876
TRIAE CS42 5DS TGACV	SEHESMLPPTEDDDHNOLGDTESSALMPKRFGSSATFVSSTPVAEYOGRLLODMPGVHOGRPAGALAVPREPLDAATVGE	880
TRIAR CS42 1BL TGACT		898
TRIAL COAL IDI TOACV	A I SUNZOW VERVERWERVOWI VOUTENUTERVERVERVERVERVERVERVERVERVERVERVERVERVE	800
TRIAE_CS42_IDL_IGACV	ALSO VSCWIEDERIEWGIRVGWIIGSVIEDVVIGIRMINRGWSVICVIGRDA FRGTAPINLIDERIGVERWAIGSVEIFF	0 9 0
TRIAE_CS42_IAL_TGACV	AISVVSCWIEEKTEWGIRVGWIIGSVIEDVVTGIRMHNRGWRSVICVIQRDAFRGTAFINLIDRLHQVLRWATGSVEIFF	898
TRIAE_CS42_2DS_TGACV	AVSVISCWYEDKTEWGDRVGWIYGSVTEDVVTGYRMHNRGWRSVYWISKRDAFLGTAPINMTDRLHQVLRWATGSVEIFF	871
TRIAE_CS42_2AS_TGACv	AVSVISCWYEDKTEWGDRVGWIYGSVTEDVVTGYRMHNRGWRSVYWISKRDAFLGTAPINMTDRLHQVLRWATGSVEIFF	871
TRIAE_CS42_2BS_TGACv	AVSVISCWYEDKTEWGDRVGWIYGSVTEDVVTGYRMHNRGWRSVYWISKRDAFLGTAPINMTDRLHQVLRWATGSVEIFF	872
TRIAE CS42 7BL TGACV	AVSVISCWYEDNTEWGLRVGWIYGSVTEDVVTGYRMHNRGWRSVYCITKRDAFRGTAPINLTDRLHQVLRWATGSVEIFF	750
TRIAE CS42 7DL TGACV	AVSVISCWYEDNTEWGLRVGWIYGSVTEDVVTGYRMQNRGWRSVYCITKRDAFRGTAPINLTDRLHQVLRWATGSVEIFF	750
TRIAE CS42 7AL TGACV	AVSVISCWYEDNTEWGLRVGWIYGSVTEDVVTGYRMHNRGWRSVYCITKRDAFRGTAPINLTDRLHOVLRWATGSVEIFF	749
TRIAE CS42 1BS TGACV		330
TRIAE CS42 5BS TGACT	PERHUPRI. PAHHHHHAVPAGAAGDOWURDHAARWUAORAVI.GDRRHORAPGCGAAGPPOGDRRRGHI.I.HAHVOAGRRRRR	956
TRIME_CO12_ODD_TOMOV		060
IRIAE_C542_5D5_IGACV	AISVISCFIEERIEWGRRIGWIIGSVIEDVVIGIRMINNGWRSVICVIRRDAFRGIAFINLIDRL <mark>EUVLRW</mark> AIGSVEFF	900
		0 7 0
TRIAE_CS42_IBL_TGACV	SRNNALFASSKMKVLQRIAYLNVGIYPFTSIFLIVYCFLPALSLFSGQFIVQTLNVTFLTYLLIITITLCLLAMLEIKWS	978
TRIAE_CS42_1DL_TGACv	SRNNALFASSKMKVLQRIAYLNVGIYPFTSIFLIVYCFLPALSLFSGQFIVQTLNVTFLTYLLIITVTLCLLAMLEIKWS	978
TRIAE_CS42_1AL_TGACv	SRNNALFASSKMKVLQRIAYLNVGIYPFTSIFLIVYCFLPALSLFSGQFIVQTLNVTFLTYLLIITVTLCLLAMLEIKWS	978
TRIAE_CS42_2DS_TGACv	SRNNAFLASRKLMFLQRVAYLNVGIYPFTSIFLLTYCFIPALSLFSGFFIVQTLNVAFLFYLLTITVTLIALGILEVKWS	951
TRIAE_CS42_2AS_TGACv	SRNNAFLASRKLMFLQRVAYLNVGIYPFTSIFLLTYCFIPALSLFSGFFIVQTLNVAFLFYLLTITITLIALGILEVKWS	951
TRIAE CS42 2BS TGACV	SRNNAFLASRKLMFLORVAYLNVGIYPFTSIFLLTYCFIPALSLFSGFFIVOTLNVAFLFYLLTITVTLIALGILEVKWS	952
TRIAE CS42 7BL TGACV	SKNNAMLASRRIMFLORMSYTNVGTYPFTSLFLTMYCLLPALSLFSGOFTVATLDPTFLCYLLLTVTLVLLCLLEVKWS	830
TRIAE CS42 7DL TGACV	SKNNAMLASRRIMFLORMSYINVGIYPFTSLFLIMYCLLPALSLFSGOFIVATLDPTFLCYLLLITVTLVLLCLEVKWS	830
TRIAE CS42 7AL TGACV	SKNNALLASRRIMFLORMSYLNVGTYPFTSLFLIMYCLLPALSLFSGOFTVATLDPTFLCYLLLITTTLVLCLEVKWS	829
TRIAF CS42 1BS TGACT		330
TRIAE_CS42_IBS_IGACV		1022
IRIAL CS42 JDS IGACV		TUZZ
	GGGHVKGAVKGAVELPDGAPKDHHDAEKGGAGGDGEDAVQKVPAVEQAAGKKLLQLLGAVPPLPL	000
TRIAE_CS42_5DS_TGACv	GGGHVRGAVRGAVELPDGAPRDHHDBERGGAGGDGEDAVQRVPAVEQAAGRKLLQLLGAVPPLPLSRNNALFATRRMKLLQRVAYFNVGMASSR	989
TRIAE_CS42_5DS_TGACv	GGGHVRGAVRGAVELFDGAFRDHHDÆRGGAGGDGEDAVQRVFAVEQAAGRKLLQLLGAVFFLFLSRNNALFATRRMKLLQRVAYFNVGMASSR	989
TRIAE_CS42_5DS_TGACv TRIAE_CS42_1BL_TGACv	GGGHVRGAVRGAVELEDGAPRDHHDÆRGGGGGDGDGDAVQRVFAVEQAAGRKLLQLGAVFFIFL	989 1058
TRIAE_CS42_5DS_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_1DL_TGACv	GGGHVRGAVRGAVELEDGAPRDHHDAERGGGGGGDGDDAVQRVFAVEQAAGRRELQLGAVFFFFF	989 1058 1014
TRIAE_CS42_5DS_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1AL_TGACv	GGGHYRGAVRGAVELEDGAPRDHHDÆRGGAGGDGDGDAVQRVFAVEQAAGRELQLGAVFFFFF SRNNALFATRRMKLLQRVAYFNVGMASSR	989 1058 1014 1048
TRIAE_CS42_5DS_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_2DS_TGACv	GGGHYRGAVRGAVELEDGAPRDHHDÆRGGAGGDGDGDAVQRVPAVEQAAGRELQLGAVPEDEL SRNNALFATRRMKLLQRVAYFNVGMASSR	989 1058 1014 1048 1021
TRIAE_CS42_5DS_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2AS_TGACv	GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIPPLTIIMVNLV GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIPPLTIIMVNLV GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIP	989 1058 1014 1048 1021 1021
TRIAE_CS42_5DS_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2AS_TGACv TRIAE_CS42_2BS_TGACv	GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFILTSKQVGDDIDDEFAELYEVKWTSLMIPPLTIIMVNLV GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKNSTK	989 1058 1014 1048 1021 1021 1022
TRIAE_CS42_5DS_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2AS_TGACv TRIAE_CS42_2BS_TGACv TRIAE_CS42_7BL_TGACv	GIGLEEWWRNEQFWLISGISAHLYAVVQGLLKVVAGIEISFILTAKAAAEDNEDIYADLYVVKWSSLIPGIGLEEWWRNEQFWLISGISAHLAAVQGLLKVAGIEISFILTAKAAAEDNEDIYADLYVVKWSSLIP	989 1058 1014 1048 1021 1021 1022 900
TRIAE_CS42_5DS_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2AS_TGACv TRIAE_CS42_2BS_TGACv TRIAE_CS42_7BL_TGACv TRIAE_CS42_7DL_TGACv	GGGHVKAVVGAVELEDGAPKDHHDÆRGGAGGDGDDAVGKVFAVEQAAGKKLEQLGAVPFIFL	989 1058 1014 1048 1021 1022 900 900
TRIAE_CS42_5DS_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2AS_TGACv TRIAE_CS42_2BS_TGACv TRIAE_CS42_7BL_TGACv TRIAE_CS42_7DL_TGACv TRIAE_CS42_7AL_TGACv	GGAVKAAVELEDGAPKDHHDÆRGGAGGDGDDAVGKVPAVEQAAGKELQLGAVPFIPL	989 1058 1014 1048 1021 1022 900 900 899
TRIAE_CS42_5DS_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV	GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIPPLTIIMVNLV GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKNSTK	989 1058 1014 1048 1021 1022 900 900 899 330
TRIAE_CS42_5DS_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_5BS_TGACV	GGGHVRAAVRAVELEDGAPKDHHDAERGGAGGGDGDGDAVQRVPAVEQAAGRKELQEDGAVPFIPE	989 1058 1014 1048 1021 1022 900 900 899 330 1022
TRIAE_CS42_5DS_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_5DS_TGACV	GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIPPLTIIMVNLV GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIPPLTIIMVNLV GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIP	989 1058 1014 1021 1022 900 900 899 330 1022 989
TRIAE_CS42_5DS_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2AS_TGACv TRIAE_CS42_2BS_TGACv TRIAE_CS42_7BL_TGACv TRIAE_CS42_7L_TGACv TRIAE_CS42_7L_TGACv TRIAE_CS42_1BS_TGACv TRIAE_CS42_5BS_TGACv TRIAE_CS42_5DS_TGACv	GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIPPLTIIMVNLV GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIPPLTIIMVNLV GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIP	989 1058 1014 1021 1021 1022 900 900 899 330 1022 989
TRIAE_CS42_5DS_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7LL_TGACV TRIAE_CS42_5DS_TGACV TRIAE_CS42_5DS_TGACV	GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIPPLTIIMVNLV GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIPPLTIIMVNLV GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIPP	989 1058 1014 1021 1021 1022 900 900 899 330 1022 989
TRIAE_CS42_5DS_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_5DS_TGACV TRIAE_CS42_5DS_TGACV TRIAE_CS42_1BL_TGACV	GGGHVKAAVELEDGAPKDHHDÆRGGAGGDGDGDAVQKVFAVEQAAGKKLQLDAVFFFF SRNNALFATRRKKLQRVAYFNVGMASSR	989 1058 1014 1021 1021 1022 900 900 900 900 900 900 1022 989
TRIAE_CS42_5DS_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_5BS_TGACV TRIAE_CS42_5DS_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1DL_TGACV	GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIPPLTIIMVNLV GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIPPLTIIMVNLV GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIP	989 1058 1014 1048 1021 1022 900 900 899 330 1022 989 1138 1014
TRIAE_CS42_5DS_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_5DS_TGACV TRIAE_CS42_5DS_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1AL_TGACV	GGGHVRGAVRGAVELEDGAPRDHHDERGGGGGGDGDGDGDAVQRVPAVEQAAGRRELQEDGAVPFEPE	989 1058 1014 1048 1021 1022 900 900 899 330 1022 989 1138 1014 1095
TRIAE_CS42_5DS_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2BS_TGACv TRIAE_CS42_7BL_TGACv TRIAE_CS42_7DL_TGACv TRIAE_CS42_7DL_TGACv TRIAE_CS42_5DS_TGACv TRIAE_CS42_5DS_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_2DS_TGACv	GGGHVKGAVKGAVELEDGAPKDHHDÆRGGGGGGDGDGDGDAVQKVPAVEQAAGKKLLQLDGAVPFIPL	989 1058 1014 1048 1021 1022 900 900 900 899 330 1022 989 1138 1014 1095 1065
TRIAE_CS42_5DS_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2BS_TGACv TRIAE_CS42_7BL_TGACv TRIAE_CS42_7AL_TGACv TRIAE_CS42_1BS_TGACv TRIAE_CS42_5DS_TGACv TRIAE_CS42_5DS_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2AS_TGACv	GGAVKAAVELEDGAFKDHHDÆRGGAGGDGDGDAVQKVFAVEQAAGKKLLQLDAVFFFF SRNNALFATRRKKLLQRVAYFNVGMASSR	989 1058 1014 1048 1021 1022 900 900 899 330 1022 989 1138 1014 1095 1068
TRIAE_CS42_5DS_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_5BS_TGACV TRIAE_CS42_5DS_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV	GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIPPLTIIMVNLV GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIPPLTIIMVNLV GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIP	989 1058 1014 1048 1021 1022 900 900 899 330 1022 989 1138 1014 1095 1068 1068
TRIAE_CS42_5DS_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_7BL_TGACv TRIAE_CS42_7DL_TGACv TRIAE_CS42_7DL_TGACv TRIAE_CS42_7DL_TGACv TRIAE_CS42_1BS_TGACv TRIAE_CS42_1BS_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2BS_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_7BL_TGACv TRIAE_CS42_7BL_TGACv	GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIPPLTIIMVNLV GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIPPLTIIMVNLV GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIP	989 1058 1014 1048 1021 1022 900 900 899 330 1022 989 1138 1014 1095 1068 1069 947
TRIAE_CS42_5DS_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_5DS_TGACV TRIAE_CS42_5DS_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_7DL_TGACV	GGGHVKAAVELEDGAFKDHHDÆRGGGGGGDGDGDAVQKVFAVEQAAGKKLLQLDGAVFFIFL	989 1058 1014 1041 1021 1022 900 900 899 330 1022 989 1138 1014 1095 1068 1069 947 947
TRIAE_CS42_5DS_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_5DS_TGACV TRIAE_CS42_5DS_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7AL_TGACV	GIGLEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIPPLTIIMVNLV GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIPPLTIIMVNLV GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIP	989 1058 1014 1048 1021 1022 900 900 900 1022 989 1138 1014 1095 1068 1069 947 947
TRIAE_CS42_5DS_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2BS_TGACv TRIAE_CS42_7BL_TGACv TRIAE_CS42_7BL_TGACv TRIAE_CS42_7AL_TGACv TRIAE_CS42_1BS_TGACv TRIAE_CS42_1BS_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2BS_TGACv TRIAE_CS42_2BS_TGACv TRIAE_CS42_7BL_TGACv TRIAE_CS42_7BL_TGACv TRIAE_CS42_7AL_TGACv TRIAE_CS42_1BS_TGACv TRIAE_CS42_1BS_TGACv TRIAE_CS42_1BS_TGACv TRIAE_CS42_1BS_TGACv TRIAE_CS42_1BS_TGACv TRIAE_CS42_1BS_TGACv TRIAE_CS42_1BS_TGACv TRIAE_CS42_1BS_TGACv TRIAE_CS42_1BS_TGACv	GGGHVKGAVELEDGAPKDHHDÆRGGGGGGDGDGDGVKVFAVEQAAGKKLLQLDGAVPFIFL	989 1058 1014 1048 1021 1022 900 900 899 330 1022 989 1138 1014 1095 1068 1069 947 947 946 330
TRIAE_CS42_5DS_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_5DS_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7AL_TGACV	GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIPPLTIIMVNLV GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIPPLTIIMVNLV GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIP	989 1058 1014 1048 1021 1022 900 899 330 1022 989 1138 1014 1095 1068 1069 947 947 947 330
TRIAE_CS42_5DS_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_5DS_TGACV TRIAE_CS42_5DS_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_5BS_TGACV TRIAE_CS42_5BS_TGACV TRIAE_CS42_5BS_TGACV TRIAE_CS42_5BS_TGACV TRIAE_CS42_5BS_TGACV TRIAE_CS42_5BS_TGACV TRIAE_CS42_5BS_TGACV TRIAE_CS42_5BS_TGACV TRIAE_CS42_5BS_TGACV	GGGHVKAAVELEDGAFKDHHDÆRGGGGGGDGDGDAVQKVFAVEQAAGKKLQLDGAVFFIFL	989 1058 1014 1021 1022 900 900 899 330 1022 989 1138 1044 1095 1068 1069 947 947 947 943 330
TRIAE_CS42_5DS_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1AL_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2BS_TGACv TRIAE_CS42_7BL_TGACv TRIAE_CS42_7DL_TGACv TRIAE_CS42_7DL_TGACv TRIAE_CS42_5DS_TGACv TRIAE_CS42_5DS_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2DS_TGACv TRIAE_CS42_2BS_TGACv TRIAE_CS42_2BS_TGACv TRIAE_CS42_2BS_TGACv TRIAE_CS42_1DL_TGACv TRIAE_CS42_2BS_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_1BL_TGACv TRIAE_CS42_1BS_TGACv TRIAE_CS42_1BS_TGACv TRIAE_CS42_1BS_TGACv TRIAE_CS42_1BS_TGACv TRIAE_CS42_5BS_TGACv TRIAE_CS42_5DS_TGACv	GGGHVKAVKAVKEAVELPDGAPKDHHDAEKGGAGGDGEDAVQKVPAVEQAAGKKLLQLLGAVPPLFL	989 1058 1014 1021 1022 900 900 899 330 1022 989 1138 1068 1068 1068 1068 1068 1068 1068 106
TRIAE_CS42_5DS_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7L_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_5BS_TGACV TRIAE_CS42_5BS_TGACV TRIAE_CS42_5BS_TGACV TRIAE_CS42_5BS_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_5BS_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1BL_TGACV	GGGHVKAVKAVKEAVELPDGAPKDHHDAEKGGAGGDGEDAVQKVPAVEQAAGKKLLQILGAVPPLFL	989 1058 1014 1048 1021 1022 900 900 899 330 1022 989 1138 1014 1068 1068 1068 1068 1068 1068 1069 947 947 946 330
TRIAE_CS42_5DS_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_1BL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DS_TGACV TRIAE_CS42_5DS_TGACV TRIAE_CS42_5DS_TGACV TRIAE_CS42_5DS_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV	GIGLEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIPPLTIIMVNLV GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIPPLTIIMVNLV GIALEEWWRNEQFWLIGGTSAHLAAVMQGLLKVVAGIEISFTLTSKQVGDDIDDEFAELYEVKWTSLMIP	989 1058 1014 1014 1021 1022 900 900 899 1138 1022 989 1138 1068 1068 1069 947 947 947 946 330 1022 989

TRIAE_CS42_IAL_TGACV	AKGLMGKRGRTPTIVIVWAGLVSITISLLWIAINPPSTAANQQLGGSFSFP-	1146
TRIAE_CS42_2DS_TGACv	AKGLMGRRGKTPTIIFVWSGLISITISLLWVALSPPEANSTGGARGGGFQFP	1120
TRIAE_CS42_2AS_TGACv	AKGLMGRRGKTPTIIFVWSGLISITISLLWVALSPPEANSTGGARSGGFQFP	1120
TRIAE_CS42_2BS_TGACv	AKGLMGRRGKTPTIIFVWSGLISITISLLWVALSPPEANSTGGARGGGFQFP	1121
TRIAE_CS42_7BL_TGACv	AKGLMGRRGRTPTIVYVWAGLISITVSLLWITISPPDDRVSQSGIEV	994
TRIAE_CS42_7DL_TGACv	AKGLMGRRGRTPTIVYVWAGLISITVSLLWITISPPDDRVSQSGIEV	994
TRIAE_CS42_7AL_TGACv	AKGLMGRRGRTPTIVYVWAGLISITVSLLWITISPPDDRVSQSGIEV	993
TRIAE_CS42_1BS_TGACv		330
TRIAE_CS42_5BS_TGACv		1022
TRIAE_CS42_5DS_TGACv		989

Appendix 6.4 List of *CslE* subfamily genes, their protein length (amino acids) and multiple sequence alignment of amino acids showing the conserved motifs (D, D, DXD, QXXRW) specific to *Cellulose synthase like* (*Csl*) family (highlighted with red boxes).

S.No	Gene name with number of splice variants (CslE)	No. of amino acids (aa)
1	TRIAE CS42 6DL TGACv1 526558 AA1687090.1	738 aa
2	TRIAE CS42 6AL TGACv1 471004 AA1500600.1	737 aa
3	TRIAE_CS42_6BL_TGACv1_499967_AA1596110.2	736 aa
4	TRIAE CS42 U TGACv1 683314 AA2158770.1	446 aa
5	TRIAE CS42 5DL TGACV1 433536 AA1415840.1	756 aa
6	TRIAE CS42 5BL TGACv1 406235 AA1342610.1	728 aa
7	TRIAE CS42 5AL TGACv1 376126 AA1232370.2	728 aa
8	TRIAE CS42 5DL TGACv1 433536 AA1415830.1	728 aa
9	TRIAE CS42 5BL TGACv1 406235 AA1342600.1	734 aa
10	TRIAE_CS42_6DS_TGACv1_543277_AA1737920.1	725 aa

Color Align Conservation results	
TRIAE CS42 5DL TGACV MVAIGRRTGQQHGHWRLAAESPPYLGPRDGEEHEAVRDGDSRGPGGVQAPRRHGGRRILLLLYYRATRVPAAGEGRAAWL	80
TRIAE CS42 5BL TGACvMERSRRLFETETHGGRAAYRLHAVTVAAGILLVLYYRATHVPAAGEGRATWL	52
TRIAE CS42 U TGACv1	0
TRIAE CS42 5AL TGACVMERTRLFETETHGGRAAYRLHAVTVAAGILLLLYYRATRVPAAGEGRAAWL	51
TRIAE CS42 6DS TGACv	50
TRIAE_CS42_5DL_TGACv	52
TRIAE_CS42_5BL_TGACvMERSRRLFETETHGGRAVYRLHAVTVAAGILLLLYYRATRVPAAGEGRAAWL	52
TRIAE_CS42_6AL_TGACvMAGSSVSGGGGRPPLFATEKPKRVLAYRVYAGTIFAGILLIWFYRATHIPARGSSSLGWR	60
TRIAE_CS42_6BL_TGACvMAGSSVSGGRPPLFATEKPKRVLAYRLYAGTIFAGILLIWFYRATHIPERGDSSLGWR	59
TRIAE_CS42_6DL_TGACvMAGSSVSGGGGGRPPLFATEKPKRVLAYRLYAGTIFAGILLIWFYRATHIPARGSSSLGWR	61
TRIAE CS42 5DL TGACv GMLAAELWYAAYWVVTOSVRWSPVRRPFIDRLAARHG-ETLPCVDIFVCTADPYSEPPSLVVSTILSLMAYNYPPE	156
TRIAE CS42 5BL TGACV GMLAAELWYAAYWVVTOSVRWSPVRRRPFIDRLAARHG-ERLPCVDIFVCTADPYSEPPSLVVSTILSLMAYNYPPE	128
TRIAE CS42 U TGACv1	0
TRIAE CS42 5AL TGACV GMLAAELWYAAYWVVTQSVRWSPVRRPFRDRLAARHG-ERLPSVDIFVCTADPYSEPPSLVVSTILSLMAYNYPPE	127
TRIAE CS42 6DS TGACV GMLAAELWYAAYWVVTQSVRWSPVRRCTFRDRLTARYG-DRLPGVDIFVCTADPLSEPPSLVISTILSVMAYNYLAE	126
TRIAE CS42 5DL TGACv GMLAAELWYAAYWAVTQSVRWSPVRRLPFIDRLAARYG-ERLPCVDIFVCTADPHSEPPSLVISTVLSLMAYNYPAE	128
TRIAE_CS42_5BL_TGACv GMLAAELCYAAYWVVTQSVRWSPLHRRPCRDRLAARYG-ERLPCVDIFVCTADPHSEPPSLVISTVLSLMAYNYPAE	128
TRIAE_CS42_6AL_TGACv AGLGLLVAEILFGLYWVLTLSVRWNPVRRTTFKDRLSERYDDDQLPGVDIFVCTADPALEPPMLVISTVLSVMAYDYPPE	140
TRIAE_CS42_6BL_TGACv AGLGLLVAELLFGLYWVLTLSVRWNPVRRTTFKDRLSERYDDDQLPGVDIFVCTADPALEPPMLVISTVLSVMAYDYPPE	139
TRIAE_CS42_6DL_TGACv AGLGLLVAELWFGLYWVLTLSVRWNPVRRATFKDRLSERYDDDQLPGVDIFVCTADPALEPPMLVISTVLSVMAYDYPPE	141
TRIAE CS42 5DL TGACv KLSVYLSDDGGSILTFYGMWEASLFAKHWLPFCKRYNIEPRSPAAYFSESDGHQELCNPKEWSLIKDMFDKMTERIDTVV	236
TRIAE CS42 5BL TGACv KLSVYLSDDGGSILTLYGMWEASLFAKHWLPFCKRYNIEPRSPAAYFSESDGHQELCTPKEWSLIKDMFDKMTERIDTAV	208
TRIAE CS42 U TGACv1	0
TRIAE CS42 5AL TGACV KLSVYLSDDGGSILTYYGMWEASLFAKHWLPFCKRYNIEPRSPAAYFSQSDGHQELCTPKEWSLIKDMFDEMTERIDTAV	207
TRIAE_CS42_6DS_TGACv KLSVYLSDDGGSVLTFYAMWEASLFAKHWLPFCKRYNIEPRSPAAYFSESYQDLCTPKEWSFIKDMYDEMTERIDTAV	204
$\texttt{TRIAE}_\texttt{CS42}_\texttt{5DL}_\texttt{TGACv} \texttt{KISVYLSDDGGSVLTFYALWEASLFAKHWIPFCKRYNIEPRSPAAYFSESDGHQDLCSPKEWSLIREMYEDMTERIDTAV}$	208
$\texttt{TRIAE}_\texttt{CS42}_\texttt{5BL}_\texttt{TGACv} KISVYLSDDGGSILTFYALWEASLFAKHWIPFCKRYNIEPRSPATYFSESDGHQDMCTPKEWSLIREMYEDMTERIDTAA$	208
$\texttt{TRIAE}_\texttt{CS42}_\texttt{6AL}_\texttt{TGACv} \\ \texttt{KLNIYLSDDAGSAVTFYALHEASEFAKHWIPFCKNYKVEPRSPAAYFAEGATPHDACSPQELLRMKELYKDLTDRVNSVV}$	220
$\texttt{TRIAE}_\texttt{CS42}_\texttt{6BL}_\texttt{TGACv}~\texttt{KLNIYLSDDAGSAVTFYALHEASEFAKHWIPFCKNYKVEPMSPAAYFAEGATPHDACSPQELLRMKELYKDLTDRVNSVV}$	219
TRIAE_CS42_6DL_TGACv KLNIYLSDDAGSAVTFYALHEASEFAKHWIPFCKNYKVEPRSPAAYFAKGATPHDACSPQEFLRMKELYKDLTDRMNSVV	221
TRIAE CS42 5DL TGACV MSGKVPEEIKASHKGFYEWNQEITSKNHQPIVQILIDSKDQNAVDNEGKVLPTLVYMAREKRPQHHHNFKAGAMNALIRV	316
TRIAE CS42 5BL TGACV MSGKVPEEIKARHKGFYEWNQEISSKNHQPIVQILIDGKDQNAVDNEGKVLPTLVYMAREKRPQHHHNFKAGAMNAL	288
TRIAE CS42 U TGACv1MQIRV	5
TRIAE_CS42_5AL_TGACV MSGKVPEEIKARQKGFHEWNQEITSKNHQPIVQILIDGKDQNAVDNEGNVLPTLVYMAREKRPQHHHNFKAGAMNAL	287
TRIAE CS42 6DS TGACV ISRKIPEEIRSNHKGFYEWNPEITSKNHQPIVQVLIDGKDQKGVDSEGNVLPTLVYMAREKRPQHHHNFKAGAMNAL	284
TRIAE_CS42_5DL_TGACv LSGKISEEVKANHKGFHEWDQENTSKNHQPIVQILIEGKDKNANDDEGNVLPTLVYMAREKRPQHHHNFKAGAMNAL	288
$\texttt{TRIAE}_\texttt{CS42}_\texttt{5BL}_\texttt{TGACv}_\texttt{LSGKISEEVKENHKGFHEWDQENTSKNHQPIVQILIEGKDKNANDDEGNVLPTLVYMAREKRPQHHNFKAGAMNALIRV}$	288
TRIAE_CS42_6AL_TGACv HSGKIPEVPECNHRGSSEWNEMITSGDHPSIVQILIDRNKRKAVDVDGNALPKLVYMSREKRPQEQHHFKAGSLNALIRV	300

TRIAE_CS42_6BL_TGACv TRIAE_CS42_6DL_TGACv	HSGKIPEVPECNHRGFSVWNETITSGDHPSIVQILIDRNKRKAVDVDGNALPKLVYMAREKRPQEQHHFKAGSLNALIRV 29 HSGKIPEVPECNHRGFSEWNETITSGDHPSVVQILIDRNKRKAVDVDGNALPKLVYMAREKRPQEQHHFKAGSLNALIRV 30)9)1
TRIAE_CS42_5DL_TGACv TRIAE_CS42_5BL_TGACv TRIAE_CS42_U_TGACv1_ TRIAE_CS42_6DL_TGACv TRIAE_CS42_6DS_TGACv TRIAE_CS42_5DL_TGACv TRIAE_CS42_5BL_TGACv TRIAE_CS42_6AL_TGACv TRIAE_CS42_6BL_TGACv TRIAE_CS42_6DL_TGACv	SSVISNSPIIMN DCD YSNNNDAV 202 LCFFLDEEYGHKIGFVOYPONYNLSKNNIYGNSLHVINDVEMGG 4D5 LCGP 39 SSVISNSPIIMN DCD YSNNNDAV 202 LCFFLDEEYGHKIGFVOYPONYNLSKNNIYGNSLHVINDVEMGG 4D5 LCGP 36 SSVISNSPIIMN DCD YSNNNDAV 202 LCFFLDEEYGHKIGFVOYPONYNLSKNDIYGNSLHVID VEMGG 4D5 LCGP 36 SSVISNSPIIMN DCD YSNNNDAV 202 LCFFLDEEYGHKIGFVOYPONYNLSKNDIYGNSLOVIN VEMGG 4D5 LCGP 36 SSVISNSPIIMN DCD YSNNNDAV 202 LCFFLDEEYGHKIGFVOYPONYNLSKNDIYGNSLOVIN VEMGG 4D5 LCGP 36 SSVISNSPIIMN DCD YSNNNDAV 202 LCFFLDEEYGHKIGFVOYPONYNLSKNDIYGNSLOVIN VEMGG 4D5 LCGP 36 SSVISNSPIIMN DCD YSNNNDAT 1502 LCFFLDEEYGHKIGFVOYPONYNLTKNNIYGNSHOVINOVLMCG 4D5 VCGP 36 SSVISNSPIIMN DCD YSNNDT 1502 LCFFLDEEYGHKIGFVOYPONYNLTKNNIYGNSHOVINOVLMCG 4D5 VCGP 36 SSVISNSPIIMN DCD YSNNDT 1502 LCFFLDEEYGHKIGFVOYPONYNLTKNNIYGNSHOVINOVLMCG 4D5 VCGP 36 SSVISNSPIIMN DCD YSNNSTSI 202 LCFFLDEEYGHKIGFVOYPONALTKNNIYGNSHOVINOVLMCG 4D5 VCGP 36 SSVISNSSVILN DCD YSNNSSSI 202 LCFFLDEE2G01GFVOYPONALTKNNIYGNSHOVINOVLMCG 4D5 VCGP 36 SSVISNSSVILN DCD YSNNSSSI 202 LCFFLDEE2G01GFVOYPONALTKNNIYMDIYNDIDNCC 4D5 WCGM 37 SSVISNSSVILN DCD YSNNSSSI 202 LCFFLDEE2C01GFVOYPONALTWN YNDIYNDIDNC 4D7 VNDIDNC 4D7 WNDIDNC 4D7 WND)6 58 57 54 58 58 58 58 58 58 59 51
TRIAE_CS42_5DL_TGACV TRIAE_CS42_5BL_TGACV TRIAE_CS42_0_TGACV1_ TRIAE_CS42_0TGACV1_ TRIAE_CS42_6DS_TGACV TRIAE_CS42_6DS_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_6AL_TGACV TRIAE_CS42_6BL_TGACV TRIAE_CS42_6DL_TGACV	MIIGTGCFHRREIICGRKETEDYOEDWAAGIKDKLOES-IDETEEKAKSLAACTYEHGTOMADEIGVKYGCVVEDVATGL 47 MIIGTGCFHRREILCGRKETEDYOEDWAAGIKDKLOES-IDETEEKAKSLAACTYEHGTOMGDEIGVKYGCVVEDVATGF 44 LIIGTGCFHRREIICGRKETKDYOEDWAAGIKDKLOES-IDETEEKAKSLAACTYEHGTOMGDEIGVKYGCAVEDVITGL 16 LIIGTGCFHRREIICGRKETKDYOEDWAAGIKDKLOES-IDETEEKAKSLAACTYEHGTOMGDEIGVKYGCAVEDVITGL 44 MYVGTGCFHRREILGRKETKDYOEDWAAGIKDKLOES-IDETEEKAKSLAACTYEHDTOMGDEIGVKYGCAVEDVITGL 44 MYVGTGCFHRREILGRRETEDYKEDWAAGIKDKLOES-IDEIEEKAKSLAASIYEHDTOMGDEIGVKYGCAVEDVITGL 44 MYVGTGCFHRREILGRRETEDYKEDWAAGIKDKTOES-IVEIEEKAKSLAASIYEHDTOMGDEIGVKYGCAVEDVITGL 44 CYVGTGCFHRREIICGRRETEDYKEDWAGGIKDKTOES-IDEIEEKAKSLAASIYEHDTOMGDEIGVKYGCAVEDVITGL 44 CYVGTGCFHRREIICGRRETEDYKEDWAGGIKDKTOES-IDEIEEKAKSLAASIYEHDTOMGDEIGVKYGCAVEDVITGL 44 CYVGTGCFHRREIIGCGRRETEDYKEDWAGGIKDKTOES-IDEIEEKAKSLAASIYEHDTOMGDEIGVKYGCPIEDVITGL 44 CYVGTGCFHRREIIGCGRRETEDYKEDWAGGIKDKTOES-IDEIEEKAKSLAASIYEHDTOMGDEIGVKYGCPIEDVITGL 44 CYVGTGCFHRRETISGOIYSKDYKEDWAGGIKDKTOES-IDEIEEKAKSLASIYCTYEHNTEMGIEKGVRYGCPIEDVITGL 45 CYVGTGCFHRRETISGOIYSKDYKEDWAAGVGIAENADELEETSKSLVTCTYEHNTEMGIEKGVRYGCPIEDVITGL 45 CYVGTGCFHRRETISGOIYSKDYKEDWAAGVGIAENADELEETSKSLVTCTYEHNTEMGIEKGVRYGCPIEDVITGL 45 CYVGTGCFHRRETISGOIYSKDYKEDWAAGVGIAENADELEETSKSLVTCTYEHNTEMGIEKGVRYGCPIEDVITGL 45	75 17 54 16 14 17 56 58
TRIAE_CS42_5DL_TGACv TRIAE_CS42_5BL_TGACv TRIAE_CS42_U_TGACv1_ TRIAE_CS42_6DL_TGACv TRIAE_CS42_6DS_TGACv TRIAE_CS42_6DL_TGACv TRIAE_CS42_5BL_TGACv TRIAE_CS42_6AL_TGACv TRIAE_CS42_6BL_TGACv TRIAE_CS42_6DL_TGACv	A THE REWDSVYNNERR PARTY OFTEL AGT I OHKRW BEGIFSIFLSK NIVELFAHEKTKURHOMEYH YELWAPNSLAT 55 A HEREWDSVYNNEKR PARTY EMEVEPTIL AGT I OHKRW BEGIFSIFLSK NIVELFAHEKTKUCHOMEYH YELWAPNSLAT 52 A HEREWDSVYNNEKR PARTY EMEVEPTIL AGT I OHKRW BEGISIFLSK NIVELFAHEKTKUCHOMEYH YELWAPNSLAT 52 A HEREWDSVYNNEKR PARTY EMEVEPTIL AGT I OHKRW BEGISIFLSK NIVELFAHEKTKURHOMEYH YELWAPNSLAT 52 A HEREWDSVYNNEKR PARTY EMEVEPTIL AGT I OHKRW BEGISIFLSK NIVELFAHEKTKURHOMEYH YELWAPNSLAT 52 CHEREWSVYNNEKR PARTY EMEVEPTIL AGT I OHKRW BEGISIFLSK NIVELFAHEKTKURHOMEY I YELWAPNSLAT 52 CHEREWSVYSNETR PARTY EMEVEPTIL AGT I OHKRW BEGISIFLSK YEPELFEHEKTNIRHOMEYSIYELWAPNSLAT 52 EHEREWKSVHSNEPR PARTY EMEVEPTIL AGT I OHKRW BEGISIFLSK YEPEMFEHEKTNIRHOMEYSIYELWAPNSIPT 52 CHEREWKSVHSNEPR PARTY EMEVEPTIL AGT I OHKRW BEGISIFLSK YEPEMFEHEKTNIRHOMEYSIYELWAPNSIPT 52 OLOCREWRSVYNPARKELEWAPTSLEGI I OHKRW BEGISISISTELSK YEPEMFEHEKTKURHOMEYSIYELWAPNSIPT 52 OLOCREWRSVYNPARKELEWAPTSLEGI I OHKRW BEGISISISTELSK YEPEMFEHEKTIKURHOMEYSIYE MANNSIPT 53 OLOCREWRSVYNPARKELEWAPTSLEGI I OHKRW BEGISISINYSFELLEHEKTKURHOMEYSIYE MANNSIPT 53 OLOCREWRSVYNPARKELEWAPTSLEGI I OHKRW BEGISUSISNISFELLEHEKTKURGLOMEYSIYE MANNSIPT 53 OLOCREWRSVYNPARKELEWAPTSLEGI I OHKRW BEGISUSISNISFELLEHEKTKURGLOMEYSIYE MANNSIPT 53	5 27 14 26 24 27 27 37 36 38
TRIAE_CS42_5DL_TGACv TRIAE_CS42_5BL_TGACv TRIAE_CS42_U_TGACv1_ TRIAE_CS42_5AL_TGACv TRIAE_CS42_6DS_TGACv TRIAE_CS42_5DL_TGACv TRIAE_CS42_5BL_TGACv TRIAE_CS42_6AL_TGACv TRIAE_CS42_6BL_TGACv TRIAE_CS42_6DL_TGACv	I YYVI IPSLALUKÇI SI PPEITSPWI APFVYVFCVKNMY SIYBAULS OTIKGWIN CORMULVKRI ISYLFCVLDNI RKL 63 IYYVI IPSLALUKÇI SI PPEITSPWI APFVYVFCVKNMY SIYBAUSS OTIKGWIN CORMULVKRI ISYLFCVLDNI RKL 64 IYYVI IPSLALUKÇI SI PPEITSPWI APFVYVFCVKNMY SIYBAUSS OTIKGWIN CORMULVKRI ISYLFCVLDNI RKL 64 IYYVI IPSLALUKÇI SI PPEITSPWI APFVYVFCVKNMY SIYBAUSS OTIKGWIN CORMULVKRI ISYLFCVLDNI RKL 64 IYYVI IPSLALUKÇI SI PPEITSPWI APFVYVFCVKNMY SIYBAUSS OTIKGWIN CORMULVKRI ISYLFCVLDNI RKL 64 IYYVI IPSLALUKÇI SI PPEITSPWI APFVYVFCVKNMY SIYBAUSS OTIKGWIN CORMULVKRI ISYLFCVLDNI RKL 64 IYYVI IPSLALUKÇI SI PPEITSPWI APFVYVFCVKNMY SIYBAUSCOTIKGWIN CORMULVKRI ISYLFCVLDNI RKL 64 IYYVI IPSLALUKÇI SI PPEITSPWI SI PIYVCVKNMY SIYBAUSCOTIKGWIN CORMUNVRI ISYLYGI TDI VRKL 64 IYYVI IPSLALUKÇI SI PPEITSPWI SIPI IYVVCVKNMY SIYBAUSCOTIKGWIN CORMUNVRI ISYLYGI TDI VRKL 64 IYYVI IPSLALUKÇI SI PPEITSPWI SIPI IYVVCVKNMY SIYBAUSCOTIKGWIN CORMUNVRI ISYLYGI TDI VRKL 64 IYYVI IPSLALUKÇI SI PPEITSPWI SIPI IYVVCVKNI SIYBAUSCOTIKGWIN CORMUNVRI ISYLYGI TDI VRKL 64 IYYVI IPSLALUKÇI SI PPEITSPWI SIPI IYVVCVKNI SIYBAUSCOTIKGWIN CORMUNVRI ISYLYGI TDI VRKL 64 IYYVI IPSLALUKÇI SI PPEITSPWI SIPI IYVVCVKNI SIYBAUSCOCITAVEWIN AQRMMURTI ISYLYGI ADIT GGM 61 IYYVI IPSLYTISPU SI SI PIYVCVVASISI SIMESI OCOTIAVEWINA QRMMURTI SYLLAA IDI GGM 61 IYYVI IPSLCFI SOVSV FPEITSPWCI PI IYVVASISI SIMESI OCOTIAVEWINA QRMMURTI SYLLAA IDI GGM 61 IYYVI IPSLCFI SOVSV FPEITSPWCI PI IYVVASISI SIMESI OCOTIAVEWINA QRMMURTI SYLLAA IDI GGM 61	35)7 24)6)4)7)7 17 16
TRIAE_CS42_5DL_TGACv TRIAE_CS42_5BL_TGACv TRIAE_CS42_U_TGACv1_ TRIAE_CS42_5AL_TGACv TRIAE_CS42_6DS_TGACv TRIAE_CS42_5DL_TGACv TRIAE_CS42_5BL_TGACv TRIAE_CS42_6AL_TGACv TRIAE_CS42_6BL_TGACv TRIAE_CS42_6DL_TGACv	LCLSKINEVVTPKVSDEDESKRYEQEIMEFGSSDPEYVIIAAIAIINIVCIMCGUSKVMKGGWN-VHDDALFPQLILCGM 71 LCLSKMNEVVIPKVSDEDESKRYEQEIMEFGSSDPEYVIIGTILINIVCILCGUSKVMKGGWN-EHDDALFPQLILCGM 68 LCLSKMNEVVSPKVSDEDESKRYEQEIMEFGSSDPEYVIIGTILINIVCILCGUSKVMKVGWNNIHDDALFPQLILCGM 68 LCLSKMTEVVTPKVSIEDESKRYEQEIMEFGSSDPEYVIIGTILINIVCILCGUSKVMKVGWNNIHDDALFPQLILCGM 68 LCLSKMTEVVTPKVSIEDESKRYEQEIMEFGSSDPEYVIIGTILINIVCILCGUSKVMKVGWNNIHDDALFPQLILCGM 68 LCLSKMTEVVTPKVSIEDESKRYEQEIMEFGSSDPEYVIIGTILINIVCILCGUSKVMKVGWNNIHDDALFPQLILCGM 68 LCLSKMTEVVTPKVSIEDESKRYEQEIMEFGSSDPEYVIIGTILATIAICNIVCILCGUSKVMKVGWNNIHDDALFPQLILCGM 68 LCLSKMTEVVTPKVSIEDESKRYEQEIMEFGSSDPEYVIIGTILATIAICNIVCILCGUSKVMKVGWNNIHDDALFPQLILCGM 68 LCLSKMTEVVTPKVSIEDESKRYEQEIMEFGSSDPEYVIIGTIAIAINNCIVVGLQIMTGGWN-ILLNVFSPQLILCGM 68 LCLSKMTEVVTPKVSUESESKRYEQEIMEFGSSDPEYVIIATVAIINNCIVVGLQIMTGGWN-ILLNVFSPQLILCGM 68 LGVSESGELTVKVDESQALERYKKGKMEFGPISGMEVIITTIAIFNIVCIVGUGRVVLREGA-AGIGPLFLQAVICVA 69 LGVSESGELTVKVDESQALERYKKGKMEFGPISGMEVIITTIAIFNIVCIVGLGRVVLREGA-AGIGPLFLQAVICVA 69	4)4)36)36)36)36)36)36)36)36)37
TRIAE_CS42_5DL_TGACv TRIAE_CS42_5BL_TGACv TRIAE_CS42_0_TGACv1_ TRIAE_CS42_5AL_TGACv TRIAE_CS42_6DS_TGACv TRIAE_CS42_5DL_TGACv TRIAE_CS42_5BL_TGACv TRIAE_CS42_6AL_TGACv TRIAE_CS42_6BL_TGACv TRIAE_CS42_6DL_TGACv	VITSIEFYEAMELRKDKCRIPAPVTLASIGFVMLALLAKIV 756 VVTSIEFYEAMELRKDKCRIPAPVTLASIGFVMLALLATIV 728 VVITSIEFYEAMELRKDKCRIPAPVTLASIGFVMLALLPAIV 446 VVTSIEFYEAMELRKDKCRIPAPVTLASIGFVMLALLPAIV 728 VVTNNEFYEAMELRNDKCKIPATVTLASIGFVMLALLPAIV 728 VVTNNEFYEAMEVRKDKCRIPASVTLASIGFVMLALLPPIV 728 VVTNIEFYEAMEVRKDKCRMERSVTLASIGFVMLALLPPIV 728 VVTNIEFYEAMEVRKDKCRMERSVTLASIGFVMLALLPPIV 728 VVTNIEFYEAMEVRKDKCRMERSVTLASIGFVMLALLPPIV 728 VVTNIEFYEAMEVRKDKCRMERSVTLASIGFVMLALLPPIV 728 VVTNIEFYEAMEVRKDKCRMERSVTLASIGFVMLALLPPIV 736 VVINAEVYEALETRRDSGIEVFVTLVSLCFVSSLCUQAE 738	

Appendix 6.5 List of *CslF* subfamily genes, their protein length (amino acids) and multiple sequence alignment of amino acids showing the conserved motifs (D, D, DXD, QXXRW) specific to *Cellulose synthase like* (*Csl*) family (highlighted with red boxes).

C No. Co			umber of	anliaa		(CalE)	No	of omino ocida	(22)	
S.NO GE	ene name	with n	umber of	splice	variants	(CSIF)	NO.	or amino acids	(aa)	
1	TRIAE_CS	42_2DL	_TGACv1_1	59781_A	A0542640.	1	845	aa		
2	TRIAE_CS	42_7BL	_TGACv1_5	80651_A	A1914920.	1	614	l aa		
3	TRIAE_CS	42_7AL	_TGACv1_5	57532_A	A1782680.	1	837	aa		
4	TRIAE_CS	42_7DL	_TGACv1_6	02590_A	A1961740.	1	835	ā aa		
5	TRIAE_CS	42_2AL	_TGACv1_0	94713_A	A0301960.	1	865	ā aa		
6	TRIAE_CS	42_2DL	_TGACv1_1	60109_A	A0546890.	1	862	2 aa		
7	TRIAE_CS	42_2BL	_TGACv1_1	30934_A	A0420130.	1	663	3 aa		
8	TRIAE_CS	42_2DS	_TGACv1_1	78985_A	A0603230.	1	870) aa		
9	TRIAE_CS	42_2AS	_TGACv1_1	12790_A	A0345230.	1	878	3 aa		
10	TRIAE_CS	42_2BS	_TGACv1_1	48027_A	A0489970.	1	877	7 aa		
11	TRIAE_CS	42_2AS	_TGACv1_1	13659_A	A0359050.	1	847	7 aa		
12	TRIAE_CS	42_2DS	_TGACv1_1	77641_A	A0581710.	2	847	/ aa		
13	TRIAE_CS	42_2BS	_TGACv1_1	48608_A	A0494060.	1	851	aa		
14	TRIAE_CS	42_U_T	GACv1_641	498_AA2	096480.1		857	7 aa		
15	TRIAE CS	42 2BS	TGACv1 1	46146 A	A0456710.	1	754	l aa		
16	TRIAE CS	42 2DS		79076 A	A0604160.	1	783	3 aa		
17	TRIAE CS	42 2AS	TGACv1 1	12322 A	A0335290.	1	878	3 aa		
18	TRIAE CS	42 ² 85		47667 A	A0486240.	1	877	7 aa		
19	TRIAE CS	42 2DS		77329 A	A0573830.	1	875	ā aa		
20	TRIAE CS	42 ² 85		48916 A	A0495580.	1	701	aa		
21	TRIAE CS4	42 2DS	TGACv1 1	78471 A	A0596060.	1	701	aa		
22	TRIAE CS	42 ² 2AS	TGACv1 1	12322 A	A0335280.	1	897	7 aa		
23	TRIAE CS	42 5BL	TGACv1 4	09916 A	A1366600.	2	815	aa		
24	TRIAE CS	42 5DL	TGACv1 4	33902 A	A1424880.	1	808	aa		
25	TRIAE CS	42 5AL	 TGACv1_3	74191 A	A1193100.	1	807	7 aa		
2.6	TRIAE CS	42.7BL	 TGACv1_5	77473 A	A1876170.	1	941	aa		
2.7	TRIAE CS	42. 7AL	 TGACv1_5	55973 A	A1751470.	1	899	aa		
28	TRIAE CS	42 7DT.	TGACV1 6	07937 A	A2011180	1	498	3 аа		
29	TRIAE_CS	42 1BS	 	49866 A	A0163180	1	856	5 aa		
Color Align TRIAE_CS42_ TRIAE_CS42_	Conservat 5DL_TGACv 5AL_TGACv	tion re 	sults 					MSMTYITKKH MSMTYITKKH	IDYAATLDEKEP IDYVASLDGKES	21 21
TRIAE_CS42_	5BL_TGACv							MSMTYISKKH	IDYAATLDEKEQ	21
TRIAE_CS42_	2AS_TGACv		MTTSP.	ATHDGAAT	GLSEPLLPNRN	IGVHAGALVVTP\	/VANGH0	GGGDKLKGDLKAKDKYW	KDVDQPDDVAA	69
TRIAE_CS42_	2BS_TGACV		MTTSP.	ATAAGAAT(GLSEPLLSNGN	IGVHAGALVVTP\	/VANGHO	GG-DKLKGDLKAKDKYW	KDVDQPDDVAA	68
TRIAE_CS42_	2DS_TGACV		MTTSP. MASA	ATDAGAAT(ACACCANA(JLSEPLLSNRN 21 ADDI LAS	IGVHAGALVVTP\	/AANGHO	GGKAKDKIW	KDVDQPGDMAV	61 43
TRIAE_CS42_	2DS TGACV		MASA	AGAGGANA	GLADPLLAS			AKKPVGAKGKHW	WAADK-DORRA	43
TRIAE CS42	2BS TGACV		MASA	VGAGGANA	GLADPLLASRE)		GGAKKPVGAKGKHW	IVAADK-DORRA	47
TRIAE CS42	2DL TGACV				MAAA	VTRRSNALRVD	PGGEA	VAVSVAADSPVAKRGLG	AKDDVWVAADE	49
TRIAE_CS42	2BL_TGACv									0
TRIAE_CS42_	2AL_TGACv				MMAAA	VTRRSNALRVD	/PDGDAV	VAVSVVADSPVAKRGLG	AKEDVWVAVDE	50
TRIAE_CS42_	2DL_TGACv				MAAA	VTRRVNALRVE	/PDG	NADTANAPAAKRII	DAKDDVWVSAD	44
TRIAE_CS42_	2BS_TGACV				MAAAVTRRA	NALRVEAPDGN	resgras	SLAADSPVAKRAVDAKD	DVWVAADEGEA	54
TRIAE_CS42_	2DS_TGACV				MAAAV'I'RRA	NALRAEAPDGNA	AESGRAS	SLAADSPAAKRAVDAKL	DVWVAADEGDT	54
TRIAE_CS42_	7AL_TGACV					MPI	RVEAL	VATUTASAAAAEGRRAA WATUTASAAAAEGRRAA	DUWWALLGDM	30
TRIAE_CS42_	7BL TGACV							MATDTVADAAEGRRARD	DVWVAAEEGDM	2.8
TRIAE CS42	U TGACv1				MPSPA	AVGGGRLADPLI	LAAD	VVVGAKDKYWVE	ADEREILASOK	43
TRIAE CS42	1BS TGACV				MASPA	AGGGGRLADPLI	LATD	VVVGPKDKYWVF	ADEREILASHR	43
TRIAE_CS42_	2AS_TGACv				MVSPATGGG	RGGNAGLAEPLI	LATNDDS	SDGAKHVFGAKAKHWVF	ADEKEMAASRE	54
TRIAE_CS42	2BS_TGACv				MVSPATSGO	RGGNAGLADPLI	LATNDDS	SDGARHVFGAKAKYWAF	ADEKEMTASRE	54
TRIAE_CS42_	2DS_TGACv				MVSPAAS	GGGNAGLADPLI	LATNDNS	SEGARHVFGAKAKYWVF	ADEKEIAASRE	52
TRIAE_CS42_	2BS_TGACV									U
TRIAE_CS42_	2DS_TGACV		CT & & & X C & C''	A CNICA CUAN						U 76
TRIAL_CS42_	ZAS_TGACV	MAPANA	SLAAANGAGH. GGGRUPSNFP	ASNGAGVAI	DŐATATENGI.(WCACTCSA AVA	LEELVAL	NGGSKVAKKISPKDKIN DMDIVAMGOICAVNDES	WUGUELCEDCE	/ 0 80
TRIAE CS42	7DL TGACV	A						CTGA 10DD3		0
TRIAE CS42	7BL TGACV	MAPAVA	GGGRVRSNEP	AAAA	ASDKPCVCGFC	VCACTGSAAVAS	SAASSLI	DMDIVAMGQIGAVNDES	WVGVELGEDGE	76
					-					
TRIAE CS42	5DL TGACV	SEDQKS	ASVKNLLVRT	TKLTTVTI	KLYRLMVFVRI	TIFVLFFKWRVS	STALTV:	ISDGTTTARAMWTMSIA	GELWFALMWVL	101

TRIAE_CS42_56L_TGACV SEBQROBSVRREDVRINGITVIRGINGAVERGENEVERMITVIGSTALIVISOSTITARAMINGIASEGUERALMAVE 101 TRIAE_CS42_5AL_TGACV PEHEKSASVERLLVRTTKLTTVIKLYRLVVFVRMIIFVLFFKWRSSTALAMISDGTTTVRAMWTMSIAGELWFALMWVL 101

TRIAE_CS42_5BL_TGACv	${\tt PKDQKSASVESLLVRTTKLTTVTIKLYRIMVFVRMAIFVLFFKWRISTALAMISDGATTVRAMWTMPIAGELWFALMWVL}$	101
TRIAE_CS42_2AS_TGACv	APDLENGGGRPLLFSNRRVKNIILYPYRVLILIRVIAVILFVGWRIKHNNSDVMWFWVMSVVADVWFSLSWLS	142
TRIAE_CS42_2BS_TGACv	APDLENGGGRPLLFSNRRVKNIILYPYRVLILIRVIAVILFVGWRIKHNNSDVMWFWVMSVVADVWFSLSWLS	141
TRIAE CS42 2DS TGACV	APDLENGGGRPLLFSNRRVKNIILYPYRVLILIRVIAVILFVGWRIKNNNSDVMWFWVISVVADVWFSLSWLS	134
TRIAE CS42 2AS TGACV	AKESGGEDGRPLLFRTYKVKGTLLHPYRALIFIRLIAVLLFFVWRIKHNKSDVMWFWTMSVVGDVWFGFSWLL	116
TRIAE CS42 2DS TGACV	AKESGGEDGRPLLFRTYKVKGTLLHPYRALIFIRLIAVLLFFVWRIKHNKSDIMWFWTLSVVGDVWFGFSWLL	116
TRIAE CS42 2BS TGACV	AKESGGEEGRPLLFRTYKVKGTLLHPYRALIFIRLIAVLLFFVWRIKHNKSDIMWFWTMSVVGDVWFGFSWLL	120
TRIAE CS42 2DL TGACV	GGIMSGDGNRPLLFRTMKVKGSILHPYRFLMLVRLVAVVAFFKWHVEHKNODSVWLWTASMTADPWFGFSWLL	122
TRIAE CS42 2BL TGACV		0
TRIAE CS42 2AL TGACV	GG-MSGDGNRPLLFRTMKVKGSILHPYRFLMLMRLVAVVIFFKWRMEHKNHDGVWLWTVSMTADVWFGFSWLL	122
TRIAE CS42 2DL TGACV	DGTSAGNGNOPLI, FRTMKVKGSTI, HPYRFL, ILVRLVAVAAFFAWRI, EHRNHDGTWL, WATSMVADAWFGFSWLI,	117
TRIAE CS42 2BS TGACV	SGSIAGDGNRTPLFRTFKVKGSILHPYRFMILVRLVAIVAFFAWRVKHKNHDGVWLWATSMVADVWFGFSWLL	127
TRIAE CS42 2DS TGACV	SGA LAGDGNR PPLERTFKVKGSTLHPYRFMILURIUA IVA FAWRVKHKNHDGVWLWATSWVA DVWFGFSWLL	127
TRIAE CS42 7AL TGACV	SGASAGRPLIFRTMKVKGSTLHPYRFLILVRLVATVAFFAWRVEHRNHDGTWLWATSMVADAWFGFSWLL	106
TRIAE CS42 7DL TGACV	SGASAGPRILERTMKVKGSTLHPYRFLTLVRLVATIAFFAWRVEHRNHDGMWLWATSMVADAWFGFSWL	104
TRIAE CS42 7BL TGACT		98
TRIAE CS42 II TGACV1	SGAG-EDGRAPLI.VETERVKGPLI.NI.VELTI.VEVTVVII.FFTWRMRHRDSDAWIJWWISVVGDI.WEGVTWII.	115
TRIAF CS42 1BS TGACT		116
TRIAE CS42 2AS TGACT		124
TRIAE_CS42_2AS_IGACV		124
TRIAE_C342_2B5_IGACV	CSGEDGRFLLIKTFRVRGFLVNITKFLNLAKLIAVIVFFAWRVQRPDSDAMWLWWISVUGDFWFGLSWWL	124
TRIAE_CS42_2DS_TGACV	CGGEDGRPLLIRTFKVKGMLVNTIRFLNLARLIAVIVFFAWRVQHPDSDAMWLWWISVVGDFWFGLSWWL	122
TRIAE_CS42_2BS_TGACV		0
TRIAE_CS42_2DS_TGACV		140
TRIAE_CS42_ZAS_TGACV	IADGGEDGRRPLLYRFFKVKGILLHPYRLSLIRLVAIVLFFVWRVRHPYADGMWLWWISMVGDLWFGVIWLL	149
TRIAE_CS42_/AL_TGACV	TDESGVAVDDRPVFRTEKIKGVLLHPYRVLIFVRLIAFTLFVIWRISHKNPDAMWLWVTSICGEFWFGFSWLL	153
TRIAE_CS42_/DL_TGACV		0
TRIAE_CS42_7BL_TGACv	TDESGAAVDDRPVFRTEKIKGVLLHPYRVLIFVRLIAFTLFVIWRISHKNPDAMWLWVTSICGEFWFGFSWLL	149
TRIAE_CS42_5DL_TGACv	DQLPKMQPVRRTVYVTALEEPRLPTMDVFVTTTDPEKEPPLVTVNTILSILAADYPPDKLTCYVSDDGGALL	173
TRIAE_CS42_5AL_TGACv	DQLPKMQPVRRTVYATALEESLLPAMDVFVTTADPEKEPPLVTVNTILSILAADYPPDKLTCYVSDDGGALL	173
TRIAE_CS42_5BL_TGACv	DQLPKMQPVRRTVFATALEEPLLPTMDVFVTTADPEKEPPLVTVNTILSILAADYPPDKLTCYVSDDGGALL	173
TRIAE_CS42_2AS_TGACv	YQLPKYNPIKMIPDLATLRKQFDTPGRSSQLPGIDVIVTTASATDEPILYTMNGVLSILAADYHIGRCNCYLSDDSGSLV	222
TRIAE_CS42_2BS_TGACv	YQLPKYNPIKMIPDLATLRKQFDTPGSSSQLPGIDVIVTTASATDEPILYTMNCVLSILAADYHIGRCNCYLSDDSGSLV	221
TRIAE_CS42_2DS_TGACv	$\verb"YQLPKYNPIKMIPDLATLRKQFDTPGRSSQLPGIDVIVTTASATDEPILYTMNCVLSILAADYHIGRCNCYLSDDSGSLV"$	214
TRIAE_CS42_2AS_TGACv	NQLPKFNPVKTIPDMVALRRQYDLPDGTSTLPGIDVFVTTADPIDEPILYTMNCVLSILASDYPVDRCACYLSDDSGALI	196
TRIAE_CS42_2DS_TGACv	NQLPKFNPVKTIPDMVALKRQYDLPDGTSTLPGIDVFVTTADPIDEPILYTMNCVLSILASDYPVDRCACYLSDDSGALI	196
TRIAE_CS42_2BS_TGACv	NQLPKFNPVKTIPDMVALRRQYDLPDGTSTLPGIDVFVTTADPIDEPILYTMNCVLSILASDYPVDRCACYLSDDSGALI	200
TRIAE CS42 2DL TGACV	NQLPKLNPIKRVPDLADRHDDATLPRIDVFVTTVDPVDEPVLYTVNTILSILAADYPIDNYACYISDDGGTLV	195
TRIAE CS42 2BL TGACV		0
TRIAE CS42 2AL TGACV	NQLPKLNPIKRVPDLAALADRHDDATLPGIDVFVTTVDPVDEPVLYTVNTILSILAADYPVDNYACYLSDDGGTLV	198
TRIAE CS42 2DL TGACV	NQLTKLNPIKRVPDLATLADQHGEAILPGIDVFVTTADPVDEPVLYTVNTVLSILAADYPIDKYACYLSDDGGTLV	193
TRIAE CS42 2BS TGACV	NQLPKLNPVKRVPDLAALADHSGDANLPGIDIFVTTVDPVDEPLLYTVNTILSILATDYPVDKYACYLSDDGGTLV	203
TRIAE CS42 2DS TGACV	NQLPKLNPVKRVPDLAALADHSGDANLPGIDIFVTTVDPVDEPLLYTVNTILSILATDYPVDKYACYLSDDGGTLV	203
TRIAE CS42 7AL TGACV	NQLPKLNPIKRVPDLVALADRHGEAILPGIDVFVTTVDPVDEPVLYTVNTILSILAADYPVDKYACYLSDDGGTLV	182
TRIAE CS42 7DL TGACV	NOLPKLNPIKRVPDLAALADLHGEAVLPGIDVFVTTVDPVDEPVMYTVNTILSILAADYPVDKYACYLSDDGGSLV	180
TRIAE CS42 7BL TGACV	NOLPKLNPIKRVPDLAALADRHGEAVLPGIDVFVTTVDPVDEPVMYTVNTILSILAADYPVDKYACYLSDDGGTLV	174
TRIAE CS42 U TGACV1	NOTTKLEPEKCVPSISVLEDHLDOPDGGSDLPLLDVFINTVDPVDEPMLYTMNSILSILATDYPVEKYATYFSDDGGSLV	195
TRIAE CS42 1BS TGACV	NOTTKLEPEKCVPSISVLEEOLOPDGGSDLPLLDVFINTVDPVDEPMLYTMNSILSILATDYPVDKYATYFSDDGGSLV	196
TRIAE CS42 2AS TGACV	NOVPKLNPTICIPTIPLLROOFDLPDGGSNLPVLDVFISTVDPVEEPMLHTMNSILSILATDYPVDKYATYLSDDGGSLL	204
TRIAE CS42 2BS TGACV	NOVPKLNPTICIPTIPLLROOFDLPDGGSNLPVLDVFISTVDPVEEPMLHTMNSILSILATDYPVDKYATYLSDDGGSLL	204
TRIAE CS42 2DS TGACV	NOVPKINPTICIPTIPILROOFDLPDGGSNLPVLDVFISTVDPVEEPMLHTMNSILSILATDYPVDKYATYLSDDGGSLL	202
TRIAE CS42 2BS TGACV		33
TRIAE CS42 2DS TGACV		33
TRIAE CS42 2AS TGACV	NOVAKI.NEVKRVENI.TI.LEOOFDI.PDGNSNI.PCI.DVFINTVDFINEPMI YTMNSI ISILADY PVDKHACYI.SDDGGSI I	229
TRIAE CS42 7AL TGACV	DOL PKINPTNEVPDI. AVI. FOR FOR POCTSTI. PCI. DT FVTTA DP TKE PTI. STANSVI. STI. A DV PVDRVTCVVSDDSCMI.	233
TRIAE CS42 7DL TGACV		0
TRIAE CS42 7BL TGACT	ΠΟΙ ΣΚΙ ΝΕΤΝΕΙΣΕΙ Α ΤΙ ΣΟΥ ΕΠΟ ΕΠΟ ΕΠΟ ΕΠΟ ΕΠΟ ΤΟ ΤΙ ΤΙ ΤΟ ΤΗ ΕΝΤΑΤΟ ΤΗ ΕΝΤΑΙΝΟΥ ΤΑ ΤΙ Α ΑΠΥ ΕΥΠΑΝΤΟΥ ΥΠΟΠΟΥ ΤΙ	229
INITE_CO42_/DD_IGACV		229
TRAF COAS 5DT TOACT		253
TRIAL_CS42_JDL_IGACV	TREAVAILARCE ARLIWEFECKNIGVEF KNEELE FOR GVRARV VSRAD HIGRSWEELEANDRRVKKEELEELEND DALLAD	253
TRIAL_CS42_JAL_IGACV	I NERVAYANNE AND WEET CHUNERDEDE E VEODVICE VORADINANSWE ELANDRIKVIKE I EELELI DADINAD	253
TRIAL_CS42_JDL_IGACV	I VEAVAINARY AND WY FORMUS VERNY EATFORST VOIDANT MIGROWE LANDARY AND TELEDAD ADDITIONAL	205
TRIAE_CS42_2AS_IGACV	LIEALVEIARFAALWVFFCRKNQIEFRAFESIFELEGILCGGASINEFIQDIRNVKIQIDEFRANDOMENII	295
TRIAE_CS42_2BS_IGACV	I LEALVEIANFAALWYF CKNRUIEFNAFARIFELEGFEGGASINEFIQUINNYMUIEFNIL I YRAI YWRYFRAI WYF CKNRUIEFNAFARIFELEGF LCGASINEFIQUINNYMUTEFNIL	294
TRIAE_CS42_2DS_IGACV	DIEALVEIARFAALWVFFCRKNQIDFRAFESIFELEGFLOGGASINEFIQDIRAVCIQIEEFRALDMENII	201
TRIAE_CS42_ZAS_TGACV	QUEALVETAFFATLWVPFCRKHCIEPRAPESFFEQEAPLiTGSAPEFFKNDHNSVIIEIDEFRECLDSLSSAI	269
TRIAE_CS42_2DS_TGACV	QUEALVETAFFATLWVPFCRKHCIEPRAPESYFELEAPLYTGSAPEDFKNDHSSVHREYDEFKEHLDSISSAI	269
TRIAE_CS42_2BS_TGACV	QTEALLETAFFATLWVPFCRKHCIEPRAPESYFELEAPLYTGSASEEFKNDHSSVHREYDEFKEHLDSLSSAI	2/3
TRIAE_CS42_2DL_TGACV	HYEAMVQVASFAALWVPFCRKHCVEPRSPESYFGIKTRSYIGGMAGEFMRDHRRVRREYEEFKVRIDSLSTTI	268
TRIAE_CS42_2BL_TGACV	SYAGGMAGEFMRDHRRVREYEEFKVRIDSLSTTI	69
TRIAE_CS42_2AL_TGACV	HYEAMVQVASFAALWVPFCRKHCVEPRSPESYFGIKTHSYAGGMAGEFMRDRRRVRREYEEFKVRIDSLSTTI	271
TRIAE_CS42_2DL_TGACv	HYEAMTQVASFAALWAPFCRKHCVEPRSPENYFGMKAQPYAGSMPGDFTRDRRVREYDEFMVRIDSLSTTI	266
TRIAE_CS42_2BS_TGACv	HYEAMIEVANFAVLWVPFCRKYCVEPRSPENYFGMKTQPYAGSMAGEFMRDHRRVRREYDELKVRVDSLSTTI	276
TRIAE_CS42_2DS_TGACv	HYEAMIEVANFAVLWVPFCRKYCVEPRSPENYFGMKTQPYAGSMAGEFMRDHRRVRREYDEFKVRVDSLSTTI	276
TRIAE_CS42_7AL_TGACV	HYEAMLQVASFAALWVPFCRKHCVEPRSPENYFGMKTRPYVGGMAGEFMSDBRRVRREYGEFKVRIDSLSTTI	255
TRIAE_CS42_7DL_TGACv	HYEAMIQIVHFAALWVPFCRKHCIEPRSPENYFGMKTRPYVGGMAGEFMSDHRRVREYGEFKVIIDSLSTTI	253
TRIAE_CS42_7BL_TGACv	HYEAMLQVASFAALWVPFCRKHCVEPRSPENYFGMKTRPYVGGMAGEFMSDHRRVRREYGEFKVRIDSLSSTI	247
TRIAE_CS42_U_TGACv1_	HYEGLQLAAEFAASWVPFCRKHCVEPRAPESYFWAKMRGEYAGSAPKEFLDDHRRMRAAYEEFKARLDGLSAAI	269
TRIAE_CS42_1BS_TGACv	HYEGLQLAAEFAASWVPFCRKHCVEPRAPESYFWAKMRGEYAGTAPKEFLDDHRRMRAAYEEFKVRLDGLSAAI	270
TRIAE_CS42_2AS_TGACv	HYDGLVETAKFAALWVPFCRKHHVEPRAPESYFGMKVRPYKGNLPEEFLDDHRRLRREYEEFKTRLDALFTVI	277
TRIAE_CS42_2BS_TGACv	HYDGLVETAKFAALWVPFCRKHHVEPRAPESYFGMKIRPYTGNLPEEFLDDHRRLRREYEEFKTRLDALFTVI	277

TRIAE_CS42_/BL_TGACV	KQRNDGYNAANAH-REGEPRPTWMADG-TQWEGTWVDASENHRRGD	HAG	IVRVLLNHPSHRRQTGPPASAD-NPLDFSGV
TRIAE CS42 5DL TGACV	DVRVPAVVYMCREKRHGRVHHRKAGAMNALLRTSAVLSNAPFTLNT	DCD	YVNNSOAL RAGVCLMLD-RGGSNVAFVOFP
TRIAE CS42 SAL TGACV	DVRVPAVVYMCREKRHGRVHHRKAGAMNALLRTSAVLSNAPFTLNI		YVNNSOALBAGVCLMLD-BGGSNVAFVOFP
TRIAE CS42 5BL TGACV	DVRVPAVVYMCREKRHGRVHHRKAGAMNALLRTSAVLSNAPFILNI		YVSNSOALRAGVCLMLD-RGGSNVAFVOFP
TRIAE CS42 2AS TGACV	DMRLPMLVYVAREKSPGVEHNKKAGALNAELRISALLSNAPFFINF	DCD	YINNSEALRAAICFMLDPREGDNTGFVOFP
TRIAE CS42 2BS TGACV	DMRLPMLVYVAREKSPGVEHNKKAGALNAELRISALLSNAPFFINF	DCD	HYINNSEALRAAVCFMLDPREGDNTGFVOFP
TRIAE CS42 2DS TGACV	DMRLPMLVYVAREKCPGVEHNKKAGALNAELRISALLSNAPFFINF	DCD	HYINNSEALHAAVCFMLDPREGDNTGFVOFP
TRIAE CS42 2AS TGACV	DVRLPMLVYISRGKNPSYDHNKKAGALNAOLRASALLSNAOFIINF	DCD	HYINNSOALRAAMCFMLDOROGDSTAFVOFP
TRIAE CS42 2DS TGACV	DVRLPMLVYISRGKNPSYDHNKKAGALNAOLRASALLSNAOFIINF	DCD	HYINNSOALRAAMCFMLDOROGDSTAFVOFP
TRIAE CS42 2BS TGACV	DVRLPMLVYISRGKNPSYDHNKKAGALNAQLRASALLSNAQFIINF	DCD	HYINNSQALRAAMCFMLDQRQGDSTAFVQFP
TRIAE CS42 2DL TGACV	DTRLPMLVYISREKRPGYDNQKKAGAMNVMLRVSVLLSNAPFVINF	DCD	HYINNSQALRAPMCFMLDPHDGQNTAFVQFP
TRIAE CS42 2BL TGACV	DTRLPMLVYISREKRPGYDNQKKAGAMNVMLRVSALLSNAPFVINF	DCD	HYINNSQALRAPMCFMLDPHDGQNTAFVQFP
TRIAE CS42 2AL TGACV	DTRLPMLVYISREKRPGYDNQKKAGAMNVMLRVSALLSNAPFVINF	DCD	HYINNSQALRAPMCFMLDPRDGQNTAFVQFP
TRIAE CS42 2DL TGACV	DMRLPMLVYISREKRLGYDNQKKAGAMNAMLRISALLSNAPFIINF	DCD	HYINNSKALRAPMCFMLDPRDGQNTAFVQFP
TRIAE CS42 2BS TGACV	DTRLPMLVYMSREKRPGYNHQKKAGAMNVMLRVSALLSNAPFVVNF	DGD	HYINNSQALCAPMCFMLDPRDGQNTAFVQFP
TRIAE CS42 2DS TGACV	DTRLPMLVYMSREKRPGYNHQKKAGAMNVMLRVSAMLSNAPFVVNF	DGD	HYINNSQALRAPMCFMLDPRDGQNTAFVQFP
TRIAE CS42 7AL TGACV	DTRLPMLVYISREKHPGYDNQKKAGAMNVMLRVSALLSNAPFVINF	DCD	HYINNSRALRAPMCFMLDPRDGQNTAFVQFP
TRIAE CS42 7DL TGACV	DTRLPMLVYISREKHPGYDNQKKAGAMNVMLRVSALLSNAPFVINF	DCD	HYINNSQALRAPMCFMLDPRDGQNTAFVQFP
TRIAE_CS42_7BL_TGACv	DTRLPMLVYISREKRPGYDNQKKAGAMNVMLRVSALLSNAPFVINF	DCD	HYINNSQALRAPMCFMLDPRDGQNTAFVQFP
TRIAE_CS42_U_TGACv1_	DARLPMLVYIAREKRPGYDHQKKAGAMNVQLRVSALLSNAPFIINF	DGD	HYVNNSQAFRAAICFMLDPRDGADTAFVQFP
TRIAE_CS42_1BS_TGACv	DARLPMLVYIAREKRPGYDHQKKAGAMNVQLRVSALLSNAPFIINF	DGD	HYVNNSQAFRAAMCFMLDPRDGADTAFVQFP
TRIAE_CS42_2AS_TGACv	DVRLPMLVYVSREKRPGYDHQKKAGALNVQLRVSALLSNAPFIINF	DCD	HYINNSQAFRAAMCFMMDRRDGDNVAFVQFP
TRIAE_CS42_2BS_TGACv	DVRLPMLVYVSREKRPGYDHQKKAGALNVQLRVSALLSNAPFIINF	DCD	HYINNSQAFRAAMCFMMDRRDGDNVAFVQFP
TRIAE_CS42_2DS_TGACv	DVRLPMLVYVSREKRPGYDHQKKAGALNVQLRVSALLSNAPFIINF	DCD	HYINNSQAFRAAMCFMMDRRDGDNVAFVQFP
TRIAE_CS42_2BS_TGACv	DVRLPMLVYISREKSPSCDHQKKAGAMNVQLRVSALLTNAPFIINF	DGD	HYVNNSKAFRAGICFMLDRREGDNTAFVQFP
TRIAE_CS42_2DS_TGACv	DVRLPMLVYISREKSPSCDHQKKAGAMNVQLRVSALLTNAPFIINF	DGD	HYVNNSKAFRAGICFMLDRREGDNTAFVQFP
TRIAE_CS42_2AS_TGACv	DVRLPMLVYISREKSPSCDHQKKAGAMNVQLRVSALLTNAPFIINF	DGD	HYVNNSKAFRAGICFMLDRREGDNTAFVQFP
TRIAE_CS42_7AL_TGACv	DVRLPMLVYVXXXXX		XXXXXX-XXXXXXXXXXXXXXXXXXXXX
TRIAE_CS42_7DL_TGACv			MVG-RDSDTVAFVQFP
TRIAE_CS42_7BL_TGACv	DARLPMLVYVSREKRPGHDHQKKAGAMNALTRASALLSNSPFILNI	DCD	HYINNSQALRAGICFMVG-RDSDTVAFVQFP
TRIAE_CS42_5DL_TGACv	QRFDGVDPADRYANHNRVFHDCTELGLDGLQGPIYVGTGCMFRRAA	LYN.	ADPPLWRPHGGDRDAGK
TRIAE_CS42_5AL_TGACv	QRFDGVDPADRYANHNRVFHDCTELGLDGLQGPIYVGTGCMFRRAA	LYN.	ADPPLWRPHG-DRDAGK
TRIAE_CS42_5BL_TGACv	QRFDGVDPADRYANHNRVFHDCTELGLDGLQGPIYVGTGCMFRRAA	LYN.	ADPPLWRPHGGDRDAGK
TRIAE_CS42_2AS_TGACv	QRFDNVDPTDR <mark>Y</mark> G <mark>NHNR</mark> VFHDGAMYGLNGQQGPTYLGTGCMFRRLA	LΥG	IDPPCWRDEDIIVDSN
TRIAE_CS42_2BS_TGACv	QRFDNVDPTDRYGNHNRVFIDGAMYGINGQQGPTYLGTGCMFRRLA	LΥG	IDPPCWRAEDMIVDSN
TRIAE_CS42_2DS_TGACv	QRFDNVDPTDRYG <mark>NHNR</mark> VFHDGAMYGINGQQGPTYLGTGCMFRRLA	LYG	IDPPCWRAEDIIVDSN
TRIAE_CS42_2AS_TGACv	QRFDNVDPSDRYG <mark>NHNR</mark> VFHDGTMLALNGLQGPSYLGTGCMFRRIA	LYG	IDEPEWRHANIVVDDK
TRIAE_CS42_2DS_TGACv	QRFDNVDPSDRYG <mark>NHNR</mark> VFHDGTMLALNGLQGPSYLGTGCMFRRIA	LYG	IDEPEWRHDNIVVDDK
TRIAE_CS42_2BS_TGACv	QRFDNVDPSDRWGNHINRVDFDGTMLALNGLQCFSYLGTGCMFRRIA	ĽУG	IDEPEWRHDNIVVDDK
TRIAE_CS42_2DL_TGACv	QRFDDVDPTDRMANHINRVDFDGTMLALNGLQCFTYLGTGTMFRRVS	LYG	LPRYKAENTKLVRK
TRIAE_CS42_2BL_TGACv	QRFDDVDPTDRYANHNRVELDGTMLALNGLQGPTYLGTGTMFRRVA	ЦYG	IPPHYRAENTKLVCK
TRIAE_CS42_2AL_TGACv	QRFDDVDFTDR MANHINRVDFDG TMLALNGLQGFTYLGTGTMFRRVS	ЦYG	IPPPRYKAENTKLVRK

TRIAE CS42 2DS TGACV HYDGLVETAKFAALWVPFCRKHHVEPRAPESYFGVKIR-----PYMGNLPEEFLDDHGRLRREYEEFKTRLDALFTLI 275 TRIAE_CS42_2B5_TGACv HYDGLLETAKFAALWVPFCRKHSIEPRAPESYFSLNTR-----PYTGNAPQDFVNDRRHMCREYDEFKERLDALFTLI 106 TRIAE CS42 2DS TGACV HYDGLLETAKFAALWVPFCRKHSIEPRAPESYFSLNTR-----PYTGNAPQDFVNDRRHMCREYDEFKERLDALFTLI 106 TRIAE_CS42_2AS_TGACv HYDGLLETAKFAALWVPFCRKHSIEPRAPESYFSLNTR-----PYTGNAPQOFVNDRRHMCREYDEFKERLDALFTLI 302 TRIAE_CS42_7AL_TGACv TYEALAESSKFATLWVPFCRKHGIEPRGPESYFELKSHP-----YMGRAQDEFVNDRRRVRKEYDEFKARINSLEHDI 306 TRIAE CS42 7DL TGACV ------ 0 TRIAE CS42 7BL TGACV TYEALAESSKFATLWVPFCRKHGIEPRGPESYFELKSHP-----YMGRAQDEFVNDRRRVRKEYDEFKARINSLEHDI 302 TRIAE_CS42_5DL_TGACv_VRPQ-----VSDLLDLSSV_295 TRIAE_CS42_5AL_TGACv_VRRQ-----VSDLLDLSSV_295 TRIAE CS42 5BL TGACV VRRQ-----VSDLLDLTSV 295 TRIAE CS42 2AS TGACV RQRSDIYSRTGTK--DEDATVTWMADG-TQWPGTWLDPTEKHRPGHHAGIVKIVQSHPEHVVPLG-VQESNDNPLNFDDV 371 TRIAE CS42 2BS TGACV RQRSDIYSKTGTK--DEDAKVTWMADG-TQWPGTWVDPAEKHRAGHHAGIVKIVQSHPEHVVPLG-VQESNDNPLNFDDV 370 TRIAE CS42 2DS TGACV RQRADIYSKTGTK--DEDAKVTWMADG-TQWPGTWLDPAEKHRAGHHAGIVKIVQSHPEHVVPLG-VHESNDSSLNFDGV 363 TRIAE CS42 2AS TGACV SKRSDAYNSMKTE--EGDANATWMANG-TQWPGSWIDTTEIHRKGHHAGIVKVVLDHSIRGHNLG-SQASTHN-LNFAST 344 TRIAE CS42 2DS TGACv SKRSDAYNSMKTE--EGDAKATWMANG-TQWPGSWIDTTEIHRKGHHAGIVKVVLGHSIRGHNLG-SQASTNN-LNFAST 344 TRIAE CS42 2BS TGACV SKRSDAYNSMKTG--EGDAKATWMANG-TOWPGSWIDTTEIHRKGHHAGIVKVVLDHSVRGHNLG-SOASTHN-LNFANT 348 TRIAE_CS42_2DL_TGACv RQRS---DAYNS-SNKGGVSATWMADG-THWPGTWVEQAENHRRGQHAGIVQVLLDHPSCKPQLGSPASTD-NPFDFSNI 342 TRIAE_CS42_2BL_TGACv RQRS---DAYNS-NNKGGVSATWMADG-TQWPGTWVEQAENHRRGQHAGIVQVLLDHPSFKPQLGSPASTD-NPFDFSNV 143 TRIAE CS42 2AL TGACV RQRS---DAYNS-KNKGGVSATWMADG-TQWPGTWVEQAENHRRGQHAGIVQVLLDHPSCEPQLGSPASTD-NPFDFSNV 345 TRIAE_CS42_2DL_TGACv RQRS---DAYN---NGDGVHATRMADG-APWPGTWIEQAENHRRAQHAGIVQVILEHPGCKPQLGSSASTD-NPFDFNNV 338 TRIAE_CS42_2BS_TGACv_RQRS---DAYNSSTKGDGVRATWMADG-TQWPGTWIEQVENHRRGQHAGIVQVILGHPSCKPQLGSPASSD-NPLDFSNV_351 TRIAE_CS42_2DS_TGACv RQRS---DAYNSSKKGDGVRATWMADG-TQWPGTWIEQVENHRRGQHAGIVQVILGHPSCKPQLGSPASAD-NPLDFSNV 351 TRIAE_CS42_7AL_TGACv RRRS---DAYN--KGDDGVHATWMADG-TQWAGTWIEQADNHRRGHHAGIVQVMLDHPSCKPQLGSSVSTN-SPIDLSNV 328 TRIAE CS42 7DL TGACV RRRS---DAYN--KRDDGVHATWMADG-TQWAGTWIEQADNHRRGQHAGIVQVMLDHPSCKPQLGSSARTN-NPIDLSNV 326 TRIAE_CS42_7BL_TGACv RRRS---DAYN--KGDDDVHATWMADG-TQWPGTWIEQADNHRRGQHAGIVKVMLDHPSCKPQLGSSASTN-KPVDLSNV 320 TRIAE CS42 U TGACv1 EQRSEACNRANGKDKEECANATWMADGSTQWQGTWIKPAKGHHPAILQVMLDQPSKDPELGMAASSD-HPLDFSAV 348 TRIAE_CS42_1BS_TGACV EQRSEACNRANG--KEEGADATWMADGSTQWQGTWIKPAKGHRKGHHPAILQVMLDQPSKDPELGMAASSG-HPLDLSAV 347 TRIAE CS42 2AS TGACV PQRSEAHGREDAK-GGGGAKATWMADG-TQWPGTWTEPAEGHRKGDHAGIIQVMLSQPSGEPQLGAPASSDDNPLDFSAV 355 TRIAE_CS42_2BS_TGACv PQRSEAHGREDAK-GGG-GKATWMADG-TQWPGTWTEPAEGHRKGDHAGIIQVMLSQPSSEPQLGEPASSDDGPLDFSAV 354 TRIAE CS42 2DS TGACV PQRSEAHGREDAK-GGG-GKATWMADG-TQWPGTWTEPAEGHRKGDHAGIIQVMLSQPSSEPQLGEPASSDHSPLDFSAV 352 TRIAE CS42 2BS TGACV PKRSDVYNHAAAK---EGAKATWMADG-TQWPGTWIDPAENHKKGQHVGIVKVMLKHPSYEPELGLGASTN-SPLDFSAV 181 TRIAE_CS42_2DS_TGACv PKRSDVYNHAAAK---EGAKATWMADG-TQWPGTWIDPAENHKKGQHVGIVKVMLKHPSYEPELGLGASTN-SPLDFSAV 181 TRIAE CS42 2AS TGACV PKRSDVYNHAAAK---EGAKATWMADG-TQWPGTWIDPAENHKKGQHVGIVKVMLKHPSYEPELGLGASTN-SPLDFSAI 377 TRIAE_CS42_7AL_TGACv KQRNDGYNAANAH-REGEPRPTWMADG-TQWQGTWVDASENHRRGDHAGIVLVLLNHPSHRRQTGPPASAD-NPLDFSGV 383 TRIAE_CS42_7DL_TGACv 0 TRIAE_CS42_7BL_TGACv KQRNDGYNAANAH-REGEPRPTWMADG-TQWEGTWVDASENHRRGDHAGIVRVLLNHPSHRRQTGPPASAD-NPLDFSGV 379

374

374 374

451 450

443

424

424

428

422

223

425

418

431

431

408 406

400 428

427

435

434 432

261

261

457

416

440

439 440

516

515

508 489

489

493

487 288

490

15 458

TRIAE_CS42_2DL_TGACV	A D D D D L L D D D D D D D D D D D D D	Company of the second sec	of the state and and	The second se		400
	QRF DDVDPTDRYANHNRVFFD	FIMLALNGLQGP:	SYLCICTME	RRVTLYGMEPI	PRYRVENIKLVDN	483
TRIAE CS42 2BS TGACv	QRFDDVDPTDRYANHNRVFFD	GTMLSLNGLQGP	SYLGTGTMF	RRVALYGMEPI	PRYRAENIKLAGK	496
TRIAE CS42 2DS TGACV	ORFDDVDPTDRYANHNRVEFD	GTMLSINGLOGPS	SYLCTCTME	RRVAT YGMEP	PRYRAENTKLAGK	496
TRIAR COAST		TTMI ST NOT OCD			DDARYENTKI ACK	173
IKIAE_C342_/AL_IGACV	QREDRODETDRESKIINKVEED	JIMLSINGLOGI	I I I GI G I MI		KINAENINEVGR	475
TRIAE_CS42_/DL_TGACV	QRFDDVDPTDRYS <mark>NHNR</mark> VFFD	FIMLSLNGLQGP	TYLGTGTME	RRVALYGMEPI	CYRAENIKLVGK	4/1
TRIAE CS42 7BL TGACV	QRFDDVDPTDRYSNHNRVFFD	GTMLSLNGLQGP	TYL <mark>GTG</mark> TMF	HRVALYGMEPÇ	QRYRAENIKLVGK	465
TRIAE CS42 U TGACv1	ORFDDVDPTDRYCNHNRMEFD	ATLLGLNGIOGP	SEVGICCME	RRVALYSADPI	RWRPDDAKEAKAS	494
TRAF COAS 100 TCACT			S FWCTC CMF			103
TRIAL_CO42_IDD_IGACV						
TRIAE_CS42_2AS_TGACV	QRFDDVDPTDRYANHNRMFFD	ATMLGMNG IQG P	SYVGIGSME	RRVALYGADPE	PRWRPDDVKVLEN	500
TRIAE_CS42_2BS_TGACv	QRFDDVDPTDRYANHNRMFFD	ATMLGMNGIQGP	SYVGTGSMF	rrvalygadpi	PRWRPDDVKVLEN	499
TRIAE CS42 2DS TGACV	ORFDDVDPTDRYANHNRMEED	ATMLGMNGTOGP:	SYVGTGSME	RRVALYGADP	PRWRPDDVKVLEN	497
TRAF COAS ODS TO ACT			CVVCEC CM			226
IKIAE_C342_2B3_IGACV	QREDDVDETDRICKIINKVEED			KKVALLGVULL	KWKFDDVKIVDS	520
TRIAE_CS42_2DS_TGACV	QRFDDVDPTDRYCNHNRVFFD	ATTTTCTNCTOCE:	SYVGIGCME	RRVALYGVDP	PRWRPDDVKIVDS	326
TRIAE CS42 2AS TGACV	QRFDDVDPTDRYCNHNRVFFD	ATLLGLNGIQGP	SYVGTGCMF	RRVALYGVDPI	PRWRPDNVKIVDS	522
TRIAE CS42 7AL TGACV	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	GTLRALDGMOGP	IYVGTGCLF	RRITVYGFDP	PRINVGGPCFPRLAGLFAKTKYEKPGLE	496
TRIAF CS42 7DL TGACT	ORFECTORTOL	TTRAIDCMOCP	TVVCTCCLE		PTNUCCPCEDPIACIEAKTKVEKPCIE	95
TRIAL_CO42_701_TOACV			TIVOTOCIL			50
TRIAL_C542_/BL_TGACV	QRFEGVDPTDL MANHNR IEFD	TLRALDGMQGP.	TINGIGCLE	RRITVIGEDEE	RINVGGPCFPRLAGLFARTRIERPSLE	238
TRIAE CS42 5DL TGACV	DVA	TEADKFGISTPF	LGSVRAALG	GLNRSEOWNTTI	TKPPRSFDGAAVGEATAEVSCGYEDREA	502
TRIAF CS42 SAL TCACT	DVA	AFADKFOTOTO.	LCSURAALN	ILNOSFOWNTTS	- PPRSEDCA AVCEATAL VSCCVEDPEA	500
TRIAL_CO42_JAL_TOACV	DVA				DDDGEDGAAVGAAAAVSCGIDDAAA	500
TRIAE_CS42_5BL_TGACV	DVA	TEADKEGISTPE.	LVSVRAALN	ILNESEQWNTTS	S-PPRSFDGAAVGDATA VSCGYDDRIA	501
TRIAE_CS42_2AS_TGACv		RFGNSLLE	lnsvlaaik	(QEEG\	/TLQPPLDDSFLEEVTKVVSSSYDDSTD	565
TRIAE CS42 2BS TGACV		RFCNSLPE	LNSVLAAIK	QEEGN	/TLPPPLDDSFLEEMTKVVSSSYDDSTD	564
TRIAE CS42 2DS TGACT		REGNSLPD	LNSVI.AATK	OFEGN		557
TREE_0012_200_TONOV						537
TRIAL_CS42_ZAS_TGACV		KFGSSIPE	LESVSKAIN	IQER:	STIPPPISETLVAEMERVVSASHDKAIG	537
TRIAE_CS42_2DS_TGACv		RFGSSIPE	lesvskain	IQERS	STIPPPISETLVAEMERVVSASHDKATG	537
TRIAE CS42 2BS TGACV		RFCSSIPE	LDSVSKAIN	IQERS	STIPPPISETLVAEMERVVSASHDKATG	541
TRIAE CS42 2DL TGACV		TGEFGYSTSP	VNSVPDAAT	ODR	STTPVLVDEHLRKDLAT MTCAYDDGSS	537
TREE_COAL ODL TOROT		TODI GIGIOI	TNEUDDAT			220
IRIAE_C542_2BL_IGACV		IGErGISISH	INSVEDMAL	QDR	SIIPVLVDERLSKULAI MICAINDGSS	330
TRIAE_CS42_2AL_TGACv		AGEFCYSTSE	VNSVPDAAI	:QDR	SITPVLVDEGLRKDLTTEMTCAYEDGSS	540
TRIAE CS42 2DL TGACV		AHEFGNSTSF	TNSMPDGAI	QERS	SITPVLVDEGLINDLATLITCAYEDGSS	533
TRIAE CS42 2BS TGACV		VNEFGSSTSF	TNSMPDOAT	OER	STTPVLVDEALSNDLAT MTCAYDDGSS	546
TRIME_0012_200_TOMOV		VIUL CODIOL	TNOME			510
TRIAL_CS42_ZDS_TGACV			INSMPDGAL	QER:	SITPVLVDEALSNDLATEMICAIEDGSS	540
TRIAE_CS42_7AL_TGACv		AAEL C NSTPE	lk <mark>s</mark> ipd g ai	QERS	SITPVLVDEALTSDLATEMTCAYDDRES	523
TRIAE CS42 7DL TGACv		AAELCNSTPE	LNSIPDGAI	QERS	SITPVLVDEGLSNDIATLMTCTYEDGSS	521
TRIAE CS42 7BL TGACV		GAELCKSTPF	LNST PDCAT	ODR	STTPVSVDEGLMSDLAT MTCAYEDRES	515
TRIAE CS42 IL TCACTI	P	VDDNMEOVOTCH	TNOMDADAN			515
IKIAE_C342_0_IGACVI_		TREMMEGROIDE.		QERSV	FSFRIVGEABLADENICATEDGIE	545
TRIAE_CS42_1BS_TGACv	R	YRPNMF G KSTSE.	INSVPAAAN	IQERS\	/PSPATVGEABLADAMTCAYBDGHE	544
TRIAE CS42 2AS TGACV		PNKFGKSMTE	INSIPVAAN	IQERS\	/MSPVSLDEPATTELADVMTCAYEDGTE	551
TRIAE CS42 2BS TGACV		PNKFCKSMTF	TNSTPVAAN	IOERSN	MSPVSIDEPATTRLADVMTCAYEDGE	550
TRIAE COAS 2DO TOACT		DNKEGKGMTE	TNGTDVAAN			5/9
IKIAE_C342_2D3_IGACV				QERSV	MSFVSLDEFATTBLADVMTCATBDGTE	040
TRIAE_CS42_2BS_TGACV		STKFGSSASE	ISSILPAAL	QERSI	LMSPPALEEPVMADLAHVMTCAYEDGTE	3//
EDITE COLO ODO ECIO						
TRIAE CS42 2DS TGACV		STKFCSSASE	IS <mark>S</mark> ILPAAD	QERSI	IMSPPALEESVMADLAHVMTCAYEDGTE	377
TRIAE CS42_2DS_TGACV TRIAE CS42_2AS_TGACV		STKFCSSASF	IS <mark>S</mark> ILPAAD ISSILPAAD	QERS1 OERS1	IMSPPALEESVMADLAHVMTCAYEDGTE IMSPPALEEPVMADLAHVMTCAYEDGTE	377 573
TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV		STKFCSSASE STKFCSSASE UPKKTVCKSDAE	IS <mark>S</mark> ILPAAD IS <mark>S</mark> ILPAAD VDSIDRASH	QERS] QERS]	IMSPPALEESVMADLAHVMTCAYEDGIE IMSPPALEEPVMADLAHVMTCAYEDGIE	377 573
TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_7AL_TGACV	MTMAKAKAAPVPAKGKHGFLP	STKFGSSASF STKFGSSASF LPKKTYGKSDAF	IS <mark>S</mark> ILPAAD ISSILPAAD VD <mark>S</mark> IPRASH	DQERSI DQERSI IPSPYAAA	IMSPPALEESVMADLAHVMTCAYEDGTE IMSPPALEEPVMADLAHVMTCAYEDGTE IAEGIVADEATIVEAVNVTAAAFEKKTG	377 573 572
TRIAE CS42 2DS TGACV TRIAE CS42 2AS TGACV TRIAE CS42 7AL TGACV TRIAE CS42 7DL TGACV	MTMAKAKAAPVPAKGKHGFLP MTMAKAKAAPVPAKGKHGFLP	STKFGSSASE STKFGSSASE LPKKTYGKSDAE LPKKTYGKSDAE	ISSILPAAD ISSILPAAD VDSIPRASH VDSIPRASH	DQERSI DQERSI IPSPYAAP IPSPYAAP	IMSPPALEESVMADLAHVMTCAYEDGTE IMSPPALEEPVMADLAHVMTCAYEDGTE AAEGIVADEATIVEAVNVTAAAFEKKTG AAEGIVADEATIVEAVNVTAAAFEKKTG	377 573 572 171
TRIAE_CS42_ZDS_TGACV TRIAE_CS42_ZAS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7BL_TGACV	MTMAKAKAAPVPAKGKHGFLP MTMAKAKAAPVPAKGKHGFLP MTMAKAKAAPVPAKGKHGFLP	STKFGSSASD STKFGSSASD LPKKTYGKSDAD LPKKTYGKSDAD LPKKTYGKSDAD	ISSILPAAC ISSILPAAC VDSIPRASH VDSIPRASH VDSIPRASH	DQERS DQERS1 IPSPYAAZ IPSPYAAZ IPSPYAAZ	IMSPPALEESVMADLAHVMTCAYDDGTE IMSPPALEEPVMADLAHVMTCAYDDGTE AAEGIVADEATIVDAVNVTAAAFDKKTG AAEGIVADEATIVDAVNVTAAAFDKKTG AAEGIVADEATIVDAVNVTAAAFEKKTG	377 573 572 171 614
TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7BL_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP	STKFCSSASD STKFCSSASD LPKKTYCKSDA LPKKTYCKSDA LPKKTYCKSDA	ISSILPAAD ISSILPAAD VDSIPRASH VDSIPRASH VDSIPRASH	DERSI DERSI IPSPYAAF IPSPYAAF IPSPYAAF	IMSPPALEESVMADLAHVMTCAYEDGTE IMSPPALEEPVMADLAHVMTCAYEDGTE AAEGIVADEATIVEAVNVTAAAFEKKTG AAEGIVADEATIVEAVNVTAAAFEKKTG AAEGIVADEATIVEAVNVTAAAFEKKTG	377 573 572 171 614
TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_5DL_TGACV	MTMAKAKAAPVPAKGKHGFLP MTMAKAKAAPVPAKGKHGFLP MTMAKAKAAPVPAKGKHGFLP	STKFCSSASD STKFCSSASD LPKKTYCKSDAF LPKKTYCKSDAF LPKKTYCKSDAF	ISSILPAAC ISSILPAAC VDSIPRASH VDSIPRASH VDSIPRASH TAPDAFRGI	QERSJ DQERSJ IPSPYAAF IPSPYAAF IPSPYAAF	IMSPPALEESVMADLAHVMTCAYDGTE IMSPPALEEPVMADLAHVMTCAYDGTE IAEGIVADEATIVBAVNVTAAAFSKKTG AAEGIVADEATIVBAVNVTAAAFSKKTG AAEGIVADEATIVBAVNVTAAAFEKKTG	377 573 572 171 614 582
TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5AL_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDLGVII YGTVID ATCHC	STKFGSSASE STKFGSSASE LPKKTYGKSDAE LPKKTYGKSDAE LPKKTYGKSDAE MHRRGMRSAYCA MHRRGMRSAYCA	ISSILPAAC ISSILPAAC VDSIPRASH VDSIPRASH VDSIPRASH TAPDAFRGI	QERSJ QERSJ IPSPYAAZ IPSPYAAZ IPSPYAAZ APINLTDRLIG	IMSPPALEESVMADLAHVMTCAYDDGTE IMSPPALEEPVMADLAHVMTCAYDDGTE AAEGIVADEATIVDAVNVTAAAFDKKTG AAEGIVADEATIVDAVNVTAAAFDKKTG AAEGIVADEATIVDAVNVTAAAFEKKTG VLRWAAGSLEIFFSRNNALLAGARLHP	377 573 572 171 614 582 580
TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5AL_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGMIYGTVTIDIATGFC MGRDIGMIYGTVTIDIATGFC	stkfcssass stkfcssass LpkktyckSda LpkktyckSda LpkktyckSda LpkktyckSda MHRRGWRSAYCA MHRRGWSSAYCA	ISSILPAAC ISSILPAAC VDSIPRASH VDSIPRASH VDSIPRASH TAPDAFRGT TAPDAFRGT	QERSJ QERSJ IPSPYAAZ IPSPYAAZ IPSPYAAZ IPSPYAAZ	IMSPPALEESVMADLAHVMTCAYDDGTE IMSPPALEEPVMADLAHVMTCAYDDGTE AAEGIVADEATIVDAVNVTAAAFEKKTG AAEGIVADEATIVDAVNVTAAAFEKKTG AAEGIVADEATIVDAVNVTAAAFEKKTG VIRWAGSLEIFFSRNNALLAGARLHP VVIRWAGSLEIFFSRNNALLAGARLHP	377 573 572 171 614 582 580
TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5AL_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGHIYGTVTIDIATGEC MGRDIGHIYGTVTIDIATGEC MGRDIGHIYGTVTIDIATGEC	STKFGSSASB STKFGSSASB LPKKTYGKSDAS LPKKTYGKSDAS LPKKTYGKSDAS MHRRGWRSAYCA MHRRGWRSAYCA MHRRGWRSAYCA	ISSILPAAC ISSILPAAC VDSIPRASH VDSIPRASH VDSIPRASH TAPDAFRGI TAPDAFRGI TAPD <mark>AF</mark> RGI	DERSI DORSSAA IPSPYAA IPSPYAA IPSPYAA IPSPYAA IPSPINLTDRLHO APINLTDRLHO APINLTDRLHO APINLTDRLHO	IMSPPALEESVMADLAHVMTCAYDDGTE IMSPPALEEPVMADLAHVMTCAYDDGTE AAEGIVADEATIVBAVNVTAAAFBKKTG AAEGIVADEATIVBAVNVTAAAFBKKTG AAEGIVADEATIVBAVNVTAAAFBKKTG VIRWAAGSLEIFFSRNNALLAGARLHP VIRWAAGSLEIFFSRNNALLAGARLHP	377 573 572 171 614 582 580 581
TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_5BL_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGWIYGTVTID ATGFC MGRDIGWIYGTVTID ATGFC MGRDIGWIYGTVTID ATGFC WGRGIGYIYNMATID VTGFR	STKFCSGASE STKFCSGASE LPKKTYCKSDAF LPKKTYCKSDAF LPKKTYCKSDAF MHRRCWRSAYCA MHRRCWRSAYCA MHRRCWRSAYCA LHCOCWRSAYATI	ISSILPAAC ISSILPAAC VDSIPRASH VDSIPRASH VDSIPRASH TAPDAFRGI TAPDAFRGI TAPDAFRGI MEREAFRGI	DERS DQERSD IPSPYAA IPSPYAA IPSPYAA IPSPYAA IPTNLTDRLH IPTNLTDRLH IPTNLTBRLH IPTNLTBRLH	IMSPPALEESVMADLAHVMTCAYDGTE IMSPPALEEPVMADLAHVMTCAYDGTE AAEGIVADEATIVBAVNVTAAAFBKKTG AEGIVADEATIVBAVNVTAAAFBKKTG VLRWAAGSLEIFFSRNNALLAGARLHP VLRWAAGSLEIFFSRNNALLAGARLHP VLRWAGSLEIFFSRNNALLAGARLHP IVRWAGSLEIFFSRNNALLAGARLHP IVRWAGSLEIFFSRNNALLAGARLHP	 377 573 572 171 614 582 580 581 645
TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_5BL_TGACV TRIAE_CS42_2AS_TGACV	MTMAKAKAAPVPAKGKHGFLP MTMAKAKAAPVPAKGKHGFLP MTMAKAKAAPVPAKGKHGFLP MGRDIGMIYGTVTID ATGFC MGRDIGWIYGTVTID ATGFC MGRDIGWIYGTVTID ATGFC MGRGIGYIYNMATID VTGFR	STKFCSGASE STKFCSGASE LPKKTYCKSDAF LPKKTYCKSDAF HRRCWRSAYCA' MHRRCWRSAYCA' MHRRCWRSAYCA' ILGCOWRSYWATT LIHCCOWRSYWYTT	ISSILPMAD ISSILPMAD VDSIPRASH VDSIPRASH VDSIPRASH TAPDARGT TAPDARGT MEREABRGT MEREABRGT	DQERS DQERS IPSPYAAA IPSPYAAA IPSPYAAA IPSPYAAA IPSPYAAA IPSPYAAA IPSPYAAA IPSPYAAA IPSPYAAA IPSPY	IMSPPALEESVMADLAHVMTCAYDGTE IMSPPALEEPVMADLAHVMTCAYDGTE AAEGIVADEATIVBAVNVTAAAFBKKTG AAEGIVADEATIVBAVNVTAAAFBKKTG VLRWAGSLEIFFSRNNALLAGARLHP VVLRWAGSLEIFFSRNNALLAGARLHP VVLRWAGGSLEIFFSRNNALLAGARLHP JVVRWSGGSLEMFFSHISPLFAGRRLSL	377 573 572 171 614 582 580 581 645 644
TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2DS_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGHIYGTVTID ATGEC MGRDIGHIYGTVTID ATGEC MGRGIGYIYNMATID VTGFR MGRGIGYIYNMATID VTGFR MGRGIGYIYNMATID VTGFR	STKFCSSASB LPKKTYCKSDAB LPKKTYCKSDAB LPKKTYCKSDAB LPKKTYCKSDAB MHRROWSSAYCA MHRROWSSAYCA MHRROWSSAYCA LIECCOWRSYYATT LIECCOWRSYYATT	ISSILPMAL ISSILPMAL VDSIPRASH VDSIPRASH TAPDARGI TAPDARGI TAPDARGI MEREARGI MEREARGI MEREARGI	DQERSD DQERSD HPSPYAA HPSPYAA APINLTDRLL APINLTDRLL APINLTDRLL APINLTBRLL APINLTBRLL APINLTBRLL	IMSPPALEESVMADLAHVMTCAYDDGTE IMSPPALEEPVMADLAHVMTCAYDDGTE AAEGIVADEATIVBAVNVTAAAFBKKTG AAEGIVADEATIVBAVNVTAAAFBKKTG VLRWAAGSLEIFFSRNNALLAGARLHP VLRWAGSLEIFFSRNNALLAGARLHP VLRWAGSLEIFFSRNNALLAGARLHP IVRWGGSLEMFFSHISPLFAGRRLSL IVRWGGSLEMFFSHISPLFAGRRLSL IVRWGGSLEMFFSHISPLFAGRRLSL	377 573 572 171 614 582 580 581 645 644 637
TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGMIYGTVTID ATGEC MGRDIGMIYGTVTID ATGEC MGRGIGVIYNMATID VTGFR MGRGIGVIYNMATID VTGFR MGRGIGVIYNMATID VTGFR	STKFGSBASB STKFGSBASB LPKKTGKSDAB LPKKTYGKSDAB LPKKTYGKSDAB MHRRGMRSBACA MHRRGMRSBACA MHRRGMRSBACA HRCGRBSMATH LHGGCMRSMATH LHGGCMRSMATH LHGGCMRSMATH	ISSILPMAL ISSILPMASH VUSIPRASH VUSIPRASH VUSIPRASH TAPDASRGT TAPDASRGT MEREASRGT MEREASRGT MEREASRGT	DERSD DQERSD IPSPYAA IPSPYAA IPSPYAA APINLTDRLL APINLTDRLL APINLTDRLL APINLTBRLL APINLTBRLL APINLTBRLL APINLTBRLL	IMSPPALEESVMADLAHVMTCAYDDGTE IMSPPALEEPVMADLAHVMTCAYDDGTE AAEGIVADEATIVBAVNVTAAAFBKKTG AEGIVADEATIVBAVNVTAAAFBKKTG VIRWAAGSLEIFFSRNNALLAGARLHP VIRWAAGSLEIFFSRNNALLAGARLHP VIRWAGGSLEMFFSNISPLFAGRRLSL IVRWSGGSLEMFFSHISPLFAGRRLSL IVRWSGGSLEMFFSNISPLFAGRRLSL IVRWSGGSLEMFFSNISPLFAGRRLSL	377 573 572 171 614 582 580 581 645 644 637 617
TRIAE_CS42_2DS_TGACV TRIAE_CS42_AS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV	MTMAKAKAAPVPAKGKHGFLP MTMAKAKAAPVPAKGKHGFLP MTMAKAKAAPVPAKGKHGFLP MGRDIGMIYGTVTID ATGFC MGRDIGMIYGTVTID ATGFC MGRGIGYIYNMATID VTGFR MGRGIGYIYNMATID VTGFR MGRGIGYIYNMATID VTGFR MGRGIGYIYDNATID VTGFR	STKFCSGASE STKFCSGASE LPKKTYCKSDAE LPKKTYCKSDAE MHRRCMSSATCA' MHRRCMSSATCA' MHRRCMSSATCA' IHCCOMRSYATT IHCCOMRSYATT IHCCOMRSYATT	ISSILPMAL ISSILPMAL VDSIPRASH VDSIPRASH VDSIPRASH TAPDARGI MEREARGI MEREARGI MEREARGI MEREARGI	ADERS	IMSPPALEESVMADLAHVMTCAYDGTE IMSPPALEEPVMADLAHVMTCAYDGTE AAEGIVADEATIVBAVNVTAAAFEKKTG AAEGIVADEATIVBAVNVTAAAFEKKTG VLRW.AGSLEIFFSRNNALLAGARLHP VLRW.AGSLEIFFSRNNALLAGARLHP VLRW.AGSLEIFFSRNNALLAGARLHP IVRW.GGSLEMFFSHISPLFAGRRLSL IVRW.GGSLEMFFSHISPLFAGRRLSL IVRW.GGSLEMFFSHISPLFAGRRLSL IVRW.GGSLEMFFSHISPLFAGRRLSL	377 573 572 171 614 582 580 581 645 644 637 617
TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2DS_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGHIYGTVTID ATGEC MGRDIGHIYGTVTID ATGEC MGRGIGYIYMATID VTGFR MGRGIGYIYMATID VTGFR MGRGIGYIYMATID VTGFR MGRGIGYIYDIATID VTGFR MGRGGVIYDIATID VTGFR	STKFCSGASB STKFCSGASB LPKKTYCKSDAB LPKKTYCKSDAB LPKKTYCKSDAB MHRRONGSAYCA MHRRONGSAYCA MHRRONGSAYCA II.GCONGSAYATI II.GCONGSAYATI II.GCONGSAYCTI II.GCONGSAYCTI II.GCONGSAYCTI	ISSILPAAL ISSILPAAL VUSIPRASH VUSIPRASH VUSIPRASH TAPDABRGT TAPDABRGT MEREABRGT MEREABRGT MEREABRGT MEREABRGT MEREABRGT MEREABRGT	DQERSD DQERSD HPSPYAA HPSPYAA APINLTDRLL APINLTDRLL APINLTDRLL APINLTBRLL APINLTBRLL APINLTBRLL APINLTBRLL APINLTBRLL APINLTBRLL	IMSPPALEESVMADLAHVMTCAYDDGTE IMSPPALEEPVMADLAHVMTCAYDDGTE AAEGIVADEATIVBAVNVTAAAFBKKTG AEGIVADEATIVBAVNVTAAAFBKKTG VLRWAGSLEIFFSRNNALLAGARLHP VLRWAGSLEIFFSRNNALLAGARLHP VLRWAGSLEIFFSRNNALLAGARLHP IVRWSGGSLEMFFSHISPLFAGRRLSL IVRWSGGSLEMFFSHISPLFAGRRLSL IVRWSGGSLEMFFSHISPLFAGRRLSL IVRWSGGSLEMFFSLINPLIGGRRIQA	377 573 572 171 614 582 580 581 645 644 637 617 617
TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGHIYGTVTID ATGEC MGRDIGHIYGTVTID ATGEC MGRGIGYIYNMATID VTGER MGRGIGYIYNMATID VTGER MGRGIGYIYNMATID VTGER MGRGIGYIYDIATID VTGER MGKGVGYIYDIATID VTGER	STKFCSGASE STKFCSGASE LPKKTYCKSDAF LPKKTYCKSDAF LPKKTYCKSDAF MHRRCWRSAYCA MHRRCWRSAYCA MHRCWRSAYCA IHCOCWRSWATT IHCOCWRSWYCTT IHCOCWRSWYCTT IHCOCWRSWYCTT	ISSILPMAL ISSILPMACH VUSIPRASH VUSIPRASH TAPDABRGT TAPDABRGT MEREABRGT MEREABRGT MEREABRGT MERDABCGI MERDABCGI	QERS QERS IPSPYAA IPSPYAA IPSPYAA APINLTDRL APINLTDRL APINLTDRL APINLTDRL APINLTDRL APINLTDRL APINLTDRL APINLTDRL	IMSPPALEESVMADLAHVMTCAYDGTE IMSPPALEEPVMADLAHVMTCAYDGTE AAEGIVADEATIVBAVNVTAAAFBKKTG AEGIVADEATIVBAVNVTAAAFBKKTG VLRWAGSLEIFFSRNNALLAGARLHP VLRWAGSLEIFFSRNNALLAGARLHP VLRWAGSLEIFFSRNNALLAGARLHP IVRWSGGSLEMFFSHISPLFAGRRLSL IVRWSGGSLEMFFSHISPLFAGRRLSL IVRWSGGSLEMFFSHISPLFAGRRLSL IVRWSGGSLEMFFSLNNPLIGGRRIQS IVRWSGGSLEMFFSLNNPLIGGRRIQA	377 573 572 171 614 582 580 581 645 644 637 617 617 621
TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGHIYGTVTID ATGFC MGRDIGHIYGTVTID ATGFC MGRGIGYIYNMATID VTGFR MGRGIGYIYNMATID VTGFR MGRGVGYIYDIATID VTGFR MGRGVGYIYDIATID VTGFR MGRGVGYIYDIATID VTGFR MGRGVGYIYDIATID VTGFR	STKFCSGASB STKFCSGASB LPKKTYCKGDAF LPKKTYCKGDAF LPKKTYCKGDAF MRRCWRSATCA MRRCWRSATCA MRRCWRSATCA MRCWRSATCA IHCOWRSATCA IHCOWRSATCA IHCOWRSATCTI IHCOWRSATCTI IHCOWRSATCTI IHCOWRSATCTI IHCOWRSATCTI IHCOWRSATCTI IHCOWRSATCTI IHCOWRSATCTI IHCOWRSATCTI IHCOWRSATCTI IHCOWRSATCTI IHCOWRSATCTI IHCOWRSATCTI IHCOWRSATCTI IHCOWRSATCTI	ISSILPAAL ISSILPAAL VUSIPRASH VUSIPRASH VUSIPRASH TAPDARGI TAPDARGI MEREARGI MEREARGI MEREARGI MEREARGI MEREARGI MERCASCGI MERCASCGI MERARGI	DQERSD DQERSD HPSPYAA HPSPYAA HPSPYAA APINLTDRLH APINLTDRLH APINLTDRLH APINLTDRLH APINLTDRLH APINLTDRLH APINLTDRLH APINLTDRLH APINLTDRLH APINLTDRLH	IMSPPALEESVMADLAHVMTCAYDDGTE IMSPPALEEPVMADLAHVMTCAYDDGTE AAEGIVADEATIVBAVNVTAAAFBKKTG AAEGIVADEATIVBAVNVTAAAFBKKTG VIRWAGSLEIFFSRNNALLAGARLHP VIRWAGSLEIFFSRNNALLAGARLHP VIRWAGGSLEMFFSHISPLFAGRRLSL IVRWSGGSLEMFFSHISPLFAGRRLSL IVRWSGGSLEMFFSHISPLFAGRRLSL IVRWSGGSLEMFFSLNNPLIGGRRIQS IVRWSGGSLEMFFSLNNPLIGGRRIQS IVRWSGGSLEMFFSLNNPLIGGRRIQS IVRWSGGSLEMFFSLNNPLIGGRRIQS IVRWSGGSLEMFFSLNNPLIGGRRIQS IVRWSGGSLEMFFSLNNPLIGGRRIQS	377 573 572 171 614 582 580 581 645 644 637 617 617 621 617
TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGHIYGTVTID ATGFC MGRDIGHIYGTVTID ATGFC MGRGIGYIYNMATID VTGFR MGRGIGYIYNMATID VTGFR MGRGKGYIYDIATID VTGFR MGRGKGYIYDIATID VTGFR MGRGKGYIYDIATID VTGFR MGRGAGWYNIATID VTGFR	STKFCSGASE STKFCSGASE LPKKTYCKSDAE LPKKTYCKSDAE LPKKTYCKSDAE MHRRCMRSAYCA MHRRCMRSAYCA MHRCMRSAYCA LHCOCMRSMYCTI LHCOCMRSMYCTI LHCCOMRSMYCTI LHCCOMRSMYCTI LHCCOMRSMYCTI LHCCOMRSMYCTI LHCCOMRSMYCTI	ISSILPAAL ISSILPAAL VUSIPRASH VUSIPRASH VUSIPRASH TAPDARGT TAPDARGT MEREARGT MEREARGT MEREARGT MEREARGT MEREARGT MEREARGT MEREARGT	QERS QERS IPSPYAA IPSPYAA IPSPYAA APINITORI APINITORI APINITORI APINITORI APINITORI APINITORI APINITORI APINITORI APINITORI APINITORI APINITORI	IMSPPALEESVMADLAHVMTCAYDDGTE IMSPPALEEPVMADLAHVMTCAYDDGTE AAEGIVADEATIVDAVNVTAAAFDKKTG AEGIVADEATIVDAVNVTAAAFDKKTG WLRWAGSLEIFFSRNNALLAGARLHP VLRWAGSLEIFFSRNNALLAGARLHP VLRWAGSLEIFFSRNNALLAGARLHP VLRWSGGSLEMFFSHISPLFAGRRLSL IVRWSGGSLEMFFSHISPLFAGRRLSL IVRWSGGSLEMFFSLNNPLIGGRRIQA IVRWSGGSLEMFFSLNNPLIGGRRIQA IVRWSGGSLEMFFSLNNPLIGGRRIQA IVRWSGGSLEVFFSHSNALIASRRLHP VLRWSGGSLEVFFSHSNALIASRRLHP	377 573 572 171 614 582 580 581 645 644 637 617 617 621 617 418
TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_5BL_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2DS_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGMIYGTVTID ATGEC MGRDIGMIYGTVTID ATGEC MGRGIGYIYNMATID VTGER MGRGIGYIYNMATID VTGER MGRGIGYIYNMATID VTGER MGRGIGYIYDIATID VTGER MGRGVGYIYDIATID VTGER MGRGVGYIYDIATID VTGER MGRDAGWYNIATID VTGER	STKFCSGASB STKFCSGASB LPKKTYCKGDAF LPKKTYCKGDAF LPKKTYCKGDAF MIRCWRSAYCA MIRCWRSAYCA MIRCWRSAYCA IIGCORSYATT IIGCORSYATT IIGCORSYATT IIGCORSYATT IIGCORSYATT IIGCORSYATT IIGCORSYATT IIGCORSYATT IIGCORSYATT IIGCORSYATT IIGCORSYATT IIGCORSYATT	ISSILPMAL ISSILPMAL VUSIPRASH VUSIPRASH VDSIPRASH TAPDABRGT TAPDABRGT MEREABRGT MEREABRGT MEREABRGT MERDABCGI MERDAGCGI MERABRGT	QERS DQERSD IPSPYAA IPSPYAA IPSPYAA APINLTORLI APINLTORLI APINLTORLI APINLTORLI APINLTORLI APINLTORLI APINLTORLI APINLTORLI APINLTORLI APINLTORLI	IMSPPALEESVMADLAHVMTCAYDGTE IMSPPALEEVMADLAHVMTCAYDGTE AAEGIVADEATIVDAVNVTAAAFDKKTG AAEGIVADEATIVDAVNVTAAAFDKKTG VLRW.AGSLEIFFSRNNALLAGARLHP VLRW.AGSLEIFFSRNNALLAGARLHP VLRW.AGSLEIFFSRNNALLAGARLHP VLRW.GGSLEMFFSHISPLFAGRRLSL IVRW.GGSLEMFFSHISPLFAGRRLSL IVRW.GGSLEMFFSHISPLFAGRRLSL IVRW.GGSLEMFFSLNNPLIGGRRIQS IVRW.GGSLEMFFSLNNPLIGGRRIQS VLRW.GGSLEMFFSLNNPLIGGRRIQA IVRW.GGSLEVFFSNNALIASRRLHP VLRW.GGSLEVFFSNNALIASRRLHP	377 573 572 171 614 582 580 581 645 645 645 645 617 617 621 617 418
TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2AL_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGHIYGTVTID ATGEC MGRDIGHIYGTVTID ATGEC MGRGIGYIYMATID VTGFR MGRGIGYIYMATID VTGFR MGRGGGYIYMATID VTGFR MGRGGGYIYDIATID VTGFR MGRGGVGYIYDIATID VTGFR MGRGGVYYNIATID VTGFR MGRDAGWYNIATID VTGFR MGRDAGWYNIATID VTGFR	STKFCSGASB STKFCSGASB LPKKTYCKSDAF LPKKTYCKSDAF LPKKTYCKSDAF MHRROWSGAYCA MHRROWSGAYCA HRCOWSGAYCA IHCCOWRSYYCT IHC	ISSILPAAL ISSILPAAL VUSIPRASH VUSIPRASH VUSIPRASH TAPDARGT TAPDARGT TAPDARGT MEREARGT MEREARGT MEREARGT MEREARGT MEREARGT MEREARGT MEPAARGT MEPAARGT	QERS QERS IPSPYAA IPSPYAA IPSPYAA APINLTDRLL APINLTDRLL APINLTDRLL APINLTBRLL APINLTBRLL APINLTBRLL APINLTBRLL APINLTBRLL APINLTBRLL APINLTBRLL APINLTBRLL	IMS PPALEES VMAD LAHVMTCAY DGTE IMS PPALEES VMAD LAHVMTCAY DGTE AAEGIVADEATIV BAVNVTAAAF BKKTG AAEGIVADEATIV BAVNVTAAAF BKKTG AUGUVADEATIV BAVNVTAAAF BKKTG VIRWA AGSLEIFFSRNNALLAGARLHP VIRWA AGSLEIFFSRNNALLAGARLHP VIRWA GGSLEMFFSHISPLFAGRRLSL IVRWS GGSLEMFFSHISPLFAGRRLSL IVRWS GGSLEMFFSLINPLIGGRRIQS IVRWS GGSLEMFFSLNNPLIGGRRIQS IVRWS GGSLEMFFSLNNPLIGGRRIQS VIRWS GGSLEVFFSHNNALIASRRLHP VIRWS GGSLEVFFSHNNALIASRRLHP VIRWS GGSLEVFFSHNNALIASRRLHP	377 573 572 171 614 582 580 581 645 644 637 617 617 621 617 418 620
TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGHIYGTVTID ATGFC MGRDIGHIYGTVTID ATGFC MGRGIGYIYNMATID VTGFR MGRGIGYIYNMATID VTGFR MGRGVGYIYDIATID VTGFR MGRGVGYIYDIATID VTGFR MGRGVGYIYDIATID VTGFR MGRDAGWYNIATID VTGFR MGRDAGWYNIATID VTGFR MGRDAGWYNIATID VTGFR MGRDAGWYNIATID VTGFR	STKFGSGASE STKFGSGASE LPKKTYGKSDAE LPKKTYGKSDAE LPKKTYGKSDAE MHRRGWRSGAYCA MHRRGWRSGAYCA HRGGWRSYGTT IHGGGWRSYGTT IHGGWRSYGTT IHGGWRSYGTT IHGGWRSYGTT IHGGWRSYGTT IHGGWRSYGTT IHGGWRSYGTT IHRGWRSYGST IHRGWRSYGST IHRGWRSYGST IHRGWRSYGST IHRGWRSYGST IHRGWRSYGST IHRGWRSYGST IHRGWRSYGST IHRGWRSYGST IHRGWRSYGST IHRGWRSYGST IHRGWRSYGST	ISSILPMAL ISSILPMAL ISSILPMAL VDSIPRASH VDSIPRASH TAPDARGT TAPDARGT MEREARGT MEREARGT MEREARGT MERDAGGI MERARGT MEPAARGT MEPAARGT MEPAARGT	QERS QERS IPSPYAA IPSPYAA IPSPYAA APINLTDRLI APINLTDRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI	IMSPPALEESVMADLAHVMTCAYDDGTE IMSPPALEEPVMADLAHVMTCAYDDGTE AAEGIVADEATIVDAVNVTAAAFDKKTG AEGIVADEATIVDAVNVTAAAFDKKTG WIRWAGSLEIFFSRNNALLAGARLHP VIRWAGSLEIFFSRNNALLAGARLHP VIRWAGSLEIFFSRNNALLAGARLHP IVRWSGGSLEMFFSHISPLFAGRRLSL IVRWSGGSLEMFFSHISPLFAGRRLSL IVRWSGGSLEMFFSLNNPLIGGRRIQS IVRWSGGSLEMFFSLNNPLIGGRRIQS IVRWSGGSLEVFFSHSNALIASRRLHP VIRWSGGSLEVFFSHSNALIASRRLHP VIRWSGGSLEVFFSHSNALIASRRLHP VIRWSGGSLEVFFSHSNALIASRRLHP VIRWSGGSLEVFFSHSNALIASRRLHP	3777 5733 572 1711 614 582 580 581 645 644 637 617 617 621 617 418 620 613
TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5BL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DS_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGWIYGTVTID ATGFC MGRDIGWIYGTVTID ATGFC MGRDIGWIYGTVTID ATGFC MGRGIGYIYNMATID VTGFR MGRGIGYIYNMATID VTGFR MGRGIGYIYNMATID VTGFR MGRGVGYIYDIATID VTGFR MGRDAGWYNIATID VTGFR MGRDAGWYNIATID VTGFR MGRDAGWYNIATID VTGFR MGRDAGWYNIATID VTGFR MGRDAGWYNIATID VTGFR MGRDAGWYNIATID VTGFR MGRDAGWYNIATID VTGFR	STKFCSGASE STKFCSGASE LPKKTYCKSDAE LPKKTYCKSDAE LPKKTYCKSDAE MIRREWRSAYCA MIRREWRSAYCA MIRREWRSAYCA IHCOMRSYACA IHCOMRSYACA IHCOMRSYACT IHCOMRSYACT IHCOMRSYACT IHCOMRSYACT IHCOMRSYACT IHCOMRSYACT IHCOMRSYACT IHCOMRSYACT IHCOMRSYACT IHCOMRSYACT IHCOMRSYACT IHCOMRSYACT IHCOMRSYACT IHCOMRSYACT IHCOMRSYACT IHCOMRSYACT IHCOMRSYACT IHCOMRSYACT	ISSILPMAL ISSILPMAL ISSILPMASH VDSIPRMSH VDSIPRMSH TAPDARRT TAPDARRT MEREARRT MEREARRT MEREARRT MEREARRT MEREARRT MEREARRT MEPAARRT MEPAARRT MEPAARRT MEPAARRT	QQERS) QQERS) IPSPYAAA IPSPYAAA IPSPYAAA APINLTDRLI APINLTDRLI APINLTDRLI APINLTDRLI APINLTDRLI APINLTDRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI	IMSPPALEESVMADLAHVMTCAYDGTE IMSPPALEEVMADLAHVMTCAYDGTE AAEGIVADEATIVDAVNVTAAAFDKKTG AAEGIVADEATIVDAVNVTAAAFDKKTG VLRW.AGSLEIFFSRNNALLAGARLHP VLRW.AGSLEIFFSRNNALLAGARLHP VLRW.AGSLEIFFSRNNALLAGARLHP VLRW.GGSLEMFFSHISPLFAGRRLSL IVRW.GGSLEMFFSHISPLFAGRRLSL IVRW.GGSLEMFFSHISPLFAGRRLSL IVRW.GGSLEMFFSHISPLFAGRRLSL IVRW.GGSLEMFFSHINPLIGGRRIQS VLRW.GGSLEVFFSHNNALIASRRLHP VLRW.GGSLEVFFSHNNALIASRLHP VLRW.GGSLEVFFSHNNALIASRLHP VLRW.GGSLEVFFSHNNALIAGRRLHP VLRW.GGSLEVFFSHNNALIAGRRLHP VLRW.GGSLEVFFSHNNALIAGRRLHP	3777 573 572 171 614 582 580 645 645 645 645 647 617 617 617 617 617 621 617 418 620 613 626
TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_5BL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGHIYSTVTID ATGEC MGRDIGHIYSTVTID ATGEC MGRGIGYIYMMATID VTGER MGRGIGYIYMMATID VTGER MGRGIGYIYMMATID VTGER MGRGIGYIYDIATID VTGER MGRGVGYIYDIATID VTGER MGRDAGWYNIATID VTGER MGRDAGWYNIATID VTGER MGRDAGWYNIATID VTGER MGRDAGWYNIATID VTGER MGRDAGWYNIATID VTGER MGRDAGWYNIATID VTGER MGRDAGWYNIATID VTGER	STKFCSBASE STKFCSBASE LPKKTYCKSDAE LPKKTYCKSDAE LPKKTYCKSDAE LPKKTYCKSDAE MHRROWSBAYCA MHRROWSBAYCA HRCOWSWATT HC	ISSILPAAL VUSIPRASH VUSIPRASH VUSIPRASH VUSIPRASH TAPDAPRGT TAPDAPRGT MEREAPRGT MEREAPRGT MEREAPRGT MEREAPRGT MEPAAPRGT MEPAAPRGT MEPAAPRGT MEPAAPRGT	QQERS	IMS PPALEES VMAD LAHVMTCAY DGTE IMS PPALEES VMAD LAHVMTCAY DGTE AAEGIVADEATIV BAVNVTAAAF BKKTG AAEGIVADEATIV BAVNVTAAAF BKKTG AEGIVADEATIV BAVNVTAAAF BKKTG VLRWA AGSLEIFFSRNNALLAGARLHP VLRWA AGSLEIFFSRNNALLAGARLHP VLRWA GGSLEMFFSHISPLFAGRRLSL IVRWS GGSLEMFFSHISPLFAGRRLSL IVRWS GGSLEMFFSLNNPLIGGRRIQA IVRWS GGSLEMFFSLNNPLIGGRRIQA IVRWS GGSLEVFFSLNNLIAGRRLHP VLRWS GGSLEVFFSHNNALIASRRLHP VLRWS GGSLEVFFSHNNALIASRRLHP VLRWS GGSLEVFFSHNNALIAGRRLHP VLRWS GGSLEVFFSHNNALIAGRRLHP VLRWS GGSLEVFFSHNNALIAGRRLHP VLRWS GGSLEVFFSHNNALIAGRRLHP VLRWS GGSLEVFFSHNNALIAGRRLHP VLRWS GGSLEVFFSHNNALIAGRRLHP VLRWS GGSLEVFFSHNNALIAGRRLHP VLRWS GGSLEVFFSHNNALMAGRRLHP	3777 573 572 171 614 582 580 645 645 644 637 617 617 617 621 617 613 620 613 626 626 626
TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGHIYGTVTID ATGEC MGRDIGHIYGTVTID ATGEC MGRGIGYIYNMATID VTGER MGRGIGYIYNMATID VTGER MGRGIGYIYNMATID VTGER MGRGIGYIYNATID VTGER MGRGVGYIYDIATID VTGER MGRDAGWYNIATID VTGER MGRDAGWYNIATID VTGER MGRDAGWYNIATID VTGER MGRDGWYNIATID VTGER MGRDGWYNIATID VTGER MGRDGWYNIATID VTGER MGRDGWYNIATID VTGER MGRDGWYNIATID VTGER	STKFGSGASU STKFGSGASU LPKKTYGKSDAU LPKKTYGKSDAU LPKKTYGKSDAU LPKKTYGKSDAU MHRRGWRSGAYCA MHRRGWRSGAYCA HGGCWRSYMCTI HGGCWRSYMCTI HGGCWRSYMCTI HGGCWRSYMCTI HRGCWRSYMCSI HRQCWRSYMCSI MHRQGWRSYMCSI MHRQCWRSYMCSI MHRQCWRSYMCSI	ISSILPAL ISSILPAL ISSILPAL VUSIPRASH VUSIPRASH VUSIPRASH TAPDARGT MEREARGT MEREARGT MEREARGT MERARGT MERARGT MERARGT MEPALARGT MEPALARGT MEPALARGT MEPALARGT MEPALARGT	QERS QERS IPSPYAA IPSPYAA IPSPYAA APINLTDRLI APINLTDRLI APINLTDRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI APINLTBRLI	IMS PPALEES VMAD LAHVMTCAY DGTE IMS PPALEES VMAD LAHVMTCAY DGTE AAEGIVADEATIV BAVNVTAAAF KKTG AAEGIVADEATIV BAVNVTAAAF KKTG WIRWAGSLEIFFSRNNALLAGARLHP VIRWAGSLEIFFSRNNALLAGARLHP VIRWAGSLEIFFSRNNALLAGARLHP IVRWSGGSLEMFFSHISPLFAGRRLSL IVRWSGGSLEMFFSHISPLFAGRRLSL IVRWSGGSLEMFFSLNNPLIGGRRIQS VIRWSGGSLEVFFSHSNALLASRRLHP VIRWSGGSLEVFFSHSNALIASRRLHP VIRWSGGSLEVFFSHSNALIASRRLHP VIRWSGGSLEVFFSHSNALIASRRLHP VIRWSGGSLEVFFSHSNALIASRRLHP VIRWSGGSLEVFFSHSNALIASRRLHP VIRWSGGSLEVFFSHSNALIASRRLHP VIRWSGGSLEVFFSHSNALIASRRLHP VIRWSGGSLEVFFSHSNALIASRRLHP VIRWSGGSLEVFFSHSNALIASRRLHP VIRWSGGSLEVFFSHSNALIASRRLHP VIRWSGGSLEVFFSHSNALIASRRLHP	3777 5733 572171 614 5822580 581 645 645 644 637 617 621 617 621 617 613 626 626 626 626 603
TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGHIYGTVTID ATGEC MGRDIGHIYGTVTID ATGEC MGRDIGHIYGTVTID ATGEC MGRGIGYIYNMATID VTGER MGRGGYYYNMATID VTGER MGRGGYYYDIATID VTGER MGRDGWYYNIATID VTGER MGRDAGWYNIATID VTGER MGRDAGWYNIATID VTGER MGRDAGWYNIATID VTGER MGRDAGWYNIATID VTGER MGRDGWYNIATID VTGER MGRDLGWYNIATID VTGER MGRDLGWYNIATID VTGER MGRDLGWYNIATID VTGER MGRDLGWYNIATID VTGER MGRDLGWYNIATID VTGER	STKFCSGASB STKFCSGASB LPKKTYCKGDAF LPKKTYCKGDAF LPKKTYCKGDAF LPKKTYCKGDAF MHRRGWRGAYCA MHRRGWRGAYCA HIGCOWRGAYCA LHGCOWRGAYCT LHGCOWRGAYCT LHGCOWRGAYCT LHGCOWRGAYCT LHGCOWRGAYCT LHGCOWRGAYCT LHGCOWRGAYCT LHGCOWRGAYCT LHGCOWRGAYCT LHGCOWRGAYCT LHGCOWRGAYCT LHRCGWRGAYCT	ISSILPAAL ISSILPAAL ISSILPAAS VUSIPRASH VUSIPRASH VUSIPRASH TAPDARGI TAPDARGI TAPDARGI MERDARGI MERDARGI MERDARGI MERDARGI MEPAARGI MEPAARGI MEPAARGI MEPAARGI MEPAARGI MEPAARGI	QERS QERS	IMS PPALEESVMAD LAHVMTCAY DGTE IMS PPALEESVMAD LAHVMTCAY DGTE AAEGIVADEATIVBAVNVTAAAF BKKTG AAEGIVADEATIVBAVNVTAAAF BKKTG AEGIVADEATIVBAVNVTAAAF BKKTG WIRWAGSLEIFFSRNNALLAGARLHP VIRWAGSLEIFFSRNNALLAGARLHP VIRWAGSLEIFFSRNNALLAGARLHP IVRWSGSLEMFFSHISPLFAGRRLSL IVRWSGSLEMFFSHISPLFAGRRLSL IVRWSGSLEMFFSLNNPLIGGRRIQS IVRWSGSLEVFFSLNNPLIGGRRIQS IVRWSGSLEVFFSHSNALLASRRLHP VIRWSGSLEVFFSHSNALLASRRLHP VIRWSGSLEVFFSHSNALLASRRLHP VIRWSGSLEVFFSHSNALLASRRLHP VIRWSGSLEMFFSHSNALLASRRLHP VIRWSGSLEMFFSHSNALLASRRLHP VIRWSGSLEMFFSHSNALLASRRLHP VIRWSGSLEMFFSHSNALLASRRLHP VIRWSGSLEMFFSHSNALLASRRLHP VIRWSGSLEMFFSHSNALLASRRLHP VIRWSGSLEMFFSHSNALLASRRLHP VIRWSGSLEMFFSHSNALLASRRLHP VIRWSGSLEMFFSHSNALLASRRLHP	3777 573 572 1711 614 582 580 581 645 644 637 617 621 621 621 621 626 626 626 626 626 626
TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGHIYGTVTID ATGFC MGRDIGHIYGTVTID ATGFC MGRGIGYIYMMATID VTGFR MGRGIGYIYMMATID VTGFR MGRGIGYIYMATID VTGFR MGRGKGVYYDIATID VTGFR MGRDGWYYNIATID VTGFR MGRDAGWYNIATID VTGFR MGRDAGWYNIATID VTGFR MGRDGWYNIATID VTGFR MGRDGWYNIATID VTGFR MGRDGWYNIATID VTGFR MGRDGWYNIATID VTGFR MGRDGWYNIATID VTGFR MGRDGWYNIATID VTGFR MGRDVGWYNIATID VTGFR	STKFCSBASE STKFCSBASE LPKKTYCKSDAE LPKKTYCKSDAE LPKKTYCKSDAE LPKKTYCKSDAE MHRROWSBAYCA MHRROWSBAYCA HRCOWSWATT IHCOWRSWYCT IHCOWRSWYCT IHCOWRSWYCT IHCOWRSWYCSI IHRCOWSWYCSI IHRCOWSWYCSI IHRCOWRSWYCSI IHRCOWRSWYCSI IHRCOWRSWYCSI IHRCOWRSWYCSI IHRCOWRSWYCSI	ISSILPAAL ISSILPAAL ISSILPAA VUSIPRASH VUSIPRASH VUSIPRASH TAPDABRGT TAPDABRGT MEREARGT MEREARGT MEREARGT MEPAABRGT MEPAABRGT MEPAABRGT MEPAABRGT MEPAABRGT MEPAABRGT MEPAABRGT	QQERS QQERS	IMS PPALEESVMAD LAHVMTCAY DGTE IMS PPALEESVMAD LAHVMTCAY DGTE AAEGIVADEATIV BAVNVTAAAFEKKTG AEGIVADEATIV BAVNVTAAAFEKKTG WIRWAGSLEIFFSRNNALLAGARLHP VIRWAGSLEIFFSRNNALLAGARLHP VIRWAGSLEIFFSRNNALLAGARLHP VIRWAGSLEMFFSHISPLFAGRRISL IVRWSGGSLEMFFSHISPLFAGRRISL IVRWSGGSLEMFFSLNNPLIGGRRIQA VIRWSGGSLEMFFSLNNPLIGGRRIQA VIRWSGGSLEVFFSHNNALIASRRLHP VIRWSGGSLEVFFSHNNALIASRRLHP VIRWSGGSLEVFFSHNNALIASRRLHP VIRWSGGSLEVFFSHNNALIASRRLHP VIRWSGGSLEVFFSHNNALIASRRLHP VIRWSGGSLEMFFSLNNPLIGGRRIQA	3777 573 572 1711 614 582 580 581 645 644 637 6177 6177 6177 617 617 612 613 620 603 626 603 603 601
TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGHIYGTVTID ATCFC MGRDIGHIYGTVTID ATCFC MGRGIGYIYMMATID VTGFR MGRGIGYIYMMATID VTGFR MGRGGVYIYDIATID VTGFR MGRGGVYIYDIATID VTGFR MGRDAGWYNIATID VTGFR MGRDAGWYNIATID VTGFR MGRDAGWYNIATID VTGFR MGRDAGWYNIATID VTGFR MGRDAGWYNIATID VTGFR MGRDUGWYNIATID VTGFR MGRDVGWYNIATID VTGFR MGRDVGWYNIATID VTGFR MGRDVGWYNIATID VTGFR MGRDVGWYNIATID VTGFR MGRDVGWYNIATID VTGFR MGRDVGWYNIATID VTGFR MGRDVGWYNIATID VTGFR MGRDVGWYNIATID VTGFR	STKFGSGASU STKFGSGASU LPKKTYGKSDAU LPKKTYGKSDAU LPKKTYGKSDAU LPKKTYGKSDAU LPKKTYGKSDAU LPKKTYGKSDAU LPKKTYGKSDAU LPKKTYGKSDAU LPKGORSYGCH LHGQORSYGCH LHGQORSYGCH LHGQORSYGCH LHGQORSYGCH LHGQORSYGCH LHGQORSYGCH LHGQORSYGCH LHGQORSYGCH LHGQORSYGCH LHGQORSYGCH LHGQORSYGCH LHGQORSYGCH LHGQORSYGCH LHGQORSYGCH LHGQORSYGCH LHGQORSYGCH LHGQORSYGCH LHGQORSYGCH	ISELLPAL ISELLPAL VUCIPRISH VUCIPRISH VUCIPRISH VUCIPRISH VDEIPRISH VDEIPRISH TAPDARGT MEREARGT MEREARGT MERARGT MERARGT MEPARGT MEPARGT MEPARGT MEPARGT MEPARGT MEPARGT	QCERS QCERS QCERS QCERS IPSPYAAF IPSPYAAF APINLTORLIC	IMS PPALEES VMAD LAHVMTCAY DGTE IMS PPALEES VMAD LAHVMTCAY DGTE AAEGIVADEATIVEAVNVTAAAFEKKTG AAEGIVADEATIVEAVNVTAAAFEKKTG WIRWAGSLEIFFSRNNALLAGARLHP VIRWAGSLEIFFSRNNALLAGARLHP VIRWAGSLEIFFSRNNALLAGARLHP IVRWSGGSLEMFFSHISPLFAGRRLSL IVRWSGGSLEMFFSHISPLFAGRRLSL IVRWSGGSLEMFFSLNNPLIGGRRIQS VIRWSGGSLEVFFSHSNALLASRRLHP VIRWSGGSLEVFFSHSNALLASRRLHP VIRWSGGSLEVFFSHSNALLASRRLHP VIRWSGGSLEVFFSHSNALLASRRLHP VIRWSGGSLEVFFSHSNALLASRRLHP VIRWSGGSLEVFFSHSNALLASRRLHP VIRWSGGSLEVFFSHSNALMAGRRLHP VIRWSGGSLEVFFSHSNALMAGRRLHP VIRWSGGSLEVFFSHSNALMAGRRLHP VIRWSGGSLEVFFSHSNALMAGRRLHP VIRWSGGSLEVFFSHSNALMASRRLHP VIRWSGGSLEVFFSHSNALMASRRLHP VIRWSGGSLEVFFSHSNALMASRRLHP VIRWSGGSLEVFFSHSNALMASRRLHP VIRWSGGSLEVFFSHSNALMASRRLHP VIRWSGGSLEVFFSHSNALMASRRLHP VIRWSGGSLEVFFSHSNALMASRRLHP VIRWSGGSLEVFFSHSNALMASRRLHP VIRWSGGSLEVFFSHSNALMASRRLHP VIRWSGGSLEVFFSHSNALMASRRLHP VIRWSGGSLEVFFSHSNALMASRRLHP VIRWSGGSLEVFFSHSNALMASRRLHP VIRWSGGSLEVFFSHSNALMASRRLHP	3777 573 572 1711 614 582 580 645 645 644 6377 6177 6177 617 617 617 617 617 617 617
TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_1D_TGACV TRIAE_CS42_1D_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGHIYGTVTID ATGGC MGRDIGHIYGTVTID ATGGC MGRDIGHIYGTVTID VTGFR MGRGIGYIYNMATID VTGFR MGRGIGYIYNMATID VTGFR MGRGKGYIYDIATID VTGFR MGRDAGWYNIATID VTGFR MGRDAGWYNIATID VTGFR MGRDAGWYNIATID VTGFR MGRDGWWNIATID VTGFR MGRDVGWYNIATID VTGFR MGRDVGWYNIATID VTGFR MGRDVGWYNIATID VTGFR	STKFCSBASB STKFCSBASB LPKKTYCKSDAF LPKKTYCKSDAF LPKKTYCKSDAF LPKKTYCKSDAF LPKKTYCKSDAF MHRRØNSBATCA MHRRØNSBATCA MHRRØNSBATCA IHCOMRSVY IHCOMRSVY IHCOMRSVY IHROMRSVY	ISELLPARE ISELLPARE ISELLPARE VUCILPRASH VUCILPRASH VUCILPRASH VUCILPRASH TAPDARET TAPDARET TAPDARET MERDARET MERDARET MERDARET MEPAARET MEPAARET MEPAARET MEPAARET MEPAARET MEPAARET MEPAARET MEPAARET MEPAARET MEPAARET	QQERS QQERS	IMS PPALEESVMAD LAHVMTCAY DGTE IMS PPALEESVMAD LAHVMTCAY DGTE AAEGIVADEATIV BAVNVTAAAF BKKTG AAEGIVADEATIV BAVNVTAAAF BKKTG AAEGIVADEATIV BAVNVTAAAF BKKTG WIRWAGSLEIFFSRNNALLAGARLHP VIRWAGSLEIFFSRNNALLAGARLHP VIRWAGSLEIFFSRNNALLAGARLHP IVRWGGSLEMFFSHISPLFAGRRLSL IVRWGGSLEMFFSHISPLFAGRRLSL IVRWGGSLEMFFSHISPLFAGRRLSL IVRWGGSLEMFFSLNNPLIGGRRIQS IVRWGGSLEVFFSHSNALIASRRLHP WIRWGGSLEVFFSHSNALIASRRLHP VIRWGGSLEVFFSHSNALIASRCHP VIRWGGSLEVFFSHSNALIASRCHP VIRWGGSLEVFFSHSNALIASRCHP VIRWGGSLEVFFSHSNALIASRCHP VIRWGGSLEVFFSHSNALIASRCHP VIRWGGSLEVFFSHSNALIASRCHP VIRWGGSLEVFFSHSNALIASRCHP VIRWGGSLEVFFSHSNALIASRCHP VIRWGGSLEVFFSHSNALIASRCHP VIRWGGSLEVFFSHSNALIASRCHP VIRWGGSLEVFFSHSNALIASRCHP VIRWGGSLEVFFSHSNALIASRCHP VIRWGGSLEVFFSHSNALIASRCHP VIRWGGSLF	3777 573572 1711614 582580 5811645 6454 6454 6454 6454 6454 6177621 6177621 6178620 6133620 6266603 6013626603 6015725 6255625
TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV1 TRIAE_CS42_1BS_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGHIYGTVTID ATGFC MGRDIGHIYGTVTID ATGFC MGRGIGYIYMMATID VTGFR MGRGIGYIYMMATID VTGFR MGRGIGYIYMATID VTGFR MGRQGYIYDIATID VTGFR MGRDAGWYNIATID VTGFR MGRDAGWYNIATID VTGFR MGRDAGWYNIATID VTGFR MGRDGWWNIATID VTGFR MGRDGWWNIATID VTGFR MGRDVGWYNIATID VTGFR	STKFCSBASE STKFCSBASE LPKKTYCKSDAE LPKKTYCKSDAE LPKKTYCKSDAE LPKKTYCKSDAE LPKKTYCKSDAE MHRRONRSAYCA MHRRONRSAYCA HRCONRSAYCA LHCONRSAYCTI LHCONRSAYCTI LHCONRSAYCTI LHCONRSAYCTI LHCONRSAYCTI LHCONRSAYCSI MHRCONRSAYCSI LHRCONRSAYCSI LHRCONRSAYCSI LHRCONRSAYCSI LHRCONRSAYCSI LHRCONRSAYCSI LHRCONRSAYCSI LHRCONRSAYCSI LHRCONRSAYCSI LHRCONRSAYCSI	ISSILPAAL ISSILPAAL VUSIPRASH VUSIPRASH VUSIPRASH TAPDARGT TAPDARGT TAPDARGT MEREARGT MEREARGT MEREARGT MERAARGT MEPAARGT MEPAARGT MEPAARGT MEPAARGT MEPAARGT MEPAARGT MEPAARGT MEPAARGT MEPAARGT MEPAARGT MEPAARGT MEPAARGT MEPAARGT MEPAARGT MEPAARGT	QQERSI QQERSI QQERSAA IPSPYAA IPSPYAA APINITORLI	IMS PPALEES VMAD LAHVMTCAY DGTE IMS PPALEES VMAD LAHVMTCAY DGTE AAEGIVADEATIV BAVNVTAAAF KKTG AAEGIVADEATIV BAVNVTAAAF KKTG AAEGIVADEATIV BAVNVTAAAF KKTG WIRWA AGSLEIFFSRNNALLAGARLHP VIRWA AGSLEIFFSRNNALLAGARLHP VIRWA AGSLEIFFSRNNALLAGARLHP VIRWA GGSLEMFFSHISPLFAGRRLSL IVRWS GGSLEMFFSHISPLFAGRRLSL IVRWS GGSLEMFFSLNNPLIGGRRIQA IVRWS GGSLEMFFSLNNPLIGGRRIQA IVRWS GGSLEVFFSHSNALIASRRLHP VIRWS GGSLEVFFSHSNALASRRLHP VIRWS GGSLEVFFSHSNALIASRRLHP VIRWS GGSLEVFFSHSNALIASRCHP VIRWS GGSLEVFFSHSNALIASRCHP VIRWS GGSLEVFFSHSNALIASRCHP VIRWS GGSLEVFFSHSNALIASRCHP VIRWS GGSLEVFFSHSNALIASRCHP VIRWS GGSLENFFSHSNALAGARLHP VIRWS GGSLENFFSHSNALAGARLHP VIRWS GGSLENFFSHSNALAGARLHP	3777 573 572 1711 614 582 580 581 645 645 645 637 617 621 617 621 617 617 621 613 626 603 601 572 2625 624
TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_5BL_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_1B_TGACV TRIAE_CS42_1B_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_1BS_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGMIYGTVTID ATGGC MGRDIGMIYGTVTID ATGGC MGRDIGMIYGTVTID ATGGC MGRGIGYIYMATID VTGGR MGRGIGYIYMATID VTGGR MGRGIGYIYMATID VTGGR MGRGGVGYIYDIATID VTGGR MGRDAGWYNIATID VTGGR MGRDAGWYNIATID VTGGR MGRDAGWYNIATID VTGGR MGRDAGWYNIATID VTGGR MGRDAGWYNIATID VTGGR MGRDVGWYNIATID VTGGR	STKFESBASE 	ISSILPAAL ISSILPAAL ISSILPAAL VUSIPRASH VUSIPRASH VUSIPRASH TAPDARGI TAPDARGI MEREARGI MEREARGI MEREARGI MEREARGI MEREARGI MEPAARGI MEPAARGI MEPAARGI MEPAARGI MEPAARGI MEPAARGI MEPAARGI MEPAARGI MEPAARGI MEPAARGI MEPAARGI MEPAARGI MEPAARGI	QCERS QCERS QCERS	IMS PPALEES VMAD LAHVMTCAY DGTE IMS PPALEES VMAD LAHVMTCAY DGTE AAEGIVADEATIV PAVNYTAAAF KKTG AAEGIVADEATIV PAVNYTAAAF KKTG AAEGIVADEATIV PAVNYTAAAF KKTG VIRWA AGSLEIFFSRNALLAGARLHP VIRWA AGSLEIFFSRNALLAGARLHP VIRWA AGSLEIFFSRNALLAGARLHP IVRWS GGSLEMFFSHISPLFAGRRLSL IVRWS GGSLEMFFSHISPLFAGRRLSL IVRWS GGSLEMFFSHISPLFAGRRLSL IVRWS GGSLEMFFSLNNPLIGGRRIQS VIRWS GGSLEVFFSHNALLASRRLHP VIRWS GGSLEVFFSHNALIASRRLHP VIRWS GGSLEVFFSHNALIASRRLHP VIRWS GGSLEVFFSHSNALIASRRLHP VIRWS GGSLEWFFSHSNALIASRRLHP VIRWS GGSLEWFFSHSNALIASRRLHP VIRWS GGSLEWFFSHSNALIASRRLHP VIRWS GGSLEWFFSHSNALIASRRLHP VIRWS GGSLEWFFSHSNALIASRRLHP VIRWS GGSLEWFFSHSNALIASRRLHP VIRWS GGSLEWFFSHSNALIASRRLHP VIRWS GGSLEWFFSHSNALIASRRLHP VIRWS GGSLEWFFSHSNALIASRRLHP VIRWS GGSLEWFFSHSNALASRRLHP VIRWS GGSLEWFFSHSNALASRRLHP VIRWS GGSLEWFFSHSNALASRRLHP VIRWS GGSLEWFFSHSNALASRLASRRLHP VIRWS GGSLEWFFSHSNALASRCHP VIRWS GGSLEWFFSHSNALASRLASRRLHP VIRWS GGSLEWFFSHSNALASRCHP VIRWS GGSLEWFSHSNALASRCHP VIRWS GGSLEWFFSHSNALASRCHP VIRWS GGSLEWFSHSNALASRCHP VIRWS GGSL	3777 573 572 1711 614 582 580 644 637 617 621 644 637 626 626 603 601 572 625 625 625 624 631
TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_1DL_TGACV_ TRIAE_CS42_1DL_TGACV_ TRIAE_CS42_1DL_TGACV_ TRIAE_CS42_1DL_TGACV_ TRIAE_CS42_1DL_TGACV_ TRIAE_CS42_1DL_TGACV_ TRIAE_CS42_1DL_TGACV_ TRIAE_CS42_1DL_TGACV_ TRIAE_CS42_1DL_TGACV_ TRIAE_CS42_1DL_TGACV_ TRIAE_CS42_1DL_TGACV_ TRIAE_CS42_1DL_TGACV_ TRIAE_CS42_1DL_TGACV_ TRIAE_CS42_1DL_TGACV_ TRIAE_CS42_1DL_TGACV_ TRIAE_CS42_1DL_TGACV_ TRIAE_CS42_1DL_TGACV_ TRIAE_CS42_1DL_TGACV_ TRIAE_CS42_2AS_TGACV_ TRIAE_CS42_2AS_TGACV_ TRIAE_CS42_2AS_TGACV_ TRIAE_CS42_1DL_TGACV_ TRIAE_CS42_1DL_TGACV_ TRIAE_CS42_1DL_TGACV_ TRIAE_CS42_2AS_TGACV_ TRI	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGHIYGTVTID ATGC MGRDIGHIYGTVTID ATGC MGRDIGHIYGTVTID VTGFR MGRGIGYIYMATID VTGFR MGRGGGYIYMATID VTGFR MGRGGGYIYDIATID VTGFR MGRDGGVYIYDIATID VTGFR MGRDGGWYNIATID VTGFR MGRDGGWYNIATID VTGFR MGRDGGWYNIATID VTGFR MGRDGWYNIATID VTGFR MGRDWGWYNIATID VTGFR MGRDWGWYNIATID VTGFR MGRDWGWYNIATID VTGFR MGRDWGWYNIATID VTGFR MGRDWGWYNIATID VTGFR MGRDWGWYNIATID VTGFR MGRDWGWYNIATID VTGFR	STKFGSASB 	ISELLPARE ISELLP	QQERS QQERS	IMS PPALEESVMAD LAHVMTCAY DGTE IMS PPALEESVMAD LAHVMTCAY DGTE AAEGIVADEATIV BAVNVTAAAF BKKTG AAEGIVADEATIV BAVNVTAAAF BKKTG AAEGIVADEATIV BAVNVTAAAF BKKTG VIRWA AGSLEIFFSRNNALLAGARLHP VIRWA AGSLEIFFSRNNALLAGARLHP VIRWA AGSLEIFFSRNNALLAGARLHP VIRWA GGSLEMFFSHISPLFAGRRLSL IVRWS GGSLEMFFSHISPLFAGRRLSL IVRWS GGSLEMFFSLINPLIGGRRIQS IVRWS GGSLEMFFSLNNPLIGGRRIQS IVRWS GGSLEVFFSHNNALLASRRLHP VIRWS GGSLEVFFSHNNALIASRRLHP VIRWS GGSLEVFFSHNNALIASRRLHP VIRWS GGSLEVFFSHNNALIASRRLHP VIRWS GGSLEVFFSHNNALIASRRLHP VIRWS GGSLEVFFSHNNALIASRRLHP VIRWS GGSLEVFFSHNNALIASRRLHP VIRWS GGSLEVFFSHNNALIASRRLHP VIRWS GGSLEVFFSHNNALIASRRLHP VIRWS GGSLEVFFSHNNALIASRRLHP VIRWS GGSLEVFFSHSNALIASRRLHP VIRWS GGSLEVFFSHSNALIASRCHP VIRWS GGSLEVFFSHSNALIASRCHP VIRWS GGSLEVFFSHSNALIASRCHP VIRWS GGSLEVFFSHSNALIASRCHP VIRWS GGSLEVFSHSNALNASRCHP VIRWS GGSLEVFSHSNALNASRCHP VIRWS GGSLEVFSHSNALNASRCHP VIRWS GGSLEVFSHSNALNASRCHP VIRWS GGSLEVFSHSNALNASRCHP VIRWS GGSLEVFSHSNALNASRCHP VIRWS GGSLEVFSHSNALNASRLHP VIRWS GGSLEVFSHSNALNASRCHP VIRWS GGSLEVFSHF	3777 573 572 572 572 572 572 572 580 645 645 645 644 637 617 617 617 617 617 613 626 603 601 572 572 625 625 625 625 625 625 625 625 625 62
TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV TRIAE_CS42_2AL_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGHIYGTVTID ATGFC MGRDIGHIYGTVTID ATGFC MGRGIGYIYMMATID VTGFR MGRGIGYIYMMATID VTGFR MGRGKGYIYDIATID VTGFR MGRDAGWYNIATID VTGFR MGRDAGWYNIATID VTGFR MGRDAGWYNIATID VTGFR MGRDAGWYNIATID VTGFR MGRDGWYNIATID VTGFR MGRDVGWYNIATID VTGFR MGDVGWYNIATID VTGFR	STKFGSGAS STKFGSGAS LPKKTYGKGDA LPKKTYGKGDA LPKKTYGKGDA LPKKTYGKGDA LPKKTYGKGDA LPKKTYGKGDA MHRRGNRSAYCA MHRGNRSAYCA HRGONRSAYAT HGCONRSAYT HGCONRSAYT HGCONRSAYT HRGONRSAYCS HRGONRSAYCS HRGONRSAYCS HRGONRSAYCS HRGONRSAYCS HRGONRSAYCS HRGONRSAYCS HRGONRSAYCS HRGONRSAYCS HRGONRSAYCS HRGONRSAYCS HRGONRSAYCS HRGONRSAYCS HRGONRSAYCS HRGONRSAYCS HRCONRS	ISELPHAE ISELPHAE ISELPHAE ISELPHAE VUCIPRASH VUCIPRASH VUCIPRASH TAPDAERGT MEREARGT MEREARGT MEREARGT MEREARGT MEPAARGT MEPAARGT MEPAARGT MEPAARGT MEPAARGT MEPAARGT MEPAARGT MEPAARGT MEPAARGT MEPAARGT MEPAARGT MEPAARGT	QCERS QCERS QCERS	IMS PPALEES VMAD LAHVMTCAY DGTE IMS PPALEES VMAD LAHVMTCAY DGTE AAEGIVADEATIV BAVNYTAAAF KKTG AAEGIVADEATIV BAVNYTAAAF KKTG MAEGIVADEATIV BAVNYTAAAF KKTG WIRWA AGSLEIFFSRNNALLAGARLHP VIRWA AGSLEIFFSRNNALLAGRRLHP VIRWA GGSLEMFFSHISPLFAGRRLSL IVRWS GGSLEMFFSHISPLFAGRRLSL IVRWS GGSLEMFFSHISPLFAGRRLSL IVRWS GGSLEMFFSLNNPLIGGRRIQA IVRWS GGSLEMFFSLNNPLIGGRRIQA IVRWS GGSLEVFFSHSNALIASRRLHP VIRWS GGSLEVFFSHSNALASRRLHP VIRWS GGSLEVFFSHSNALIASRRLHP VIRWS GGSLEVFFSHSRCPLLAGRRLHP VIRWS GGSLEVFFSHSRCPLLAGRRLHP VIRWS GGSLEVFFSRFCPLLAGRRLHP	3777 573 572 572 572 582 581 645 645 645 645 645 621 621 621 621 621 626 603 626 603 626 626 625 625 625 625 625 631 630
TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGHIYGTYTD ATCEC MGRDIGHIYGTYTD ATCEC MGRDIGHIYGTYTD VTCER MGRGIGYIYNMATD VTCER MGRGIGYIYNMATD VTCER MGRGGYTYDIATD VTCER MGRGKGYTYDIATD VTCER MGRDAGWYNIATD VTCER MGRDAGWYNIATD VTCER MGRDAGWYNIATD VTCER MGRDAGWYNIATD VTCER MGRDAGWYNIATD VTCER MGRDGWYNIATD VTCER MGRDGWYNIATD VTCER MGRDGWYNIATD VTCER MGRDWGWYNIATD VTCER MGRDWGWYNIATDD VTCER	STKFGSASB 	ISELLPARE ISELLPARE VUCILERASH VUCILERASH VUCILERASH VUCILERASH VUCILERASH TAPDARCT TAPDARCT MEREARCT MEREARCT MEREARCT MEREARCT MERASTRCT MEPASTRCT MEPASTRCT MEPASTRCT MEPASTRCT MEPASTRCT MEPASTRCT MEPASTRCT MEPASTRCT MEPASTRCT MEPASTRCT MEPASTRCT MEPASTRCT MEPASTRCT MEPASTRCT	QQERSD QQERSD QQERSAA HPSPYAA HPSPYAA APINLTORLI	IMS PPALEESVMAD LAHVMTCAY DGTE IMS PPALEESVMAD LAHVMTCAY DGTE AAEGIVADEATIV BAVNVTAAAF BKKTG AAEGIVADEATIV BAVNVTAAAF BKKTG AAEGIVADEATIV BAVNVTAAAF BKKTG WIRWA AGSLEIFFSRNNALLAGARLHP VIRWA AGSLEIFFSRNNALLAGARLHP VIRWA AGSLEIFFSRNNALLAGARLHP VIRWA GGSLEMFFSHISPLFAGRRLSL IVRWS GGSLEMFFSHISPLFAGRRLSL IVRWS GGSLEMFFSLNNPLIGGRRIQS VIRWS GGSLEMFFSLNNPLIGGRRIQS VIRWS GGSLEVFFSHSNALLASRRLHP VIRWS GGSLEVFFSHSNALIASRRLHP VIRWS GGSLEVFFSHSPCPLLAGRRLHP IIRWS GGSLEVFFSHSRCPLLAGRRLHP IIRWS GGSLEVFFSHSRCPLLAGRRLHP IIRWS GGSLEVFFSHSRCPLLAGRRLHP VIRWS GGSLEVFFSHSRCPLLAGRRLHP VIRWS GGSLEVFFSHSRCPLAGRRLHP VIRWS GGSLEVFFSHSRCPLAGRRLHP VIRWS GGSLEVFFSHSRCPLAGRRLHP VIRWS GGSLEVFFSHSRCPLAGRRLHP VIRWS GGSLEVFFSHSRCPLAGRRLHP VIRWS GGSLEVFFSHSRCPLAGRRLHP VIRWS GGSLEVFFSHSRCPLAGRRLHP VIRWS GGSLEVFFSRCPLAGRRLHP VIRWS GGSLEVFFSRCPLAG	3777 573 572 1711 614 582 580 645 645 645 645 647 617 621 617 621 617 621 617 622 613 626 603 626 603 572 625 624 630 630 632 632 632 632
TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5BL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_1B_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2BS_TGACV	MTMAKAKAA PV PAKGKHGFLP MTMAKAKAA PV PAKGKHGFLP MTMAKAKAA PV PAKGKHGFLP MTMAKAKAA PV PAKGKHGFLP MTMAKAKAA PV PAKGKHGFLP MGRDIGHIYGTVTID ATGEC MGRDIGHIYGTVTID ATGEC MGRDIGHIYGTVTID ATGEC MGRGIGYIYMMATID VTGER MGRGIGYIYMMATID VTGER MGRGGVYIYDIATID VTGER MGRDIGWYYNIATID VTGER MGRDIGWYNIATID VTGER MGRDIGWYNATID VTGER MGRDGWYDMATID VTGER MGDGWWYDMATID VTGER MGRDGWWYDMATID VTGER MGRDGWWYDMATID VTGER MGRDGWWYDMATID VTGER	STKFCSBASE STKFCSBASE LPKKTYCKSDAE LPKKTYCKSDAE LPKKTYCKSDAE LPKKTYCKSDAE LPKKTYCKSDAE MHRROWSBAYCA MHRROWSBAYCA IECOWRSYYCA IECOWRSYYCA IECOWRSYYCA IECOWRSYYCA IECOWRSYYCA IEROWRSYYCSI IEROWRSYYCDI IEROWRSYYCDI IEROWRSYYCDI IEROWRSYYCSI	ISSILPAAL ISSILPAAL VUSIPRASH VUSIPRASH VUSIPRASH VUSIPRASH TAPDAPRT TAPDAPRT TAPDAPRT MEREAPRT MEREAPRT MEREAPRT MEPAAPRT MEPAAPRT MEPAAPRT MEPAAPRT MEPAAPRT MEPAAPRT MEPAAPRT MEPAAPRT MEPAAPRT MEPAAPRT MEPAAPRT MEPAAPRT MEPAAPRT MEPAAPRT MEPAAPRT MEPAAPRT MEPAAPRT	QQERS	IMS PPALEESVMAD LAHVMTCAY DGTE IMS PPALEESVMAD LAHVMTCAY DGTE AAEGIVADEATIV BAVNVTAAAF BKKTG AAEGIVADEATIV BAVNVTAAAF BKKTG AEGIVADEATIV BAVNVTAAAF BKKTG WIRWAGSLEIFFSRNNALLAGARLHP VIRWAGSLEIFFSRNNALLAGARLHP VIRWAGSLEIFFSRNNALLAGARLHP VIRWAGSLEMFFSHISPLFAGRRISL VVRWSGGSLEMFFSHISPLFAGRRISL VVRWSGGSLEMFFSLNNPLIGGRRIQA VVRWSGGSLEMFFSLNNPLIGGRRIQA VVRWSGGSLEVFFSNNALLASRRLHP VIRWSGGSLEVFFSNNALIASRRLHP VIRWSGGSLEVFFSNNALIASRRLHP VIRWSGGSLEVFFSNNALIASRRLHP VIRWSGGSLEVFFSNNALIASRRLHP VIRWSGGSLEVFFSNNALIASRRLHP VIRWSGGSLEVFFSNNALIASRRLHP VIRWSGGSLEVFFSNNALIASRRLHP VIRWSGGSLEVFFSNNALIASRRLHP VIRWSGGSLEVFFSNNALIASRRLHP VIRWSGGSLEVFFSNNALIASRRLHP VIRWSGGSLEVFFSNNALIASRRLHP VIRWSGGSLEVFFSNNALIASRRLHP VIRWSGGSLEVFFSNNALIASRRLHP VIRWSGGSLEVFFSNNALIASRRLHP VIRWSGGSLEVFFSNALASRRLHP VIRWSGGSLEVFFSNALIASRRLHP VIRWSGGSLEVFFSRFCPLLAGRRLHP IIRWSGGSLEVFFSRFCPLLAGRRLH	3777 5732 5722 5722 5722 5722 572 572 572 572 57
TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_5BL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_10S_TGACV TRIAE_CS42_10S_TGACV TRIAE_CS42_10S_TGACV TRIAE_CS42_10S_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV	MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MTMAKAKAA PVPAKGKHGFLP MGRDIGHIYGTVID ATGFC MGRDIGHIYGTVID ATGFC MGRGIGYIYNMATID VTGFR MGRGIGYIYNMATID VTGFR MGRGGYYYNIATID VTGFR MGRGGYYYDIATID VTGFR MGRDAGWYNIATID VTGFR MGRDAGWYNIATID VTGFR MGRDAGWYNIATID VTGFR MGRDIGWYNIATID VTGFR MGDGWGWYNIATID VTGFR MGDGWGWYNIATID VTGFR MGDGWGWYNIATID VTGFR MGDGWGWYDMATID VTGFR MGDGWGWYDMATID VTGFR	STKFCSGASB STKFCSGASB LPKKTYCKSDAF LPKKTYCKSDAF LPKKTYCKSDAF LPKKTYCKSDAF MIRCWRSAYCA MIRCWRSAYCA MIRCWRSAYCA IIGCORSAYATI IIGCORSAYATI IIGCORSAYCTI IIGCO	ISELLPAL ISELLPAL ISELLPAL ISELLPAL VUCIPRASH VUCIPRASH VUCIPRASH TAPDARGI TAPDARGI MEREARGI MEREARGI MEREARGI MEREARGI MEPAARGI	QQERSJ QQERSJ QQERSJ JQESPYAAF IPSPYAAF IPSPYAAF APINLTORLI	IMS PPALEES VMAD LAHVMTCAY DGTE IMS PPALEES VMAD LAHVMTCAY DGTE AAEGIVADEATIV PAVNYTAAAF KKTG AAEGIVADEATIV PAVNYTAAAF KKTG AAEGIVADEATIV PAVNYTAAAF KKTG WIRWA AGSLEIFFSRNNALLAGARLHP VIRWA AGSLEIFFSRNNALLAGARLHP VIRWA GGSLEMFFSNISPLFAGRRLSL IVRWS GGSLEMFFSNISPLFAGRRLSL IVRWS GGSLEMFFSNISPLFAGRRLSL IVRWS GGSLEMFFSLINPLIGGRRIQS VIRWS GGSLEMFFSLNNPLIGGRRIQS VIRWS GGSLEVFFSNNALLASRRLHP VIRWS GGSLEVFFSNNALIASRRLHP VIRWS GGSLEVFFSNSNALIASRRLHP VIRWS GGSLEVFFSNSNALIASRRLHP VIRWS GGSLEVFFSNSNALIASRRLHP VIRWS GGSLEVFFSNSNALIASRRLHP VIRWS GGSLEVFFSNSNALIASRRLHP VIRWS GGSLEVFFSNSNALIASRRLHP VIRWS GGSLEVFFSNSNALIASRRLHP VIRWS GGSLEVFFSNSNALIASRRLHP VIRWS GGSLEVFFSNSPCPLLAGRRLHP VIRWS GGSLEVFFSNCPLLAGRRLHP VIRWS GGSLEVFF	3777 5732 5721 614 5822 5801 645 645 645 645 645 647 617 621 617 621 617 626 613 626 626 601 5722 625 624 630 630 630 632 632 634 632 632 632 632 632 632 633 632 633 633
TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_1AL_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_2DS_TGACV	MTMAKAKAA PV PAKGKHGFLP MTMAKAKAA PV PAKGKHGFLP MTMAKAKAA PV PAKGKHGFLP MTMAKAKAA PV PAKGKHGFLP MTMAKAKAA PV PAKGKHGFLP MGRDLGHIYGTVTID ATGGC MGRDLGHIYGTVTID ATGGC MGRDLGHIYGTVTID ATGGC MGRDLGHIYGTVTID VTGGP MGRGLGYIYNMATID VTGGP MGRGLGYIYNMATID VTGGP MGRGLGYIYNMATID VTGGP MGRDAGWYINIATID VTGGP MGRDAGWYINIATID VTGGP MGRDAGWYINIATID VTGGP MGRDLGWVINIATID VTGGP MGRDLGWVINIA		ISELLPARE ISELLP	QQERS QQERS	IMS PPALEESVMAD LAHVMTCAY DGTE IMS PPALEESVMAD LAHVMTCAY DGTE AAEGIVADEATIV BAVNVTAAAF BKKTG AAEGIVADEATIV BAVNVTAAAF BKKTG IAEGIVADEATIV BAVNVTAAAF BKKTG IVRW AGSLEIFFSRNALLAGARLHP VIRW AGSLEIFFSRNALLAGARLHP VIRW AGSLEIFFSRNALLAGARLHP IVRW GGSLEMFFSHISPLFAGRRLSL IVRW GGSLEMFFSHISPLFAGRRLSL IVRW GGSLEMFFSHISPLFAGRRLSL IVRW GGSLEMFFSNNPLIGGRRIQS IVRW GGSLEMFFSNNALLASRRLHP VIRW GGSLEVFFSNALLASRRLHP VIRW GGSLEVFFSNALLASRRLHP VIRW GGSLEVFFSNALLASRRLHP VIRW GGSLEVFFSNALLASRRLHP IVRW GGSLEVFFSNALLASRRLHP IVRW GGSLEVFFSNALLASRRLHP IVRW GGSLEVFFSNALLASRRLHP IIRW GGSLEVFFSNALLASRRLHP VIRW GGSLEVFFSNALLASRRLHP VIRW GGSLEVFFSNALLASRRLHP VIRW GGSLEVFFSNALLASRRLHP IIRW GGSLEVFFSRACPLLAGRRLHP IIRW GGSLEVFFSRACPLAGRACHP IIRW GGSLEVFFSRACPLAGRACHP IIRW GGSLEVFFSRACPLAGRACHP IIRW GGSLEVFFSRACPLAGRACHP IIRW	3777 5732 1711 614 5820 5811 6454 6454 6457 6617 418 6200 6137 6216 6177 418 6200 613 6266 6033 6266 603 6057 2262 625 625 625 625 625 625 625 625 62
TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV	MTMAKAKAA PV PAKGKHGFLP MTMAKAKAA PV PAKGKHGFLP MTMAKAKAA PV PAKGKHGFLP MTMAKAKAA PV PAKGKHGFLP MTMAKAKAA PV PAKGKHGFLP MGRDIGHIYSTVTID ATGEC MGRDIGHIYSTVTID ATGEC MGRDIGHIYSTVTID ATGEC MGRGIGYIYMMATID VTGER MGRGIGYIYMMATID VTGER MGRGIGYIYMMATID VTGER MGRGKGVYYDIATID VTGER MGRDAGWYNIATID VTGER MGRDAGWYNIATID VTGER MGRDAGWYNIATID VTGER MGRDGWYNIATID VTGER MGRDGWYNIATID VTGER MGRDVGWYNIATID VTGER MGRDVGWYNATID VTGER MGRDVGWYNATID VTGER MGRDVGWYNATID VTGER MGRDVGWYNATID VTGER MGRDVGWYNATID VTGER MGRDVGWYNIATID VTGER MGREVGWYNIATID VTGER MGREVGWYNIATID VTGER MGREVGWYNIATID VTGER	STKFCSBASE STKFCSBASE LPKKTYCKSDAE LPKKTYCKSDAE LPKKTYCKSDAE LPKKTYCKSDAE LPKKTYCKSDAE LPKKTYCKSDAE MHRROWSAYCA HRCOWSAYCA LHCOWRSYCT LHCOWRSYCT LHCOWRSYCT LHCOWRSYCT LHCOWRSYCT LHCOWRSYCSI	ISELLPACE ISELLPACE VUCIPASE VUCIPASE VUCIPASE VUCIPASE VUCIPASE TAPDAERGI TAPDAERGI MEREARGI MEREARGI MEREARGI MEREARGI MEPASERGI MEPASERGI MEPASERGI MEPASERGI MEPASERGI MEPASERGI MEPASERGI MEPASERGI MEPASERGI MEPASERGI MEPASERGI MEPASERGI MEPASERGI MEPASERGI MEPASERGI MEPASERGI	QQERS QQERS	IMS PPALEES VMAD LAHVMTCAY DGTE IMS PPALEES VMAD LAHVMTCAY DGTE AAEGIVADEATIV BAVNVTAAAF KKTG AAEGIVADEATIV BAVNVTAAAF KKTG AEGIVADEATIV BAVNVTAAAF KKTG WLRWA AGSLEIFFSRNNALLAGARLHP VLRWA AGSLEIFFSRNNALLAGARLHP VLRWA AGSLEIFFSRNNALLAGARLHP VLRWA GGSLEMFFSHISPLFAGRRLSL IVRWS GGSLEMFFSHISPLFAGRRLSL IVRWS GGSLEMFFSHISPLFAGRRLSL IVRWS GGSLEMFFSLNNPLIGGRRIQA IVRWS GGSLEMFFSLNNPLIGGRRIQA IVRWS GGSLEVFFSHNNALIASRRLHP VLRWS GGSLEVFFSHNNALIASRRLHP VLRWS GGSLEVFFSHNNALIASRRLHP VLRWS GGSLEVFFSHSNALIASRRLHP VLRWS GGSLEVFFSHSNALIASRRLHP VLRWS GGSLEVFFSHSNALIASRRLHP IVRWS GGSLEVFFSHSNALIASRRLHP IVRWS GGSLEVFFSHSNALIASRRLHP IVRWS GGSLEVFFSHSNALIASRRLHP IVRWS GGSLEVFFSHSNALIASRRLHP IVRWS GGSLEVFFSHSNALIASRRLHP IVRWS GGSLEVFFSHSNALIASRRLHP IIRWS GGSLEVFFSHSNALIASRRLHP IIRWS GGSLEVFFSHSNALIASRRLHP IIRWS GGSLEVFFSHSNALIASRRLHP IIRWS GGSLEVFFSRFCPLLAGRRLHP IIRWS GGSLEVFFSRFCPLLAGRRLHP IIRWS GGSLEVFFSRFCPLLAGRRLHP IIRWS GGSLEVFFSRFCPLLAGRRLHP IIRWS GGSLEMFFSRNCPLLAGRRLHP IIRWS GGSLEMFFSRNCPLLAGRRLHP	3777 5732 5722 1711 614 5820 5811 645 645 645 645 645 6617 6217 6217 6216 613 6266 6603 6011 617 6216 6266 6266 603 6012 6266 633 6026 645 645 645 645 645 645 645 645 645 64
TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5AL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_1BS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV	MTMAKAKAA PV PAKGKHGFLP MTMAKAKAA PV PAKGKHGFLP MTMAKAKAA PV PAKGKHGFLP MTMAKAKAA PV PAKGKHGFLP MTMAKAKAA PV PAKGKHGFLP MGRDIGMIYGTVID ATGEC MGRDIGMIYGTVID ATGEC MGRDIGMIYGTVID ATGEC MGRDIGMIYGTVID VTGER MGRGIGYIYMATID VTGER MGRGIGYIYMATID VTGER MGRDGWYIYIATID VTGER MGRDAGWYNIATID VTGER MGRDAGWYNIATID VTGER MGRDAGWYNIATID VTGER MGRDAGWYNIATID VTGER MGRDGWYNIATID VTGER MGRDVGWYNIATID VTGER MGDGVGWYDMATID VTGER MGDGVGWYNIATID VTGER MGRDVGWYNIATID VTGER MGRDVGWYNIATID VTGER MGRDVGWYNIATID VTGER MGREVGWYNIATID VTGER MGREVGWYNIATID VTGER MGREVGWYNIATID VTGER MGREVGWYNIATID VTGER MGREVGWYNIATID VTGER MGREVGWYNIATID VTGER	STKFGSGASE 	ISELLPHAL ISELLPHAL ISELLPHAL ISELLPHASH VUCILPHASH VUCILPHASH VUCILPHASH VUCILPHASH TAPDARGT TAPDARGT MEREARGT MEREARGT MEREARGT MEREARGT MEPAARGT	QCERS QCERS QCERS QPSPYAAF PSPYAAF PSPYAAF PSPYAAF APINLTDRLIC APINL	IMS PPALEES VMAD LAHVMTCAY DGTE IMS PPALEES VMAD LAHVMTCAY DGTE AAEGIVADEATIV PAVN TAAAF KKTG AAEGIVADEATIV PAVN TAAAF KKTG AAEGIVADEATIV PAVN TAAAF KKTG WIRWA AGSLEIFFSRNNALLAGARLHP VIRWA AGSLEIFFSRNNALLAGARLHP VIRWA GGSLEMFFSHISPLFAGRRLSL IVRWS GGSLEMFFSHISPLFAGRRLSL IVRWS GGSLEMFFSHISPLFAGRRLSL IVRWS GGSLEMFFSLNNPLIGGRRIQS VIRWS GGSLEMFFSLNNPLIGGRRIQS VIRWS GGSLEVFFSHSNALLASRRLHP VIRWS GGSLEVFFSHSPCPLLAGRRLHP VIRWS GGSLEVFFSHSPCPLLAG	3777 5732 57227 1711614 58225800 64556445 64566456666666666666666666666666666666
TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_5DL_TGACV TRIAE_CS42_SAL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_1DS_TGACV TRIAE_CS42_1DS_TGACV TRIAE_CS42_1DS_TGACV TRIAE_CS42_1DS_TGACV TRIAE_CS42_1DS_TGACV TRIAE_CS42_2DS_TGACV	MTMAKAKAA PV PAKGKHGFLP MTMAKAKAA PV PAKGKHGFLP MTMAKAKAA PV PAKGKHGFLP MTMAKAKAA PV PAKGKHGFLP MGRDIGHI YGTVTID ATGCO MGRDIGHI YGTVTID ATGCO MGRDIGHI YGTVTID ATGCO MGRGIGYI YMATID VTGFR MGRGIGYI YMATID VTGFR MGRGIGYI YMATID VTGFR MGRGIGYI YMATID VTGFR MGRDAGWY YNIATID VTGFR MGRDAGWY YNIATID VTGFR MGRDAGWY YNIATID VTGFR MGRDAGWY YNIATID VTGFR MGRDAGWY YNIATID VTGFR MGRDAGWY YNIATID VTGFR MGRDVGWY YNIATID VTGFR MGREVGWY YNIATID VTGFR MGREVG	STKFGSBASB STKFGSBASB LPKKTYGKSDAF LPKKTYGKSDAF LPKKTYGKSDAF LPKKTYGKSDAF LPKKTYGKSDAF LPKKTYGKSDAF MHRROWSATCA' MHRROWSAYCA' HHGGONSMYCTI HGGONSMYCTI HGGONSMYCTI HGGONSMYCTI HGGONSMYCTI HGGONSMYCTI HGGONSMYCTI HGGONSMYCTI HRGONSMYCTI HRGONSMYCSI HRGONSMYCSI HRGONSMYCSI HRGONSMYCSI HRGONSMYCSI HRGONSMYCSI HRGONSMYCSI HRGONSMYCSI HRGONSMYCSI HRGONSMYCSI HRGONSMYCSI HRGONSMYCSI HRGONSMYCSI HRTONSMYCSI HRTONSMYCSI HRTONSMYCSI HRTONSMYCSI HRTONSMYCSI HRTONSMYCSI HRTONSMYCSI HRTONSMYCSI HRTONSMYCSI HRNONSMYCH	ISELLPARE ISELLPARE VUCILPARE VUCILPARE TAPDARE TAPDARE TAPDARE TAPDARE TAPDARE TAPDARE TAPDARE MERARE MERARE MERARE MERARE MEPARE MEPARE TMEPARE TMEPARE TMEPARE TMEPARE TMEPARE TMEPARE TMEPARE TMEPARE TMEPARE TMEPARE TMEPARE TMEPARE TMEPARE TMEPARE TMEPARE TMEPARE TMEPARE	QQERS QQERS QQERS	IMS PPALEESVMAD LAHVMTCAY DGTE IMS PPALEESVMAD LAHVMTCAY DGTE AAEGIVADEATIV BAVNVTAAAF BKKTG AAEGIVADEATIV BAVNVTAAAF BKKTG AEGIVADEATIV BAVNVTAAAF BKKTG WIRWAGSLEIFFSRNALLAGARLHP VIRWAGSLEIFFSRNALLAGARLHP VIRWAGSLEIFFSRNALLAGARLHP VIRWAGSLEMFFSHISPLFAGRRLSL VVRWSGGSLEMFFSHISPLFAGRRLSL VVRWSGGSLEMFFSHISPLFAGRRLSL VVRWSGGSLEMFFSLNNPLIGGRRIQA VVRWSGGSLEMFFSLNNPLIGGRRIQA VVRWSGGSLEVFFSHNALLASRRLHP VIRWSGGSLEVFFSHNALLAGRRLHP VIRWSGGSLEVFFSHNALLAGRRLHP VIRWSGGSLEVFFSHNALLAGRRLHP VIRWSGGSLEVFFSHNALIASRRLHP VIRWSGGSLEVFFSHNALIASRRLHP VIRWSGGSLEVFFSHNALIASRRLHP VIRWSGGSLEVFFSHNALIASRRLHP VIRWSGGSLEVFFSHNALIASRRLHP VIRWSGGSLEVFFSHSNALIASRRLHP VIRWSGGSLEVFFSHSNALIASRRLHP VIRWSGGSLEVFFSHSNALIASRRLHP VIRWSGGSLEVFFSHSNALIASRRLHP VIRWSGGSLEVFFSHSPCPLLAGRRLHP ILRWSGGSLEVFFSRCPLLAGRRLHP ILRWSGGSLEVFFSRCPLLAGRRLHP ILRWSGGSLEVFFSRCPLLAGRRLHP ILRWSGGSLEVFFSRCPLLAGRRLHP ILRWSGGSLEVFFSRCPLLAGRRLHP ILRWSGGSLEMFFSRCPLLAGRCHP ILRWSGGSLEMFFSRCPLAGRCPLAGRCHP ILRWSGGSLEMFFSRCPLAGRCPLAGRCH	3777 5732 57121 614 5820580 5811645 645465644 6177621 6176617 612166060 6176617 6266603 62626502 622502 622502 622502 622502 62525020 6252502 62525020 62525020 62525020 62525
TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7DL_TGACV TRIAE_CS42_7BL_TGACV TRIAE_CS42_5BL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2DL_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_7AL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_1DL_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2DS_TGACV	MTMAKAKAA PV PAKGKHGFLP MTMAKAKAA PV PAKGKHGFLP MTMAKAKAA PV PAKGKHGFLP MTMAKAKAA PV PAKGKHGFLP MTMAKAKAA PV PAKGKHGFLP MGRDIGHIYSTVTD ATGEC MGRDIGHIYSTVTD ATGEC MGRDIGHIYSTVTD VTGER MGRGIGYIYMATID VTGER MGRGIGYIYMATID VTGER MGRGGYYYDIATID VTGER MGRDAGWYYDIATID VTGER MGRDAGWYNIATID VTGER MGRDAGWYNIATID VTGER MGRDAGWYNIATID VTGER MGRDGWYNIATID VTGER MGRDGWYNIATID VTGER MGRDGWYNIATID VTGER MGRDGWYNIATID VTGER MGRDGWYNIATID VTGER MGRDGWYNIATID VTGER MGRDGWYNIATID VTGER MGRDGWYNIATID VTGER MGRDVGWYNIATID VTGER MGRDGWWYNIATID VTGER MGRDVGWYNIATID VTGER MGRDGWWYNIATID VTGER MGREVGWYNIATID VTGER MGREVGWYNIATID VTGER MGREVGWYNIATID VTGER MGREVGWYNIATID VTGER MGREVGWYNIATID VTGER MGREVGWYNIATID VTGER MGREVGWYNIATID VTGER MGREVGWYNIATID VTGER MGREIGWYDTYTD VTGER MGREIGWYDTYTD VTGER MGREIGWYDTYTD VTGER	STKFCSBASE STKFCSBASE LPKKTYCKSDAE LPKKTYCKSDAE LPKKTYCKSDAE LPKKTYCKSDAE LPKKTYCKSDAE LPKKTYCKSDAE LPKKTYCKSDAE LPKTYCKSDAE MHRCORSAYCA LHCORSAYCT	ISELLPACE ISELLPACE VUSIPRASH VUSIPRASH VUSIPRASH VUSIPRASH TAPDAERGI TAPDAERGI MEREARGI MEREARGI MEREARGI MEREARGI MEREARGI MEPAARGI	QUERS QUERS QUERS	IMS PPALEES VMAD LAHVMTCAY DGTE IMS PPALEES VMAD LAHVMTCAY DGTE AAEGIVADEATIV BAVNVTAAAF KKTG AAEGIVADEATIV BAVNVTAAAF KKTG AAEGIVADEATIV BAVNVTAAAF KKTG WIRWAGSLEIFFSRNNALLAGARLHP VIRWAGSLEIFFSRNNALLAGARLHP VIRWAGSLEIFFSRNNALLAGARLHP VIRWAGSLEMFFSHISPLFAGRRLSL IVRWSGGSLEMFFSHISPLFAGRRLSL IVRWSGGSLEMFFSLNNPLIGGRRIQA IVRWSGGSLEMFFSLNNPLIGGRRIQA IVRWSGGSLEVFFSHSNALIASRRLHP VIRWSGGSLEVFFSHSNCPLLAGRRLHP VIRWSGGSLEVFFSHSNCPLLAGRRLHP VIRWSTGSLEFFSKNNPLFGSTYLHP VIRWSTGSLEFFSKNNPLFGSTYLHP VIRWSTGSLEFFSKNNPLFGSTYLHP VIRWSTGSLEFFSKNNPLFGSTYLHP VIRWSTGSLEFFSKNNPLFGSTYLHP VIRWSTGSLEFFSKNNPLFGSTYLHP VIRWSTGSLEFFSKNNPLFGSTYLHP VIRWSTGSLEFFSKNNPLFGSTYLHP VIRWSTGSLEFFSKNNPLFGSTYLHP VIRWSTGSLEFFSKNNPLFGSTYLHP VIRWSTGSLEFFSKNNPLFGSTYLHP VIRWSTGSLEFFSKNNPLFGSTYLHP VIRWSTGSLEFFSKNNPLFGSTYLHP VIRWSTGSLEFFSKNNPLFGSTYLHP VIRWSTGSLEFFSKNNPLFGSTYLHP VIRWSTGSLEFFSKNNPLF	3777 5732 5722 1711 614 5822 580 6454 645 6445 6377 6217 6217 6217 6217 6216 617 6216 617 6216 613 6266 6266 6266 6257 6255 6255 631 632 632 632 632 632 632 632 632 632 632

 TRIAE_CS42_5DL_TGACv
 LQRLAYLNTTVYPFTSIFLLLYCLLPAIPLVTRSASASAFSVTMPPSGTYMGFVAALMLTLAMVAVLEVRWSGITLGEWW
 662

 TRIAE_CS42_5AL_TGACv
 LQRLAYLNTTVYPFTSIFLLLYCLLPAIPLVTRNASTSAFSVNTPPSATYIAFVAALMLTLAMVAVLEVRWSGITLGEWW
 660

TRIAE_CS42_5BL_TGACv	$\verb"LQRLAYLNTTVYPFTSIFLLLYCLLPAIPLVTRSASTSAFSVNTPPSATYIGFVAALMLTLAMVAALEVRWSGITLGEWW"$	661
TRIAE_CS42_2AS_TGACv	VQRLSYINFTIYPLTSLFILMYAFCPVMWLLPTEILVQRPYTRYIVYLIIVIAMIHVIGMFEIMWAGITWLDWW	719
TRIAE_CS42_2BS_TGACv	VQRLSYINFTIYPLTSLFILMYAFCPVMWLLPTEILIQRPYTRYIVYLLIVIAMIHVIGMFEIMWAGITWLDWW	718
TRIAE_CS42_2DS_TGACv	VQRLSYINFTIYPLTSLFILMYAFCPVMWLLPTEILVQRPYTRYIVYLIIVIVMIHVIGMFEIMWAGITWLDWW	711
TRIAE_CS42_2AS_TGACv	LQRVSYLNMTVYPVTSFFILLYALSPVMVLIPDEVYIQRPFTKYVVFLLVIILMIHVIGMLEIKWAGVTWLDYW	691
TRIAE_CS42_2DS_TGACV	LQRVSYLNMTVYPVTSLFILLYALSPVMWLIPDEVYLQRPFTXIVVFLLVILMIHVIGHEIKWAGVTWLDIW	691 691
TRIAE_CS42_2BS_IGACV	LQRVSILMMIVIEVISSETLLIALSEVIMULEDEVIIQREFINIVVELLVILMINVIGHLETIMAGVISTLMAGVINLDIW	691
TRIAE CS42 2BL TGACV	LORITYLNMSTYPIYYPIISYNFFYMWLFSEOLYIORPEGTMAYL/UGITAMVHLIGMEEVRWSGTTLLDWF	492
TRIAE CS42 2AL TGACV	LORIAYLNMSTYPIVTVFILSYNFFPVMWLFSEOLYIORPFGTYMAYLVAIIAMVHLIGMFEVRWSGITLLDWF	694
TRIAE CS42 2DL TGACV	LORIAYFNMSTYPIVTVFILAYNFFPVMWLFSEOLYIORPFGTYIAYLVAVIAMMHVIGMFEVKWAGITLLDWC	687
TRIAE CS42 2BS TGACV	LQRIAYLNMSTYPIVTVFILAYNLFPVLWLFSEQFYIQRPFAWGFFTDQARHVLLGMLFNVWILVLL	693
TRIAE CS42 2DS TGACV	LQRVAYLNMSTYPIVTVFILAYNLFPVLWLFSEQFYIQRPFGTYIMYLVAVIAMIHVIGMFEVKWAGITLLDWC	700
TRIAE_CS42_7AL_TGACv	LQRIAYLNMSIYPIATMFILAYSFFPVMWLFSEESYYIQRPFGTFIMYLVAVIAMMHVIGMFEVKWAGITLQDWW	678
TRIAE_CS42_7DL_TGACv	LQRIAYLNMSIYPIATMFILAYSFFPVMWLFSEQSYYIQRPFGTFIMYLVVVIAMMHVIGMFEVKWAGITLQDWW	676
TRIAE_CS42_7BL_TGACv		572
TRIAE_CS42_U_TGACVI_	MORVAYINMTTYPVSTFFICMVYLYPVMULFQGEFYIQRPFQFFALFVVVIIATVELIGMVEIRWAGLTLLDWV	699
TRIAE_CS42_IBS_TGACV	MQRIAIINMITIPVSIFFICMIFFIPVMWLFQGEFILQRFFQIFALFVVIVIAIVELIGMVEIRWAGLTPLDWF	090 705
TRIAE CS42 2RS TGACV	MORVATIANTETE DALE VVCTHLEDIMVENGRETICKETETVMIVLITIVSNEVIGAVETVMAGI.TILDWE	704
TRIAE CS42 2DS TGACV	MORVATIANTITI DISALEVUCYHLIPI.MWVIN GATTA GATTA MARATINI MITTI MITTI MARATINI MARAT	702
TRIAE CS42 2BS TGACV	MORIAYANMTAYPVSSVFLVFYLLFPVIWIFRGOFYIOKPFPTYVLYLVIVIALTELIGMVEIKWAGLTLLDWI	531
TRIAE CS42 2DS TGACV	MQRIAYANMTAYPVSSVFLVFYLLFPVIWIFRGQFYIQKPFPTYVLYLVIVIALTELIGMVEIKWAGLTLLDWI	531
TRIAE CS42 2AS TGACV	MQRIAYANMTAYPVSSVFLVFYLLFPVIWIFRGQFYIQKPFPTYVLYLVIVIALTELIGMVEIKWAGLTLLDWI	727
TRIAE_CS42_7AL_TGACv	LQRVAYINITTYPFTAIFLIFYTTVPALSFVTGHFIVQRPTTMFYVYLGIVLSTLLVIAVLEVKWAGVTVFEWF	726
TRIAE_CS42_7DL_TGACv	$\verb"LQRVAYINITTYPFTAIFLIFYTTVPALSFVTGHFIVQRPTTMFYVYLGIVLSTLLVIAVLEVKWAGVTVFEWF"$	325
TRIAE_CS42_7BL_TGACv	LQRVAYINITTYPFTAIFLIFYTTVPALSFVTGHFIVQRPTTMFYVYLGIVLSTLLVIAVLEVKWAGVTVFEWF	768
TRIAE_CS42_5DL_TGACV	RNEQFWMVSATSAYAAAVVQVALKVSAGKEIAFKLTSKQRAS-SPGGGVREFFAELYAVRWTVLMVPTAVVLAVNVMSMA	741
TRIAE_CS42_SAL_TGACV	RNEQFWMVSATSATAATAAVVQVALKVAAGKEIAFKLTSKNKASDSGGVVADRFAELTAVKWTVLMVPTAVVLAVNVTSMA	740
TRIAE_CS42_JBL_IGACV	INFO FEMICISTA V PRAVILIMATINI TERECI HEDVITERO DI ANGARI SAGGA VICENZA VICENZA DA D	79/
TRIAE CS42 2BS TGACV	RNEQFFMIGSVTAY PTAVLHWVNLLTKKGI HERVTKOPVADTDDKYAEMYEVHWVPMVPAVVULESNILAIG	793
TRIAE CS42 2DS TGACV	RNEOFFMIGSVTAYPTAVLHMVVNILTKKGIHFRVTTKOPVADTDDKYAEMYEVHWVPMMIPAVVVLFSNILAIG	786
TRIAE CS42 2AS TGACV	RNEQFFMIGSTSAYPAAVLHMVVNLLTKKGIHFRVTSKQTAADTNDKFADLYDMRWVPMLIPTTVVLIANVGAIG	766
TRIAE CS42 2DS TGACV	RNEQFFMIGSTSAYPAAVLHMVVNLLTKKGIHFRVTSKQTAADTNDKFADLYDMRWVPMLIPTTVVLIANVGAIG	766
TRIAE_CS42_2BS_TGACv	${\tt RNEQFFMIGSTSAYPAAVLHMVVNLLTKKGIHFRVTSKQTAADTNDKFADLYDMRWVPMLIPTTVVLIANVGAIG}$	770
TRIAE_CS42_2DL_TGACv	RNEQFYMIGATGVYPTAVLYMLLKLATGKGIYFRLTSKQTEACSNDKFADLYTVRWVPLLIPTTAVIIVNVAAVG	766
TRIAE_CS42_2BL_TGACv	RNEQFYMIGATGVYPTAVLYMLKLVTGKGIYFRLTSKQTEGCSNDKFADLYTVRWVPLLIPTAAVIIVNVAAIG	567
TRIAE_CS42_ZAL_TGACV	RNEQFYMIGATGVY PTAVLYMLLKLYTGKGI FFRLTSKQTEACSNDKFADLYTVRWVPLLIPTTAVI VNVAAVG	769
TRIAE_CS42_2DL_IGACV	RNEQFILIAAIGVIPIAVLIMALKEVIGGGMAFKLISKQIEACSNAFTTTTANLIVRWYELLIPIAVLIVAAVAAVG	754
TRIAE CS42 2DS TGACV	RNEOFYMIGATGYYPTAVLYMALKLYTGKGIYFRLTSKOTDACSNDKXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	775
TRIAE CS42 7AL TGACV	RNEQFYMITATGVYPTAVLYMALKLIRGKGIYFRLTSKQTEACSGEKFADLYTVRWVPLLIPTVAVLVVNIAAIG	753
TRIAE CS42 7DL TGACV	RNEQFYMIAATGVYPTAVLYMALKLIRGKGIYFRLTSKQTEACSDEKFADLYTVRWVPLLIPTVAVLIVNVTAVG	751
TRIAE_CS42_7BL_TGACv	GSTLYSASLTSTCRSTRSPRCSSPTATLSSPAVGSTLYSASLTSTCRSTRSPRCSS	614
TRIAE_CS42_U_TGACv1_	RNEQFYIIGTTGVYPMAMLHILLRSLGIKGVSFKLTAKKLTGGARERLAELYDVQWVPLLVPTVVVMAVNVAAIG	774
TRIAE_CS42_1BS_TGACv	RNEQFYIIGTTGVYPMAMLHIILRSLGIKGVSFKLTAKKLTSGTRERLAELYDVQWVPLLVPTVVMAVNVAAIG	773
TRIAE_CS42_2AS_TGACV	RNEQFYMICATGVYPTAVLHVVLRSLGLKGMSFKMTAKQLATGARERFAELYNVQWAPLLIPTLVVIAVNVAIG	780
TRIAE_CS42_2BS_TGACV	RNEQFIMICATGVIFIAVLHVVLRSLGLRGLSFRMIARQLAIGAREKFAELIDVQWAPLLIFILVVIAVNVAIG	ררר ררר
TRIAE_CS42_2DS_IGACV	RNEQFINICATOVI FTAVENUVIKSUSUKSUSUKTARQUAIGARAK—— FAELI DVQWAFIDIFI TUVIKVIV VATO	606
TRIAE CS42 2DS TGACV	RNEOFY I I GATAVY PTAVFHI VLKLFGLKGVSFKLTAKOVASSTSDKFAELYAVOWAPMLI PTMVVI AVNVCAIG	606
TRIAE CS42 2AS TGACV	RNEOFYIIGATAVYPTAVFHIVLKLFGLKGVSFKLTAKOVASSTSDKFAELYAVOWAPMLIPTMVVIAVNVCAIG	802
TRIAE CS42 7AL TGACV	RNGQFWMTASCSAYLAAVCQVLTKVIFRRDISFKLTSKLPSGDEKKDPYADLYVVRWTPLMITPIIIIFVNIIGSA	802
TRIAE_CS42_7DL_TGACv	RNGQFWMTASCSAYLAAVCQVLTKVIFRRDISFKLTSKLPSGDEKKDPYADLYVVRWTPLMITPIIIIFVNIIGSA	401
TRIAE_CS42_7BL_TGACv	${\tt RNGQFWMTASCSAYLAAVCQVLTKVIFRRDISFKLTSKLPSGDEKKDPYADLYVVRWTPLMITPIIIIFVNIIGSA}$	844
TRIAE_CS42_5DL_TGACV	AAVQEGRWRKGPAAVLAMAFNAWVVVHLHPFALGLMGRWSKTLSPLLLLVVGFTVLSLCFVLHLHML	808
TRIAE_CS42_SAL_TGACV	AAVQEGRWRRGPAAVLAMAFNAWVVVHLYPFALGLMGRWSRTLSPLLLLVVVFTVLSLCFVLHLHML	807
TRIAE_CS42_SBL_IGACV	35G ING ING ALCH	874
TRIAE CS42 2BS TGACV	VATGKSVLYMGTWSAACKRHGALGLI.FNMWINVIJLYFALATIGKWAKRTGTI.FTILPTAFI.STALMYTGTHTFI.HFFP	873
TRIAE CS42 2DS TGACV	VAIGKSILYMGTWSAAOKRHGALGLLFNLWIMVLLYPFALAIIGRWAKRTGILFILLPIAFLSTSLMYIGVHTFLLHFFP	866
TRIAE CS42 2AS TGACV	VAMGKTIVYMGAWTIAQKTHAALGLLFNVWIMVLLYPFALAIMGRWAKRPVILVVLLPVAFTIVCLVYVAVHILLLSYLT	846
TRIAE_CS42_2DS_TGACv	${\tt VAMGKTIVYMGAWTIAQKTHAALGLLFNVWIMVLLYPFALAIMGRWAKRPVILLVLLPVAFTIVCLVYVAVHILLSYLT}$	846
TRIAE_CS42_2BS_TGACv	$\verb VAMGKTIVYMGAWTIAQKTHAALGLLFNVWIMVLLYPFALAIMGRWAKRPVILLVLLPVAFTIVCLVYVAVHILLLSYLT VAMGKTIVYMGAWTIAQKTHAALGLLFNVWIMVLLYPFALAIMGRWAKRPVILLVLLPVAFTIVCLVYVAVHILLLSYLT VAMGKTAVHILLSYLT VAMGKTAVHIT VAMGKTAVHIT VAMGKTAVHIT VAMGKTAVHIT VAMGKTAVHIT VAMGKTAVHTAVHIT VAMGKTAVHTAVHTAVHTAVHTAVHTAVHTAVHTAVHTAVHTAVH$	850
TRIAE_CS42_2DL_TGACv	AAIGKAATWGFFTDEARHALLGMVFNMGILVLLYPFALGIMGKWAKRPIILFIVLVMAISVVGLLYVSLHAPYTGEWS	844
TKIAE_CS42_2BL_TGACV	AAIGKAATWGFFTDEAKHALLGMVFNMGILVLLYPFALGIMGKWGKKPIILFIVLVMAISVVGLLYVTLHAPYTGEWS	645 017
TRIAL CS42 2AL IGACV	ADIGKADAWGFSTDOARHVLLGMVFNWGTUVUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU	840
TRIAE CS42 2BS TGACV		754
TRIAE CS42 2DS TGACV	SRPSWCSS	783
TRIAE CS42 7AL TGACV	AAIGKAATWGFFTDQAWHAVLGMVFNVGTLVLLYPFALGIMGQWGKRPGILLVMLVMAIGTVGLLYVTLQQDGHRMSF	831
TRIAE_CS42_7DL_TGACv	AAIGKAATWGFFTDQAWHAVLGMVFNVGTLVLLYPFALGIMGQWGKRPGILLVMLVMAIGTVGLLYVTLQQDGHRMSF	829
TRIAE_CS42_7BL_TGACv		614
TRIAE_CS42_U_TGACv1_	AAAGKAIVGRWSAAQVAGAASGLVFNVWMLLLLYPFALGIMGRWSKRPYILFIVLVTAVAATASMYVALAGSLPYLHS	852
TRIAE CS42 1BS TGACV		
TOTAE COAS SNO MONO	AAAGKAIAGRWSAAQVAGAASGLVFNVWMLLLLYPFALGIMGRWSKRPYILFIVLVTAVAATASVYVALAGSLPYLHS	851
TRIAE_CS42_2AS_TGACV	AAAGKAIAGRWSAAQVAGAASGLVFNVWMLLLLYPFALGIMGRWSKRPYILFIVLVTAVAATASVYVALAGSLPYLHS AAVGKAITWGWSAGQVVEAASGLMFNVWILLMFYPFALGVIGRWGKRPYVLFAMFVAAFAAIAAVYVAVQAALAGNLL AAVGKAITWGWSAGQVVEAASGLMFNVWILLMFYPFALGVIGRWGKRPYVLFAMFVAAFAAIAAVYVAVQAAIAGNLL	851 858 857

TRIAE_CS42_2DS_TGACV TRIAE_CS42_2BS_TGACV TRIAE_CS42_2DS_TGACV TRIAE_CS42_2AS_TGACV TRIAE_CS42_2AS_TGACV	VAVGKAITWGWSAGQVVEAJ ASIGKAIVGGWSLMQMADAG ASIGKAIVGGWSLMQMADAG ASIGKAIVGGWSLMQMADAG	ASGLMENVWILLMFYPFALGVIGRWGKRPFVLFAMFVAAFAAIAAVYVAVQAALAGNLP SLGLVFNAWILVLIYPFALGMIGRWSKRPYILFILFVIAFILIALVDIAIQAMRSGFVR SLGLVFNAWILVLIYPFALGMIGRWSKRPYILFILFVIAFILIALVDIAIQAMRSGFVR LGLVFNAWILVLIYPFALGMIGRWSKRPYILFILFVIAFILIALVDIAIQAMRSGFVR	855 684 684 880
TRIAE_CS42_7AL_IGACV	VAFAKVLDGEWTHWLKV	AGGVEENEWULEHLIPEAKGILGKHGKTEVVVLVWWAFTEVITAVLIINTEHMISSGGK	477
TRIAE_CS42_7BL_TGACV	VAFAKVLDGEWTHWLKV	AGGVEENEWULEHLITEAKGILGKHGKTEVVULVWWAFTEVITAVLIINTEHMISSGGK	920
			520
TRIAE CS42 5DL TGACV		808	
TRIAE CS42 5AL TGACV		807	
TRIAE CS42 5BL TGACV	G	815	
TRIAE CS42 2AS TGACV	SMLI	878	
TRIAE CS42 2BS TGACV	SMLI	877	
TRIAE CS42 2DS TGACV	SMLI	870	
TRIAE_CS42_2AS_TGACv	F	847	
TRIAE_CS42_2DS_TGACv	F	847	
TRIAE_CS42_2BS_TGACv	F	851	
TRIAE_CS42_2DL_TGACv	QVAVSLGKASLTGPSGSG	862	
TRIAE_CS42_2BL_TGACv	QVAVSLGKASLTGPSGSG	663	
TRIAE_CS42_2AL_TGACv	QVAVSLGKASLTGPSGSG	865	
TRIAE_CS42_2DL_TGACv	TFLSW	845	
TRIAE_CS42_2BS_TGACv		754	
TRIAE_CS42_2DS_TGACv		783	
TRIAE_CS42_7AL_TGACv	LTRPSG	837	
TRIAE_CS42_7DL_TGACv	LTRPSG	835	
TRIAE_CS42_7BL_TGACv		614	
TRIAE_CS42_U_TGACv1_	GIKLV	857	
TRIAE_CS42_1BS_TGACv	GIKLV	856	
TRIAE_CS42_2AS_TGACv	YFQLGHWSIGGAVSLPSRRV-	878	
TRIAE_CS42_2BS_TGACv	YFQLGHWSIGGAVSLPSRRV-	877	
TRIAE_CS42_2DS_TGACv	YFQLGHRSIGGAVSLASRRV-	875	
TRIAE_CS42_2BS_TGACv	FHFKSSGGATFPTSWGL	701	
TRIAE_CS42_2DS_TGACv	FHFKSSGGATFPTSWGL	701	
TRIAE_CS42_2AS_TGACv	FHFKSSGGATFPTSWGL	897	
TRIAE_CS42_7AL_TGACv	HTTVHGHHGKKFVDAGYYNWP	899	
TRIAE_CS42_7DL_TGACv	HTTVHGHHGKKFVDAGYYNWP	498	
TRIAE_CS42_7BL_TGACv	HTTVHGHHGKKFVDAGYYNWP	941	

Appendix 6.6 List of *CslH* subfamily genes, their protein length (amino acids) and multiple sequence alignment of amino acids showing the conserved motifs (D, D, DXD, QXXRW) specific to *Cellulose synthase like* (*Csl*) family (highlighted with red boxes).

S.No	Gene name with number of splice variants (CslH)	No. of amino acids (aa)	
1	TRIAE CS42 3DS TGACv1 271739 AA0907200.1	714 aa	
2	TRIAE_CS42_3AS_TGACv1_212952_AA0704280.1	331 aa	
3	TRIAE_CS42_3B_TGACv1_222234_AA0760340.1	751 aa	
4	TRIAE_CS42_3B_TGACv1_221049_AA0728260.1	458 aa	
5	TRIAE_CS42_3DS_TGACv1_273502_AA0931770.1	579 aa	
6	TRIAE CS42 2AL TGACv1 094351 AA0296300.3	752 aa	
7	TRIAE CS42 2DL TGACv1 158387 AA0517170.1	752 aa	
8	TRIAE_CS42_2BL_TGACv1_129372_AA0380770.1	799 aa	

TRIAE_CS42_2AL_TGACv TRIAE_CS42_2BL_TGACv TRIAE_CS42_2DL_TGACv TRIAE_CS42_3AS_TGACv TRIAE_CS42_3B_TGACv1	MAGGKKLHERVALGRTAWMLADFVILLLLALV MHRGEDSLSGLYKCTLAFVACGCGWSCGVVLLASLLLVASYLSATAMAGGKKLQERVALGRSAWMLADFVILFLVLALV 	33 80 33 0 30
TRIAE_CS42_3DS_TGACv	MSSAMKLQERVIVPRTAWKLADIFILCLLFALL	33
TRIAE CS42 3B TGACv1	MSSAMKLQERVTVPRTAWKLADIFILCLLLVLL	33
TRIAE_CS42_3DS_TGACv	MGSAMKLQERVILPRTAWKLADIFILCLLFALL	33
TRIAE_CS42_2AL_TGACv	ARRAASLGERGGTWLAALVCEAWFAFVWILNMNGKWSPVRFDTYPENLSHRLEELPAVDMFVTTADPALEPPLITVNT	111
TRIAE CS42 2BL TGACV	ARRAASLGERGGTWLAALVCEAWFAFVWILNMNGKWSPVRFDTYPENLSHRMEELPAVDMFVTTADPALEPPLITVNT	158
TRIAE_CS42_2DL_TGACv	$\label{eq:linear} \texttt{ARRAASLGERGGTWLAALVCEAWFAFVWILNMNGKWSPVRFDTYPDNLSHRMEELPAVDMFVTTADPALEPPLITVNT}$	111
TRIAE CS42 3AS TGACV		0
TRIAE_CS42_3B_TGACv1	${\tt ALLSCRVASLREGGASVAALVCEAWFTFVWIINMNIKWNPVRFNTYPENLSQRTDELPAVDMLVTTADPELEPPLMTVNT}$	110
TRIAE CS42 3DS TGACV	${\tt SCRVLSLGEGGAGAASVAALVCEAWFTFVWILNMNIRWNPVRFHTYPENLSQRMDGLPAVDMLVTTADPELEPPLMTVNT}$	113
TRIAE CS42 3B TGACv1	SCRVASLGEGGAGAAALVCEAWFTFVWILNMNIKWNPVRFHTYPENLSQRMDELPAVDMLVTTADPELEPPLMTVNT	110
TRIAE_CS42_3DS_TGACv	${\tt SCRVASLGDGGAGAASVAALVCEAWFTFVWILNMNIKWNPVRFHTYPENLSQRMDELPAVDMLVTTADPELEPPLMTVNT}$	113
TRIAE_CS42_2AL_TGACv	$\verb+VLSLLALDYPDVGKLACYVSDDGCSPVTCYALREAAKFASLWIPFCKRYDVGVRAPFMYFSSAPEVGTGTADHEFLESWAPFMYFSSAPEVGTGTADHFFLESWAPFMYFSSAPEVGTGTADHFFLESWAPFMYFSSAPEVGTGTADHFFLESWAPFMYFSSAPEVGTGTADHFFLESWAPFMYFSSAPEVGTGTADHFFLESWAPFMYFSSAPEVGTGTADHFFLESWAPFMYFSSAPEVGTGTADHFFLESWAPFMYFSSAPEVGTGTADHFFLESWAPFMYFSSAPEVGTGTADHFFLESWAPFMYFSSAPEVGTGTADHFFLESWAPFMYFSSAPEVGTGTADHFFLESWAPFMYFSSAPEVGTGTADHFFLESWAPFMYFTMYFTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGT$	191

TRIAE_CS42_2BL_TGACv TRIAE_CS42_2DL_TGACv	VLSLLALDYPHVGKLACYVSDDGCSPLTCYSLREAAKFASLWVPFCKRHDVGVRAPFMYFSSAPEVDTGTVDHEFLESWA VLSLLALDYPDVGRLACYVSDDGCSPVTCYALREAAKFAGLWVPFCKRHDVGVRAPFMYFSSAPEVGNGTVDHEFLESWA	238 191
TRIAE CS42 3AS TGACV TRIAE CS42 3B TGACV1	VI.SLLAVDYPDVDKLACYVSDDGCSPVTCYALREAAGFARLWVPFCKRHGVGVRAPFMYFASSRPEPELAGDWTFI	0 186
TRIAE_CS42_3DS_TGACv	VLSLLAMDYPDVDKLACYVSDDGCSPVTCYALHEAARFAGLWVPFCKRHGVGVRAPFMYFASRPEPELAGDNFSDEWT	191
TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3DS_TGACV1	VLSLLAVDYPDVDKLACYVSDDGCSPVTCYALREAAGFARLWVFFCKRHGVGVRAPFIYFASS-RPEPDLAGDKFSDDWI VLSLLAVDYPDVDKLACYVSDDGCSPATCYALREAAWFARLWVFFCKRHDVRVRAPIIYFASRLEPELAGDTFSDEWT	189 191
TRIAE CS42 2AL TGACV	LMKTEYEKLASRIENADEVSILR-DGGEEFAEFIDAERGNHPTIVKVLWDNSKSK-AGEGFPHLVYLSREKSPRHRHNFK	269
TRIAE_CS42_2BL_TGACV	LMKSEYEKLASRIENADEVSILR-DGGDEFAEFIDAERGNHPTIVKVLWDNSKNK-TGEGFPHLVYLSREKSPRHRHNFK	316 269
TRIAE_CS42_3AS_TGACv		0
TRIAE_CS42_3B_TGACV1 TRIAE_CS42_3DS_TGACV1	RSEIDRLVSRIESADEGSLERDDAADFIEFREARRGDHPAIVNVLWDRSRSSRIGSGDGFPNLVIVSRERIRRDHHIN FIKSEYDKLVSRIESADEGSLERDDDAGEFTEFMEAKRGDHPGIVKVLWDNSKSSRIGEGFPNLVYVSREKSRKHDHHIN	266 271
TRIAE_CS42_3B_TGACv1 TRIAE_CS42_3DS_TGACv	FIKSEYDKLVSLIESADEASLLRHDHAGEFTEFKGAECGDHPAIVKVLWDNSKSSGTGEGFPNLVYVSREKSRKHDHHYK FIKSEYDKLVSRIESADEGSLLRHDDAGEFTEFMEAERTDHPAIVKVLWDNSKSSRTGEAFPHLVYVSSEKSRKHHHHYK	269 271
TRIAE CS42 2AL TGACV	AGAMNVLTRVSAVMTNAPIMLNVDCDMFANNPOVALHAMCLLLGFDDEIHSGFVOAPOKFYGGLKDDPFGNOMOVITKKI	349
TRIAE_CS42_2BL_TGACV	AGAMNVLTRVSAVMTNAPIMLNYDCD MFANNPQVALHAMCLLLGFDDEIHSGFVQAPQKFYGGLKDDPFGNQMQVITKKI	396 349
TRIAE_CS42_2DL_IGACV TRIAE_CS42_3AS_TGACV		0
TRIAE_CS42_3B_TGACv1	AGAMNVLARVSAVMTNAPIILNIDCD IFVNNPQVVLHAMCLLLGFNDETCSGFVQVPQRFYAKLKDDPFGNQIEVLREKL	346 351
TRIAE_CS42_3B_TGACv1	AGAMNVLARVSAVMTNAPIILN <mark>V</mark> DCD IFVNNPQVVLHATCLLLGFDDETCSGFVQVPQRFYGKLKDDPFGNQMEVLRS	347
TRIAE_CS42_3DS_TGACv	AGAMNVLARVSAVMTNAPIILN <mark>YDCD</mark> MFVNNSQVVLHAMCLLLGFDDETCSGFVQVPQRFYGKLKDDPFGNQMEVLREKL	351
TRIAE_CS42_2AL_TGACV	GGGLAGIQGTFYGGTGCFHRRKVIYGMPPPDTVKHETRGSPSYKELQAKFGSSKELIESSRNIISGDLLARPTVDISS	427
TRIAE_CS42_2BL_IGACV TRIAE_CS42_2DL_TGACV	GGGLAGIQGIFIGGIGCFHRRKVIIGHFPD-TVKHENKGSPSIKLLQAKFGSSKLLLESSKNIISGDLLAKFIVDIS GGGLAGIQGMFYGGTGCFHRRKVIYGVPPD-TVKHEMKGSPSYKELQAKFGSSKLLESSRNIISGDLLARFIVDIS	427
TRIAE_CS42_3AS_TGACv TRIAE_CS42_3B_TGACv1	MIDISS LGGLSGLOGIYYLGTGCFHRRKIIYGVAPPSFAAVKHEROGSLTYEDLRTKFGASVELAESARNIYSREIPLKPMIDISS	6 42.6
TRIAE_CS42_3DS_TGACv	FGGLAGLQGIYYLGMGCFHRRKIIYGVAPSSSAAIKHEREGSRSYEDLRTKFGASVELVESARNIYSGEIPPSPMIDISS	431
TRIAE_CS42_3B_TGACv1 TRIAE_CS42_3DS_TGACv1	LSYEDLLTKFGASMELVESSRNIYSVEIPPKPMLDITS LGGLSGLQGIFYLGTGCFHRRKIIYGVAPSSFAAVKHEREGSLSYEDLRTKFGASVELVESTRNIYSREIPPKPMVNISS	385 431
TRIAE CS42 2AL TGACV	RWEMAKOWGDCNYEAGECORCORCORCENCEWYGSMTEDILTGORIOAAGWESALLDTDPPAFLGCAPTGGPASL	507
TRIAE_CS42_2BL_TGACv	RVEM <mark>AKQV</mark> GD <mark>CNYE</mark> AGTCNGQEIGWVYGSMTEDILTGQRIQAAGWESALLDTDPPAFLGCAPTGGPASL ^T QFKRWATGLL	554
TRIAE_CS42_2DL_TGACV TRIAE_CS42_3AS_TGACV	RVEMAKQVGDCNYEAGTCMGQEIGWVYGSMTEDILTGLRIHAAGWESALLDTEPPAFLGCAPTGGPASLTQFKRWATGLL RIOVAKOVSSCNYETDHWGOEIGWSYGSMAEDILTGORTHSSGWKSTLLDTNPPAFLGCAPTGGPASLTOYKRWATGLL	507 86
TRIAE_CS42_3B_TGACv1	RIQV <mark>AKQV</mark> SS <mark>CNYE</mark> TGTH <mark>NGQEI</mark> GWSYGSMAEDILTGQRIHSAGWKSTSPDTNPPAFLGCAPTGGPASL ^T QYKRWATGLL	506
TRIAE_CS42_3DS_TGACv TRIAE_CS42_3B_TGACv1	RLQVAKQVSSCNYETDHMGQBLGWSYGSMAEDILTGQRIHSSGWKSTLLDTNPPAFLGCAPTGGPASLTQYKRWATGLL RLOVAKOVSTCNYETGHMGBBASNHG	511 412
TRIAE_CS42_3DS_TGACv	CIQV <mark>AKQV</mark> SS <mark>CNYE</mark> TG <mark>HHNGQBI</mark> GWSYGSMAEDILTGQRIHSAGWKSTLLDTNPPAFLGCAPTGGPASLI <mark>QYKRWA</mark> TGVL	511
TRIAE_CS42_2AL_TGACv	EILISRNSPILGTIFKCLØLROCLGYLIVDAWPVRAPFELCYPLLGPFCLLTNQSFLPTASDEGFHIPAALFLTYNIYHL	587
TRIAE_CS42_2BL_TGACV TRIAE_CS42_2DL_TGACV	EILISRNSPILGTIFRRLQLRQCLAYLIVNAWPMRAPFPMCYALLGPFCLLTNQSFLPTTSNEGFRIPAALFLSYHVYHL EILISONSPILGTIFRRLOLROCLAYLIVEAWPVRAPFELCYALLGPFCLLTNOSFLPTASDEGFRIPAALFLTCHIYHL	634 587
TRIAE_CS42_3AS_TGACV	EILLGQNSPIIATIFKRLQFRQFLAYLVFYVWSMRAPFELCYALLGPFCLFRNQSFLLKASNHGFSIQLALFLSYNIYNF	166
TRIAE_CS42_3B_TGACVI TRIAE_CS42_3DS_TGACVI	EILLGPNIPIIATIERREOFROTEGIUVEIVWSMRAPFELCIELLGEFCLFRNESFLLASNHGFSIQLALFLSINIINF EILLGQNSPIMATVFKRLQFROSLAYLVFYVWSMRAPFELCYALLGPFCLFRNQSFLLKASNHGFSIQLALFLSINIYNF	586 591
TRIAE_CS42_3B_TGACv1	-FSIQLALFUSYNTYNFVEYKECELSARTWWNNMRNINLLAPCFP	458 579
		6.67
TRIAE_CS42_2AL_TGACV TRIAE_CS42_2BL_TGACV	MEYKECGLSVRAWWNNHRMQRITSASAWLLAFLTVILKTLGLSETVFEVTRKESSTSSDGGAGTDDADPGLFTFDSAPVF MEYKECGLSVRAWWNNHRMQRITSASAWLLAFLTVILKTLGLSETVFEVTRKESSTSSDGGTGTDEADTGLFTFDSAPVF	667 714
TRIAE_CS42_2DL_TGACV	MEYKECGLSVRAWWNNHRMQRITSASAWLLAFLTVILKTLGLSETVFEVTRKESSTSSDGGAGTDEADPGLFTFDSAPVF	667 245
TRIAE_CS42_3AS_IGACV TRIAE_CS42_3B_TGACV1	VEIMEGELSARIWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW	665
TRIAE_CS42_3DS_TGACV	VEYMECGLSARTWWNNMRMQRIVSISSWLLDFLSVVLKTIGLSKTVFEVTRKDKST-SDGDPSTHETDLGWFTFDSSPVF	670 458
TRIAE_CS42_3DS_TGACv		579
TRIAE_CS42_2AL_TGACv	IPVTALSVLNIVALTVAAWRAVVGTVAG-VHGGPGVGEFVCCGWMVLCFWPFVRGLVSSGKYGIPWSVRVKAGLIVAAFV	746
TRIAE_CS42_2BL_TGACV	IPVTALSMLNIVALAVAAWRAVVGTAAG-VHGGPGVGEFVCCGWMVLCFWPFMRGLVSSGKYGIPWSVRVKAGLIVAAFV	793 746
TRIAE_CS42_3AS_TGACv	IPVTVAILNIATIAIGVWRHAIFWMITGNHDWQNIGEFICCGWAILYFWPFIKGLVGRGRYGIPWNVKLKAWVIVVAFL	325
TRIAE_CS42_3B_TGACv1 TRIAE_CS42_3DS_TGACv1	IPMTAVAILNIVTIAIGVWRHAIFWMTTGNHDCQNIGEFLCCGLMILYFWPFIKGLVGRGRYGIPWNVKLKAWVIVVAFL	745 714
TRIAE_CS42_3B_TGACv1		458
TRIAE_CS42_3DS_TGACv		579
TRIAE_CS42_2AL_TGACV	HLCTRN 752	
TRIAE_CS42_2DL_TGACV	HICTRN 752	
TRIAE_CS42_3AS_TGACV	YFCRGD 331 YFCRGD 751	
TRIAE_CS42_3DS_TGACv	714	
TRIAE_CS42_3B_TGACv1 TRIAE_CS42_3DS_TGACv	458 579	
Appendix 6.7 List of *CslJ* subfamily genes, their protein length (amino acids) and multiple sequence alignment of amino acids showing the conserved motifs (D, D, DXD, QXXRW) specific to *Cellulose synthase like* (*Csl*) family (highlighted with red boxes).

S.No	Gene name	with nu	mber of	splice	variants	(CslJ)	No.	of	amino	acids	(aa)			
1	TRIAE_C	S42_3DS	_TGACv1	272297	_AA0918580	.1		738	aa					
2	TRIAE_C	S42_3AS	_TGACv1	210908	_AA0681280	.2		766	aa					
3	TRIAE_C	S42_3B_	TGACv1_2	221705_2	AA0747940.	1		734	aa					
4	TRIAE_C	S42_3DS	_TGACv1	_272756	_AA0924850	.1		734	aa					
Color	Align Consei	vation	results											
TRIAE_	CS42_3B_TGAG	Cv1 MAAK	PSQDAPL	QLHTVEV	DQPIATVNRI	LAVLHVAI	LAAAAIA	HRGA	HVMLA	DLVLLE	LWALSQ	APMWRPV	SR <mark>AAFPSRI</mark>	80
TRIAE_	CS42_3DS_TGA	ACV MATK	PSQDAPL	PLHTVQT	DQPLATVNRI	LAAVHLAI	LGAAAIA	HRGA	HVMLA	ADLVLLE	LWALSQ	APMWRPV	SRTAFPSRI	80
TRIAE_	CS42_3AS_TGA	ACV MAAR	PSQDAPL	QLHTVQT Stumvom	DQPLATVNR			HRGA	AHVMLA	DLVLL	LWALSQ	APMWRPV	SRAAFPSRI	80
TRIAL_	CS42_3DS_TGA	ACV MAAE	PSQDAPL	2 lhi vQi	DQPLATVNR	LAALHVA	LAAAAIA	HRGP	AHVMLA		LWALSQ	APMWRPV	SRAAFPSRI	80
TRIAE_	CS42_3B_TGAG	Cv1 SRAA	LPAVDVM	VVTADPD	KEPAAKVMNI	VVSAMAL	NYPGGRI	SVYI	LS DI AGS	SPRTLLA	ARKAYA	FARAWVP	FCRKYGVRC	160
TRIAE_	CS42_3DS_TGA	ACV SRAA	LPAVDVM	VVTADPE.	KEPAAKVMNI	VVSAMAL	DYPGGRI	SVYI	LS DIDAGS	SPRTLLA	ARKAYA	FARAWVP	FCRKYGVRC	160
TRIAE_	CS42_3AS_TGA	ACV SRPA	LPAVDVM	VVTADPD.	KEPAAKVMN'I	VVSAMALI	DYPGGRL	SVYI	LS DIDAGS	SPRTLLA	ARKAYA		FCRKYGVRC	160
TRIAL	CS42_3DS_TGA	ACV SRAA	LPAVDVM	VVTADPD.	KEPAAK VMN 1	VVSAMALI	DIPGGRL	SVII	12 DI AG	SPRILLP	ARKAIA	ARAWVP	FURKIGVRU	100
TRIAE_	CS42_3B_TGAG	Cv1 PCPD	RFFAGDD	QI <mark>DI</mark> DGH	hrq <mark>el</mark> ddri	RIKKMYE	rf <mark>k</mark> egvf	EVMS	DAALS	QS <mark>W</mark> TKAI	hDAHVE	IITGDE-	QDSSNSNSG	239
TRIAE_	CS42_3DS_TGA	ACV PCPD	RFFAGDD	KI <mark>.D.</mark> GSH	hhh <mark>el</mark> addri	RIKNMYE	IFNEGVR	REVMS	BDADLS	2SCTKAI	HDAHVE	IITGDE-	QDSS <mark>N</mark> SNSG	239
TRIAE	CS42_3AS_TGA	ACV PCPD	RFFAGDD	QI DI GDH	HRQELDDDRI	RIKNVYE	IFKEGVE		JDATLS(2SWTKAI	HAHVE	IITDEQG	QDSSHSNSC	240
TRIAL	CS42_3DS_TGA	ACV PCPD	RFFAGDD	2. <mark>D.</mark> GGA	RQLLDDDRI	RIKINMIE.	TEREGVE	' V MIL	DAALS	25WTKAL	HDAHVE		QDSS <mark>NSNSC</mark>	233
TRIAE	CS42_3B_TGAG	Cv1 DGEE	DEDATPL	LVYVSRG	KRRSS <mark>T</mark> HHFF	AGALNVLI	LRVSSL	ISNSI	PYVMVI	DCD 4YCN	ISRSSIL	EAMCFHL	DGRRRADLA	319
TRIAE	CS42_3DS_TG#	ACV DGEE	DEDAMPL	LVYVSRE	KRRSS <mark>T</mark> HHFF	KAGALNVLI	LRVSSLI	SNSI	PYVMVI	DCD 4YCN	ISRSSIL	EAMCFHL	DGRRRADLA	319
TRIAE_	CS42_3AS_TGA	ACV DGDG	DEDAMPL	LVYVSRE	KRRSSTHHFF	AGALNVLI	LRVSSL	SNSI	PYVMVI	DCD 4YCN	SRSSLL	EAMCFHL	DGRRRADLA	320
TRIAE_	CS42_3DS_TGA	ACV DGBE	DEDA <mark>M</mark> PL.	LVYVSRE	KRRSS <mark>A</mark> HHFF	AGALNVLI	LRVSSLM	ISNSI	PYVMVI	DCD 1YCN	SWSSVL	EAMCFHL	DGRRRADLA	313
TRIAE	CS42_3B_TGAG	Cv1 FVQF	PQMFHNL	STSDIYA	NELR <mark>SIFW</mark> T-					RWF	GLDGLR	GPILSGT	GFCARRDAI	371
TRIAE	CS42_3DS_TG#	ACV FVQF	PQMFHNL	SSSDIYA	NELRSIFWT-					RWF	GLDGLR	GPILSGT	GFCARRDAI	371
TRIAE_	CS42_3AS_TGA	ACV FVQF	PQMFHNL	SSSDIYA	NELRPIFWVF	RKKTNRPC	IASVIFS	EFSS	SNLGACI	MVQTRWF	GLDGLR	GPILSGT	GFCVRRDAV	400
TRIAE_	CS42_3DS_TGA	ACV FVQF	PQMFHNL	SSSDIYA	NELLR <mark>SI EW</mark> AG	SPTG				-LRDAVE	RRGRPP	GPILSGT	GFCVRRDAV	372
TRIAE	CS42_3B_TGAG	Cv1 YGAL	PASSQDQ	-FSGVEV	GELKRRFGVS	SNGHIASLE	RRPGTGS	TIV	RDALP	QDAE	LVACCD	YETGTEW	GEEVGFLYÇ	447
TRIAE	CS42_3DS_TGA	ACV YGAR	PASSQDQ	FSGVEV	GELKRRFGVS	SNGHIASLE	RRSGTGS	STIVA	RDALP	2EDAF	LVASCA	YETGTEW	GEQVGFLYÇ	448
TRIAE	CS42_3AS_TGA	ACV YGAG	PGSSQEQ	-FSGVEV	GELKRRFGVS	SNGHIASLE	RRSGTGS	TIV	AGDVLI	PQDAE	LVASCD	YETGTEW	GEDVGFLYC	477
TRIAE_	CS42_3DS_TGA	acv <u>Yga</u> g	PGSSQDH	2SSGVEV	GELKRRFGVS	SNGHIASLE	RRSGTGS	511174	RDG	2PQBDAE	LVASCD	YETGTEW	GEEVGELYÇ	452
TRIAE	CS42 3B TGAG	Cv1 SVVE	DYFTGYR	QLYC <mark>R</mark> GW'	TSVYCFPAT	SRPPFLGS	SVPTNLN	IDAL\	QNKRWI	SGMLAV	GLSRHC	PLAS <mark>A</mark> AA	ISVPESMGE	527
TRIAE	CS42_3DS_TG#	ACV SVVE	DYFTGYR	QLYCRGW'	TSVYCFPA <mark>A</mark> A	ASRPPFLGS	SVPTNLN	IDAL	QNKRWI	SGMLAV	GLSRHC	plas-aa	ICVPQSMGE	527
TRIAE_	CS42_3AS_TGA	ACV SVVE	DYFTGYR	QLYCRGW'	ISVYCFPATO	SRPPFLGS	SVPTNLN	IDAL\	QNKRWI	SGLLAV	GLSRHC	PLASAAA	ISVPQSMGE	557
TRIAE_	CS42_3DS_TGA	ACV SVVE	DYFTGYR	QLYC <mark>P</mark> GW	TSVYCFPAT	TRPPFLGS	SVPTNLN	IDAL\	ONKRWI	SGMLAV	GLSRHC	PLAS <mark>A</mark> AA	VSVPQSMGE	532
TRIAE	CS42_3B_TGAG	Cv1 AYYA	FMALYAF	PVLCYAI	VPQLCFFRG	TSFP-EAS	STLWFAA	VFVS	SSSLQHI	LVEVSVA	KRGLAA	RTCWNEQ	RFWALNAVI	606
TRIAE	CS42_3DS_TGA	ACV AYYA	FMALYAF	PVLCYAT	VPQLCFLRG	TSFPGAA	STLWFAA	VFAS	SSSLQHI	LVEVSVA	KRGLAL	RTWWNEQ	RFWALNAVI	607
TRIAE_	CS42_3AS_TGA	ACV AYYA	FTPLYAF	PLLCYAT	VPQLCFLRG/	TSFPEAAS	STLWFAA	VFAS	SSSLQHI	LVEVSVA	KRGLAA	RTWWNEQ	RFWALNAVI	637
TRIAE_	CS42_3DS_TGA	ACV AYYA	FMALYAF	PVLCYAT	VPQLCFLRGC	TSFP-GE	SALWFAA	VLA:	SSLQH	LVEVSFA	KRGLAA	RAWWNEQ	RFWALNAVI	611
TRIAE_	CS42_3B_TGAG	Cv1 GQLF	ACLSVAL	NLVDGAG	GRAVDFDLTS	SKAS <mark>DDRL</mark>	YRD <mark>GVF</mark> E	FAG	STLLL	PATTLCI	LNAAAL	VGGVWKM	VGRGGNMP-	- 685
TRIAE	CS42_3DS_TGA	ACV GQLF	ACLGVAL	NLVG-AG	GRAVDFDLTS	SKASDDRLY	YRDGVFD	FAG	TTLLI	PATTLCI	LNAAAL	VGGVWKM	VGRGGSVS-	- 685
TRIAE	CS42_3AS_TGA	ACV GQIF	ACLGVAL	SLVG-AG	GRAVDFDLTS	SKASGDRLY	YRDGVFD	FAG	SALLLI	PATTLCI	LNAAAL	VGGVWKM	VGRGGNVSC	5 716
TRIAE_	CS42_3DS_TGA	AUV GQLE	ACVSVAL	SLVG-AG	GRAVDFDLTS	SKASDDRLY	YRDSVFD	FAG	SALLL	PATTLCI	LNIAAL	VGGVWKM	VGRGGSVS-	- 689
TRIAE	CS42_3B_TGAG	Cv1 - <mark>GE</mark> L	FLLCYIA	ALSYPLL	QGMFLRRDL	ARVPARIT <i>I</i>	AMSVAMV	ATLI	LSLFG	734				
TRIAE	CS42_3DS_TGA	ACV - <mark>GEL</mark>	FLLCYVA	ALSYPLL	QGMFLRRDPA	ARVPAPIT/	amsvamv	ALI	LSLFG	734				
TRIAE	CS42_3AS_TGA	ACV TGEL	FLLCYVA	ALSYPLL	JGMFLRRDPA	ARVPARITA	AVSVAIV	ATLI	SLFG	/66 739				
TUTHE	COAS_ODD_IGE		г п п с т v А	THE REPORT	- OPHI LIKKDP		at 10 V At 10	±21	ырть С	,				

IX. REFERENCES

- Achyuthan KE, Achyuthan AM, Adams PD, Dirk SM, Harper JC, Simmons BA, Singh AK. 2010. Supramolecular self-assembled chaos: polyphenolic lignin's barrier to cost-effective lignocellulosic biofuels. *Molecules* 15: 8641-8688.
- Ahmad S, Kaur S, Lamb-Palmer ND, Lefsrud M, Singh J. 2015. Genetic diversity and population structure of Pisum sativum accessions for marker-trait association of lipid content. *The Crop Journal* **3**: 238-245.
- Andersen JR, Lubberstedt T. 2003. Functional markers in plants. Trends Plant Sci 8: 554-560.
- Appenzeller L, Doblin M, Barreiro R, Wang H, Niu X, Kollipara K, Carrigan L, Tomes D, Chapman M, Dhugga KS. 2004. Cellulose synthesis in maize: isolation and expression analysis of the cellulose synthase (CesA) gene family. *Cellulose* 11: 287-299.
- Aramrak A, Kidwell KK, Steber CM, Burke IC. 2015. Molecular and phylogenetic characterization of the homoeologous EPSP Synthase genes of allohexaploid wheat, Triticum aestivum (L.). *BMC genomics* 16: 1.
- Arioli T, Peng L, Betzner AS, Burn J, Wittke W, Herth W, Camilleri C, Höfte H, Plazinski J, Birch
 R. 1998. Molecular analysis of cellulose biosynthesis in Arabidopsis. *Science* 279: 717-720.
- Baenziger PS, Bakhsh A, Lorenz A, Walia H. 2014. Bridging Conventional Breeding and Genomics for A More Sustainable Wheat Production. doi:10.1007/978-94-007-7575-6_7: 185-209.
- Baldwin L, Głazowska S, Mravec J, Fangel J, Zhang H, Felby C, Willats WG, Schjoerring JK.
 2017. External nitrogen input affects pre-and post-harvest cell wall composition but not the enzymatic saccharification of wheat straw. *Biomass and Bioenergy* 98: 70-79.

- Bassi FM, Bentley AR, Charmet G, Ortiz R, Crossa J. 2016. Breeding schemes for the implementation of genomic selection in wheat (Triticum spp.). *Plant Science* **242**: 23-36.
- Battenfield SD, Guzmán C, Gaynor RC, Singh RP, Peña RJ, Dreisigacker S, Fritz AK, Poland JA.
 2016. Genomic selection for processing and end-use quality traits in the CIMMYT spring bread wheat breeding program. *The Plant Genome* 9.
- Bennypaul HS, Mutti JS, Rustgi S, Kumar N, Okubara PA, Gill KS. 2012a. Virus-induced gene silencing (VIGS) of genes expressed in root, leaf, and meiotic tissues of wheat. *Functional & integrative genomics* 12: 143-156.
- Bennypaul HS, Mutti JS, Rustgi S, Kumar N, Okubara PA, Gill KS. 2012b. Virus-induced gene silencing (VIGS) of genes expressed in root, leaf, and meiotic tissues of wheat. *Funct Integr Genomics* 12: 143-156.
- Bentsen NS, Felby C, Thorsen BJ. 2014. Agricultural residue production and potentials for energy and materials services. *Progress in energy and combustion science* **40**: 59-73.
- Bernal AJ, Jensen JK, Harholt J, Sørensen S, Moller I, Blaukopf C, Johansen B, De Lotto R, Pauly M, Scheller HV. 2007. Disruption of ATCSLD5 results in reduced growth, reduced xylan and homogalacturonan synthase activity and altered xylan occurrence in Arabidopsis. *The Plant Journal* 52: 791-802.
- Bernardo R, Yu J. 2007. Prospects for genomewide selection for quantitative traits in maize. *Crop Science* **47**: 1082-1090.
- Bhullar R, Nagarajan R, Bennypaul H, Sidhu GK, Sidhu G, Rustgi S, von Wettstein D, Gill KS.
 2014. Silencing of a metaphase I-specific gene results in a phenotype similar to that of the Pairing homeologous 1 (Ph1) gene mutations. *Proceedings of the National Academy of Sciences* 111: 14187-14192.

- Boerjan W, Ralph J, Baucher M. 2003. Lignin biosynthesis. *Annual review of plant biology* **54**: 519-546.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*: btu170.
- Bolot S, Abrouk M, Masood-Quraishi U, Stein N, Messing J, Feuillet C, Salse J. 2009. The 'inner circle' of the cereal genomes. *Curr Opin Plant Biol* **12**: 119-125.
- Bolser DM, Kerhornou A, Walts B, Kersey P. 2015. Triticeae resources in ensembl plants. *Plant* and Cell Physiology **56**: e3-e3.
- Bottley A, Xia G, Koebner R. 2006. Homoeologous gene silencing in hexaploid wheat. *The Plant Journal* **47**: 897-906.
- Brachi B, Morris GP, Borevitz JO. 2011. Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biol* **12**: 232.
- Bradbury LM, Niehaus TD, Hanson AD. 2013. Comparative genomics approaches to understanding and manipulating plant metabolism. *Curr Opin Biotechnol* **24**: 278-284.
- Bray N, Pimentel H, Melsted P, Pachter L. 2015. Near-optimal RNA-Seq quantification. *arXiv* preprint arXiv:150502710.
- Bregitzer P, Cooper LD, Hayes PM, Lemaux PG, Singh J, Sturbaum AK. 2007. Viability and bar expression are negatively correlated in Oregon Wolfe Barley Dominant hybrids. *Plant biotechnology journal* 5: 381-388.
- Breton C, Šnajdrová L, Jeanneau C, Koča J, Imberty A. 2006. Structures and mechanisms of glycosyltransferases. *Glycobiology* **16**: 29R-37R.

- Broom D, Galindo F, Murgueitio E. 2013. Sustainable, efficient livestock production with high biodiversity and good welfare for animals. *Proceedings of the Royal Society of London B: Biological Sciences* 280: 20132025.
- Buhrow LM, Clark SM, Loewen MC. 2016. Identification of an attenuated barley stripe mosaic virus for the virus-induced gene silencing of pathogenesis-related wheat genes. *Plant methods* 12: 12.
- Burton RA, Collins HM, Kibble NA, Smith JA, Shirley NJ, Jobling SA, Henderson M, Singh RR, Pettolino F, Wilson SM. 2011a. Over-expression of specific HVCSLF cellulose synthaselike 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**: 117-135.
- Burton RA, Collins HM, Kibble NA, Smith JA, Shirley NJ, Jobling SA, Henderson M, Singh RR, Pettolino F, Wilson SM et al. 2011b. Over-expression of specific HvCsIF cellulose synthase-like genes in transgenic barley increases the levels of cell wall (1,3;1,4)-beta-dglucans and alters their fine structure. *Plant Biotechnol J* **9**: 117-135.
- Burton RA, Fincher GB. 2014a. Evolution and development of cell walls in cereal grains.
- Burton RA, Fincher GB. 2014b. Plant cell wall engineering: applications in biofuel production and improved human health. *Current Opinion in Biotechnology* **26**: 79-84.
- Burton RA, Gibeaut DM, Bacic A, Findlay K, Roberts K, Hamilton A, Baulcombe DC, Fincher GB. 2000. Virus-induced silencing of a plant cellulose synthase gene. *The Plant Cell* 12: 691-705.
- Burton RA, Jobling SA, Harvey AJ, Shirley NJ, Mather DE, Bacic A, Fincher GB. 2008. The genetics and transcriptional profiles of the cellulose synthase-like HvCslF gene family in barley. *Plant Physiol* **146**: 1821-1833.

- Burton RA, Shirley NJ, King BJ, Harvey AJ, Fincher GB. 2004. The CesA gene family of barley. Quantitative analysis of transcripts reveals two groups of co-expressed genes. *Plant Physiol* **134**: 224-236.
- Burton RA, Wilson SM, Hrmova M, Harvey AJ, Shirley NJ, Medhurst A, Stone BA, Newbigin EJ, Bacic A, Fincher GB. 2006a. Cellulose synthase-like CslF genes mediate the synthesis of cell wall (1,3;1,4)-beta-D-glucans. *Science* **311**: 1940-1942.
- Burton RA, Wilson SM, Hrmova M, Harvey AJ, Shirley NJ, Medhurst A, Stone BA, Newbigin EJ, Bacic A, Fincher GB. 2006b. Cellulose synthase-like CslF genes mediate the synthesis of cell wall (1, 3; 1, 4)-B-D-glucans. *Science* **311**: 1940-1942.
- Cakir C, Tör M. 2010. Factors influencing Barley Stripe Mosaic Virus-mediated gene silencing in wheat. *Physiological and Molecular Plant Pathology* **74**: 246-253.
- Carpita NC. 1996. Structure and biogenesis of the cell walls of grasses. *Annual review of plant biology* **47**: 445-476.
- Carroll A, Mansoori N, Li S, Lei L, Vernhettes S, Visser RG, Somerville C, Gu Y, Trindade LM. 2012. Complexes with mixed primary and secondary cellulose synthases are functional in Arabidopsis plants. *Plant Physiol* 160: 726-737.
- Chan J, Calder G, Fox S, Lloyd C. 2007. Cortical microtubule arrays undergo rotary movements in Arabidopsis hypocotyl epidermal cells. *Nature Cell Biology* **9**: 171-175.
- Chan J, Crowell E, Eder M, Calder G, Bunnewell S, Findlay K, Vernhettes S, Höfte H, Lloyd C. 2010. The rotation of cellulose synthase trajectories is microtubule dependent and influences the texture of epidermal cell walls in Arabidopsis hypocotyls. *Journal of cell science* 123: 3490-3495.

- Chandra R, Takeuchi H, Hasegawa T. 2012. Methane production from lignocellulosic agricultural crop wastes: A review in context to second generation of biofuel production. *Renewable and Sustainable Energy Reviews* **16**: 1462-1476.
- Chantreau M, Chabbert B, Billiard S, Hawkins S, Neutelings G. 2015. Functional analyses of cellulose synthase genes in flax (Linum usitatissimum) by virus-induced gene silencing. *Plant biotechnology journal.*
- Chen L, Hao L, Parry MA, Phillips AL, Hu YG. 2014. Progress in TILLING as a Tool for Functional Genomics and Improvement of Crops. J Integr Plant Biol doi:10.1111/jipb.12192.
- Chen Y, Cao J. 2014. Comparative genomic analysis of the Sm gene family in rice and maize. *Gene* **539**: 238-249.
- Ching A, Dhugga KS, Appenzeller L, Meeley R, Bourett TM, Howard RJ, Rafalski A. 2006. Brittle stalk 2 encodes a putative glycosylphosphatidylinositol-anchored protein that affects mechanical strength of maize tissues by altering the composition and structure of secondary cell walls. *Planta* **224**: 1174-1184.
- Cho S, Garvin DF, Muehlbauer GJ. 2006. Transcriptome analysis and physical mapping of barley genes in wheat-barley chromosome addition lines. *Genetics* **172**: 1277-1285.
- Choulet F, Alberti A, Theil S, Glover N, Barbe V, Daron J, Pingault L, Sourdille P, Couloux A, Paux E. 2014a. Structural and functional partitioning of bread wheat chromosome 3B. *Science* 345: 1249721.
- Choulet F, Caccamo M, Wright J, Alaux M, Šimková H, Šafář J, Leroy P, Doležel J, Rogers J, Eversole K et al. 2014b. The Wheat Black Jack: Advances Towards Sequencing the 21 Chromosomes of Bread Wheat. doi:10.1007/978-94-007-7572-5_17: 405-438.

- Ciesielski PN, Resch MG, Hewetson B, Killgore JP, Curtin A, Anderson N, Chiaramonti AN, Hurley DC, Sanders A, Himmel ME et al. 2014. Engineering plant cell walls: tuning lignin monomer composition for deconstructable biofuel feedstocks or resilient biomaterials. *Green Chemistry* **16**: 2627.
- Cockram J, White J, Zuluaga DL, Smith D, Comadran J, Macaulay M, Luo Z, Kearsey MJ, Werner
 P, Harrap D et al. 2010. Genome-wide association mapping to candidate polymorphism
 resolution in the unsequenced barley genome. *Proc Natl Acad Sci U S A* 107: 21611-21616.
- Cocuron JC, Lerouxel O, Drakakaki G, Alonso AP, Liepman AH, Keegstra K, Raikhel N, Wilkerson CG. 2007. A gene from the cellulose synthase-like C family encodes a beta-1,4 glucan synthase. *Proc Natl Acad Sci U S A* **104**: 8550-8555.
- Cohen J. 1992. Statistical power analysis. *Current directions in psychological science* 1: 98-101.
- Consortium IWGS. 2014. A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science* **345**: 1251788.
- Coomey JH, Hazen SP. 2015. Brachypodium distachyon as a Model Species to Understand Grass Cell Walls. In *Genetics and Genomics of Brachypodium*, pp. 197-217. Springer.
- Cosgrove DJ. 2000. Expansive growth of plant cell walls. *Plant Physiology and Biochemistry* **38**: 109-124.
- Cosgrove DJ. 2005. Growth of the plant cell wall. *Nature reviews molecular cell biology* **6**: 850-861.
- Crossa J, Pérez P, de los Campos G, Mahuku G, Dreisigacker S, Magorokosho C. 2011. Genomic selection and prediction in plant breeding. *Journal of Crop Improvement* **25**: 239-261.

- Crowell EF, Bischoff V, Desprez T, Rolland A, Stierhof Y-D, Schumacher K, Gonneau M, Höfte H, Vernhettes S. 2009. Pausing of Golgi bodies on microtubules regulates secretion of cellulose synthase complexes in Arabidopsis. *The Plant Cell* 21: 1141-1154.
- Csuros M, Rogozin IB, Koonin EV. 2011. A detailed history of intron-rich eukaryotic ancestors inferred from a global survey of 100 complete genomes. *PLoS Comput Biol* **7**: e1002150.
- Desprez T, Juraniec M, Crowell EF, Jouy H, Pochylova Z, Parcy F, Hofte H, Gonneau M, Vernhettes S. 2007a. Organization of cellulose synthase complexes involved in primary cell wall synthesis in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* 104: 15572-15577.
- Desprez T, Juraniec M, Crowell EF, Jouy H, Pochylova Z, Parcy F, Höfte H, Gonneau M, Vernhettes S. 2007b. Organization of cellulose synthase complexes involved in primary cell wall synthesis in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences* **104**: 15572-15577.
- Devos KM, Wu X, Qi P. 2017. Genome Structure and Comparative Genomics. In *Genetics and Genomics of Setaria*, pp. 135-147. Springer.
- Dhaliwal AK, Mohan A, Gill KS. 2014. Comparative analysis of ABCB1 reveals novel structural and functional conservation between monocots and dicots. *Frontiers in plant science* 5: 657.
- Dhugga KS. 2001. Building the wall: genes and enzyme complexes for polysaccharide synthases. *Current opinion in plant biology* **4**: 488-493.
- Dhugga KS. 2007. Maize biomass yield and composition for biofuels. *Crop Science* **47**: 2211-2227.
- Dhugga KS. 2012. Biosynthesis of non-cellulosic polysaccharides of plant cell walls. *Phytochemistry* **74**: 8-19.

- Dhugga KS, Barreiro R, Whitten B, Stecca K, Hazebroek J, Randhawa GS, Dolan M, Kinney AJ, Tomes D, Nichols S. 2004a. Guar seed β-mannan synthase is a member of the cellulose synthase super gene family. *Science* **303**: 363-366.
- Dhugga KS, Barreiro R, Whitten B, Stecca K, Hazebroek J, Randhawa GS, Dolan M, Kinney AJ, Tomes D, Nichols S et al. 2004b. Guar seed beta-mannan synthase is a member of the cellulose synthase super gene family. *Science* **303**: 363-366.
- Ding S-Y, Zhao S, Zeng Y. 2013. Size, shape, and arrangement of native cellulose fibrils in maize cell walls. *Cellulose* **21**: 863-871.
- Doblin MS, De Melis L, Newbigin E, Bacic A, Read SM. 2001. Pollen tubes of Nicotiana alata express two genes from different β-glucan synthase families. *Plant Physiology* **125**: 2040-2052.
- Doblin MS, Pettolino F, Bacic A. 2010. Plant cell walls: the skeleton of the plant world. *Functional Plant Biology* **37**: 357-381.
- Doblin MS, Pettolino FA, Wilson SM, Campbell R, Burton RA, Fincher GB, Newbigin E, Bacic
 A. 2009. A barley cellulose synthase-like CSLH gene mediates (1,3;1,4)-beta-D-glucan synthesis in transgenic Arabidopsis. *Proc Natl Acad Sci U S A* 106: 5996-6001.
- Douche T, San Clemente H, Burlat V, Roujol D, Valot B, Zivy M, Pont-Lezica R, Jamet E. 2013.
 Brachypodium distachyon as a model plant toward improved biofuel crops: Search for secreted proteins involved in biogenesis and disassembly of cell wall polymers. *Proteomics* 13: 2438-2454.
- Douchkov D, Lueck S, Hensel G, Kumlehn J, Rajaraman J, Johrde A, Doblin MS, Beahan CT, Kopischke M, 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*.

Ebringerová A, Hromádková Z, Heinze T. 2005. Hemicellulose. 186: 1-67.

- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**: 1792-1797.
- Endler A, Kesten C, Schneider R, Zhang Y, Ivakov A, Froehlich A, Funke N, Persson S. 2015. A mechanism for sustained cellulose synthesis during salt stress. *Cell* **162**: 1353-1364.
- Endler A, Persson S. 2011. Cellulose synthases and synthesis in Arabidopsis. *Mol Plant* **4**: 199-211.
- Endo T, Fedorov A, de Souza SJ, Gilbert W. 2002. Do introns favor or avoid regions of amino acid conservation? *Molecular biology and evolution* **19**: 521-252.
- Eudes A, Liang Y, Mitra P, Loque D. 2014. Lignin bioengineering. *Curr Opin Biotechnol* **26**: 189-198.
- Famoso AN, Zhao K, Clark RT, Tung CW, Wright MH, Bustamante C, Kochian LV, McCouch SR. 2011. Genetic architecture of aluminum tolerance in rice (Oryza sativa) determined through genome-wide association analysis and QTL mapping. *PLoS Genet* 7: e1002221.
- FAN W-x, HOU Y-x, FENG S-w, ZHU F-k, RU Z-g. 2012. Study on Cellulose and Lodging Resistance of Wheat Straw. *Journal of Henan Agricultural Sciences* **9**: 010.
- Faris JD. 2014. Wheat domestication: Key to agricultural revolutions past and future. *Genomics of Plant Genetic Resources*: 439.
- Farrokhi N, Burton RA, Brownfield L, Hrmova M, Wilson SM, Bacic A, Fincher GB. 2006. Plant cell wall biosynthesis: genetic, biochemical and functional genomics approaches to the identification of key genes. *Plant Biotechnol J* 4: 145-167.

- Fincher GB. 2009. Revolutionary times in our understanding of cell wall biosynthesis and remodeling in the grasses. *Plant physiology* **149**: 27-37.
- Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic acids research* 44: D279-D285.
- Fry SC. 2001. Plant cell walls. eLS.
- Fry SC. 2004. Primary cell wall metabolism: tracking the careers of wall polymers in living plant cells. *New phytologist* **161**: 641-675.
- Fujii S, Hayashi T, Mizuno K. 2010. Sucrose synthase is an integral component of the cellulose synthesis machinery. *Plant and cell physiology* 51: 294-301.
- Gabhane J, William SP, Gadhe A, Rath R, Vaidya AN, Wate S. 2014. Pretreatment of banana agricultural waste for bio-ethanol production: individual and interactive effects of acid and alkali pretreatments with autoclaving, microwave heating and ultrasonication. *Waste Manag* **34**: 498-503.
- Gil-Humanes J, Pistón F, Barro F, Rosell CM. 2014. The Shutdown of Celiac Disease-Related Gliadin Epitopes in Bread Wheat by RNAi Provides Flours with Increased Stability and Better Tolerance to Over-Mixing. *PloS one* **9**: e91931.
- Girio FM, Fonseca C, Carvalheiro F, Duarte LC, Marques S, Bogel-Lukasik R. 2010. Hemicelluloses for fuel ethanol: A review. *Bioresour Technol* **101**: 4775-4800.
- Goodstein DM, Batra S, Carlson J, Hayes R, Phillips J, Shu S, Schmutz J, Rokhsar D. 2014. Phytozome Comparative Plant Genomics Portal. Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA (US).

- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U,
 Putnam N. 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic acids research* 40: D1178-D1186.
- Goubet F, Barton CJ, Mortimer JC, Yu X, Zhang Z, Miles GP, Richens J, Liepman AH, Seffen K, Dupree P. 2009. Cell wall glucomannan in Arabidopsis is synthesised by CSLA glycosyltransferases, and influences the progression of embryogenesis. *Plant J* 60: 527-538.
- Guerriero G, Hausman JF, Strauss J, Ertan H, Siddiqui KS. 2016. Lignocellulosic biomass:
 Biosynthesis, degradation, and industrial utilization. *Engineering in Life Sciences* 16: 1-16.
- Gupta P, Rustgi S. 2004. Molecular markers from the transcribed/expressed region of the genome in higher plants. *Functional & integrative genomics* **4**: 139-162.
- Gupta PK, Mir RR, Mohan A, Kumar J. 2008. Wheat genomics: present status and future prospects. *Int J Plant Genomics* **2008**: 896451.
- Gurung S, Mamidi S, Bonman JM, Xiong M, Brown-Guedira G, Adhikari TB. 2014. Genomewide association study reveals novel quantitative trait Loci associated with resistance to multiple leaf spot diseases of spring wheat. *PLoS One* **9**: e108179.
- Gutierrez R, Lindeboom JJ, Paredez AR, Emons AMC, Ehrhardt DW. 2009. Arabidopsis cortical microtubules position cellulose synthase delivery to the plasma membrane and interact with cellulose synthase trafficking compartments. *Nature Cell Biology* **11**: 797-806.
- Haigler CH, Davis JK, Slabaugh E, Kubicki JD. 2016. Biosynthesis and Assembly of Cellulose. Molecular Cell Biology of the Growth and Differentiation of Plant Cells: 120.

- Hamann T, Osborne E, Youngs HL, Misson J, Nussaume L, Somerville C. 2004. Global expression analysis of CESA and CSL genes in Arabidopsis. *Cellulose* **11**: 279-286.
- Hatfield RD, Marita JM, Frost K, Grabber J, Ralph J, Lu F, Kim H. 2009. Grass lignin acylation:
 p-coumaroyl transferase activity and cell wall characteristics of C3 and C4 grasses. *Planta* 229: 1253-1267.
- Hazen SP, Scott-Craig JS, Walton JD. 2002. Cellulose synthase-like genes of rice. *Plant Physiol* 128: 336-340.
- He XY, Zhang YL, He ZH, Wu YP, Xiao YG, Ma CX, Xia XC. 2008. Characterization of phytoene synthase 1 gene (Psy1) located on common wheat chromosome 7A and development of a functional marker. *Theor Appl Genet* **116**: 213-221.
- Held MA, Penning B, Brandt AS, Kessans SA, Yong W, Scofield SR, Carpita NC. 2008. Smallinterfering RNAs from natural antisense transcripts derived from a cellulose synthase gene modulate cell wall biosynthesis in barley. *Proceedings of the National Academy of Sciences* 105: 20534-20539.
- Hernández-Blanco C, Feng DX, Hu J, Sánchez-Vallet A, Deslandes L, Llorente F, Berrocal-Lobo
 M, Keller H, Barlet X, Sánchez-Rodríguez C. 2007. Impairment of cellulose synthases
 required for Arabidopsis secondary cell wall formation enhances disease resistance. *The Plant Cell* 19: 890-903.
- Hill JL, Hammudi MB, Tien M. 2014. The Arabidopsis Cellulose Synthase Complex: A Proposed Hexamer of CESA Trimers in an Equimolar Stoichiometry. *The Plant Cell Online*: tpc. 114.131193.

- Holland N, Holland D, Helentjaris T, Dhugga KS, Xoconostle-Cazares B, Delmer DP. 2000. A comparative analysis of the plant cellulose synthase (CesA) gene family. *Plant Physiology* 123: 1313-1323.
- Holzberg S, Brosio P, Gross C, Pogue GP. 2002. Barley stripe mosaic virus-induced gene silencing in a monocot plant. *The Plant Journal* **30**: 315-327.
- Hong Y, Chen L, Du LP, Su Z, Wang J, Ye X, Qi L, Zhang Z. 2014. Transcript suppression of TaGW2 increased grain width and weight in bread wheat. *Funct Integr Genomics* 14: 341-349.
- Hood EE. 2016. Plant-based biofuels. F1000Research 5.
- Houston K, Burton RA, Sznajder B, Rafalski AJ, Dhugga KS, Mather DE, Taylor J, Steffenson BJ, Waugh R, Fincher GB. 2015. A Genome-Wide Association Study for Culm Cellulose Content in Barley Reveals Candidate Genes Co-Expressed with Members of the CELLULOSE SYNTHASE A Gene Family. *PloS one* 10.
- Hu Z, Song N, Xing J, Chen Y, Han Z, Yao Y, Peng H, Ni Z, Sun Q. 2013. Overexpression of three TaEXPA1 homoeologous genes with distinct expression divergence in hexaploid wheat exhibit functional retention in Arabidopsis. *PloS one* 8: e63667.
- Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z. 2010. Genomewide association studies of 14 agronomic traits in rice landraces. *Nature genetics* 42: 961-967.
- Hunter CT, Kirienko DH, Sylvester AW, Peter GF, McCarty DR, Koch KE. 2012. Cellulose Synthase-Like D1 is integral to normal cell division, expansion, and leaf development in maize. *Plant physiology* 158: 708-724.

- Iehisa JC, Okada M, Sato K, Takumi S. 2017. Detection of splicing variants in the leaf and spike transcripts of wild diploid wheat Aegilops tauschii and transmission of the splicing patterns to synthetic hexaploid wheat. *Plant Gene* **9**: 6-12.
- Jaiswal V, Gahlaut V, Meher PK, Mir RR, Jaiswal JP, Rao AR, Balyan HS, Gupta PK. 2016. Genome Wide Single Locus Single Trait, Multi-Locus and Multi-Trait Association Mapping for Some Important Agronomic Traits in Common Wheat (T. aestivum L.). *PloS* one 11: e0159343.
- Jonas E, de Koning DJ. 2013. Does genomic selection have a future in plant breeding? *Trends Biotechnol* **31**: 497-504.
- Jordan KW, Wang S, Lun Y, Gardiner L-J, MacLachlan R, Hucl P, Wiebe K, Wong D, Forrest KL, Sharpe AG. 2015. A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. *Genome biology* **16**: 1.
- Joshi CP, Mansfield SD. 2007. The cellulose paradox—simple molecule, complex biosynthesis. *Current opinion in plant biology* **10**: 220-226.
- Kaur R, Singh K, Singh J. 2013. A root-specific wall-associated kinase gene, HvWAK1, regulates root growth and is highly divergent in barley and other cereals. *Functional & integrative genomics* 13: 167-177.
- Kaur S, Dhugga KS, Gill K, Singh J. 2016. Novel Structural and Functional Motifs in cellulose synthase (CesA) Genes of Bread Wheat (Triticum aestivum, L.). *PLoS One* **11**.

Keegstra K. 2010. Plant cell walls. *Plant physiology* 154: 483-486.

Kim CM, Park SH, Je BI, Park SH, Park SJ, Piao HL, Eun MY, Dolan L, Han C-d. 2007a. OsCSLD1, a cellulose synthase-like D1 gene, is required for root hair morphogenesis in rice. *Plant Physiology* 143: 1220-1230.

- Kim E, Magen A, Ast G. 2007b. Different levels of alternative splicing among eukaryotes. *Nucleic acids research* 35: 125-131.
- Kollers S, Rodemann B, Ling J, Korzun V, Ebmeyer E, Argillier O, Hinze M, Plieske J, Kulosa D, Ganal MW. 2013. Genetic architecture of resistance to Septoria tritici blotch (Mycosphaerella graminicola) in European winter wheat. *Molecular breeding* 32: 411-423.
- Kotake T, Aohara T, Hirano K, Sato A, Kaneko Y, Tsumuraya Y, Takatsuji H, Kawasaki S. 2011.
 Rice Brittle culm 6 encodes a dominant-negative form of CesA protein that perturbs cellulose synthesis in secondary cell walls. *Journal of Experimental Botany* 62: 2053-2062.
- Krasileva KV, Vasquez-Gross HA, Howell T, Bailey P, Paraiso F, Clissold L, Simmonds J, Ramirez-Gonzalez RH, Wang X, Borrill P. 2017. Uncovering hidden variation in polyploid wheat. *Proceedings of the National Academy of Sciences*: 201619268.
- Kumar M, Wightman R, Atanassov I, Gupta A, Hurst CH, Hemsley PA, Turner S. 2016. S-Acylation of the cellulose synthase complex is essential for its plasma membrane localization. *Science* **353**: 166-169.
- Kurek I, Kawagoe Y, Jacob-Wilk D, Doblin M, Delmer D. 2002. Dimerization of cotton fiber cellulose synthase catalytic subunits occurs via oxidation of the zinc-binding domains. *Proceedings of the National Academy of Sciences* **99**: 11109-11114.

Laird NM, Lange C. 2011. The fundamentals of modern statistical genetics. Springer.

- Lairson L, Henrissat B, Davies G, Withers S. 2008. Glycosyltransferases: structures, functions, and mechanisms. *Biochemistry* **77**: 521.
- Lee WS, Hammond-Kosack KE, Kanyuka K. 2012. Barley stripe mosaic virus-mediated tools for investigating gene function in cereal plants and their pathogens: virus-induced gene

silencing, host-mediated gene silencing, and virus-mediated overexpression of heterologous protein. *Plant Physiol* **160**: 582-590.

- Lei L, Zhang T, Strasser R, Lee CM, Gonneau M, Mach L, Vernhettes S, Kim SH, Cosgrove DJ, Li S. 2014. The jiaoyao1 mutant is an allele of korrigan1 that abolishes endoglucanase activity and affects the organization of both cellulose microfibrils and microtubules in Arabidopsis. *The Plant Cell* 26: 2601-2616.
- Li K, Wang H, Hu X, Liu Z, Wu Y, Huang C. 2016. Genome-wide association study reveals the genetic basis of stalk cell wall components in maize. *PloS one* **11**: e0158906.
- Li M, Xiong G, Li R, Cui J, Tang D, Zhang B, Pauly M, Cheng Z, Zhou Y. 2009. Rice cellulose synthase-like D4 is essential for normal cell-wall biosynthesis and plant growth. *The Plant Journal* **60**: 1055-1069.
- Li S, Bashline L, Lei L, Gu Y. 2014a. Cellulose synthesis and its regulation. *Arabidopsis Book* **12**: e0169.
- Li S, Logan Bashline LL, Gu Y. 2014b. Cellulose Synthesis and Its Regulation. *The Arabidopsis* book/American Society of Plant Biologists **12**.
- Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658-1659.
- Liepman AH, Cavalier DM. 2012. The CELLULOSE SYNTHASE-LIKE A and CELLULOSE SYNTHASE-LIKE C families: recent advances and future perspectives. *Frontiers in plant science* **3**: 109.
- Liepman AH, Wilkerson CG, Keegstra K. 2005. Expression of cellulose synthase-like (Csl) genes in insect cells reveals that CslA family members encode mannan synthases. *Proc Natl Acad Sci U S A* **102**: 2221-2226.

- Lin F, Manisseri C, Fagerström A, Peck ML, Vega-Sánchez ME, Williams B, Chiniquy DM, Saha P, Pattathil S, Conlin B. 2016. Cell Wall Composition and Candidate Biosynthesis Gene Expression Across Rice Development. *Plant and Cell Physiology*: pcw125.
- Lin Y, Kao Y-Y, Chen Z-Z, Chu F-H, Chung J-D. 2014. cDNA cloning and molecular characterization of five cellulose synthase A genes from Eucalyptus camaldulensis. *Journal of plant biochemistry and biotechnology* 23: 199-210.
- Lindedam J, Andersen SB, DeMartini J, Bruun S, Jørgensen H, Felby C, Magid J, Yang B, Wyman C. 2012. Cultivar variation and selection potential relevant to the production of cellulosic ethanol from wheat straw. *biomass and bioenergy* **37**: 221-228.
- Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z. 2012. GAPIT: genome association and prediction integrated tool. *Bioinformatics* **28**: 2397-2399.
- Liu X, Huang M, Fan B, Buckler ES, Zhang Z. 2016. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet* 12: e1005767.
- Liu X, Wang Q, Chen P, Song F, Guan M, Jin L, Wang Y, Yang C. 2012. Four Novel Cellulose Synthase (CESA) Genes from Birch (Betula platyphylla Suk.) Involved in Primary and Secondary Cell Wall Biosynthesis. *Int J Mol Sci* 13: 12195-12212.
- Lopes M, Dreisigacker S, Peña R, Sukumaran S, Reynolds M. 2015. Genetic characterization of the wheat association mapping initiative (WAMI) panel for dissection of complex traits in spring wheat. *Theoretical and Applied Genetics* **128**: 453-464.
- Ma J, Stiller J, Berkman PJ, Wei Y, Rogers J, Feuillet C, Dolezel J, Mayer KF, Eversole K, Zheng Y-L. 2013. Sequence-based analysis of translocations and inversions in bread wheat (Triticum aestivum L.). *PloS one* **8**: e79329.

- Madramootoo C. 2015. Emerging Technologies for Promoting Food Security: Overcoming the World Food Crisis. Woodhead Publishing.
- Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI. 2014. CDD: NCBI's conserved domain database. *Nucleic acids research*: gku1221.
- Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M, Jakobsen KS, Wulff BB, Steuernagel B, Mayer KF, Olsen O-A. 2014. Ancient hybridizations among the ancestral genomes of bread wheat. *Science* 345: 1250092.
- Mayer KF, Rogers J, Doležel J, Pozniak C, Eversole K, Feuillet C, Gill B, Friebe B, Lukaszewski AJ, Sourdille P. 2014. A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science* **345**: 1251788.
- McFarlane HE, Doring A, Persson S. 2014. The cell biology of cellulose synthesis. *Annu Rev Plant Biol* **65**: 69-94.
- McGrann GR, Smith PH, Burt C, Mateos GR, Chama TN, MacCormack R, Wessels E, Agenbag G, Boyd LA. 2014. Genomic and genetic analysis of the wheat race-specific yellow rust resistance gene Yr5. *Journal of Plant Science and Molecular Breeding* **3**.
- McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C et al. 2009. Genetic properties of the maize nested association mapping population. *Science* **325**: 737-740.
- Meuwissen THE, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genomewide dense marker maps. *Genetics* **157**: 1819-1829.

- Michel S, Ametz C, Gungor H, Epure D, Grausgruber H, Löschenberger F, Buerstmayr H. 2016.
 Genomic selection across multiple breeding cycles in applied bread wheat breeding.
 Theoretical and Applied Genetics 129: 1179-1189.
- Mittler R, Shulaev V. 2013. Functional genomics, challenges and perspectives for the future. *Physiologia Plantarum* **148**: 317-321.
- Mochida K, Shinozaki K. 2010. Genomics and bioinformatics resources for crop improvement. *Plant Cell Physiol* **51**: 497-523.
- Mochida K, Shinozaki K. 2013. Unlocking Triticeae genomics to sustainably feed the future. *Plant Cell Physiol* **54**: 1931-1950.
- Mochida K, Yamazaki Y, Ogihara Y. 2004. Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. *Molecular Genetics and Genomics* **270**: 371-377.
- Mohan A, Schillinger WF, Gill KS. 2013. Wheat seedling emergence from deep planting depths and its relationship with coleoptile length. *PloS one* **8**: e73314.
- Molnár I, Vrána J, Burešová V, Cápal P, Farkas A, Darkó É, Cseh A, Kubaláková M, Molnár-Láng M, Doležel J. 2016. Dissecting the U, M, S and C genomes of wild relatives of bread wheat (Aegilops spp.) into chromosomes and exploring their synteny with wheat. *The Plant Journal* 88: 452-467.
- Monaco MK, Stein J, Naithani S, Wei S, Dharmawardhana P, Kumari S, Amarasinghe V, Youens-Clark K, Thomason J, Preece J et al. 2014. Gramene 2013: comparative plant genomics resources. *Nucleic Acids Research* **42**: D1193-D1199.
- Morgan JL, Strumillo J, Zimmer J. 2013. Crystallographic snapshot of cellulose synthesis and membrane translocation. *Nature* **493**: 181-186.

Moriarty P, Honnery D. 2016. Can renewable energy power the future? Energy Policy 93: 3-7.

- Motte JC, Escudié R, Beaufils N, Steyer JP, Bernet N, Delgenès JP, Dumas C. 2014. Morphological structures of wheat straw strongly impacts its anaerobic digestion. *Industrial Crops and Products* 52: 695-701.
- Muthamilarasan M, Theriappan P, Prasad M. 2013. Recent advances in crop genomics for ensuring food security. *Curr Sci* **105**: 155-158.
- Nemeth C, Freeman J, Jones HD, Sparks C, Pellny TK, Wilkinson MD, Dunwell J, Andersson AAM, Aman P, Guillon F et al. 2010. Down-Regulation of the CSLF6 Gene Results in Decreased (1,3;1,4)-beta-D-Glucan in Endosperm of Wheat. *Plant Physiology* **152**: 1209-1218.
- Neumann K, Kobiljski B, Denčić S, Varshney RK, Börner A. 2010. Genome-wide association mapping: a case study in bread wheat (Triticum aestivum L.). *Molecular Breeding* 27: 37-58.
- O'brien T, Feder N, McCully ME. 1964. Polychromatic staining of plant cell walls by toluidine blue O. *Protoplasma* **59**: 368-373.
- Ong RG, Chundawat SPS, Hodge DB, Keskar S, Dale BE. 2014. Linking Plant Biology and Pretreatment: Understanding the Structure and Organization of the Plant Cell Wall and Interactions with Cellulosic Biofuel Production. doi:10.1007/978-1-4614-9329-7_14: 231-253.
- Paredez AR, Somerville CR, Ehrhardt DW. 2006. Visualization of cellulose synthase demonstrates functional association with microtubules. *Science* **312**: 1491-1495.
- Pauly M, Gille S, Liu L, Mansoori N, de Souza A, Schultink A, Xiong G. 2013. Hemicellulose biosynthesis. *Planta* 238: 627-642.

- Pauly M, Keegstra K. 2008. Cell-wall carbohydrates and their modification as a resource for biofuels. *The Plant Journal* 54: 559-568.
- Penning BW, Sykes RW, Babcock NC, Dugard CK, Klimek JF, Gamblin D, Davis M, Filley TR, Mosier NS, Weil CF et al. 2014. Validation of PyMBMS as a High-throughput Screen for Lignin Abundance in Lignocellulosic Biomass of Grasses. *BioEnergy Research* doi:10.1007/s12155-014-9410-3.
- Perera FP. 2016. Multiple threats to child health from fossil fuel combustion: Impacts of air pollution and climate change. *Environ Health Perspect*.
- Persson S, Paredez A, Carroll A, Palsdottir H, Doblin M, Poindexter P, Khitrov N, Auer M, Somerville CR. 2007. Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in Arabidopsis. *Proc Natl Acad Sci U S A* **104**: 15566-15571.
- Petty I, Hunter B, Wei N, Jackson A. 1989. Infectious barley stripe mosaic virus RNA transcribed in vitro from full-length genomic cDNA clones. *Virology* **171**: 342-349.
- Pingault L, Choulet F, Alberti A, Glover N, Wincker P, Feuillet C, Paux E. 2015. Deep transcriptome sequencing provides new insights into the structural and functional organization of the wheat genome. *Genome biology* **16**: 29.
- Poczai P, Varga I, Laos M, Cseh A, Bell N, Valkonen JP, Hyvönen J. 2013. Advances in plant gene-targeted and functional markers: a review. *Plant methods* **9**: 6.
- Porth I, Klapšte J, Skyba O, Hannemann J, McKown AD, Guy RD, DiFazio SP, Muchero W, Ranjan P, Tuskan GA. 2013. Genome-wide association mapping for wood characteristics in Populus identifies an array of candidate single nucleotide polymorphisms. *New Phytologist* 200: 710-726.

- Poursarebani N, Ariyadasa R, Zhou R, Schulte D, Steuernagel B, Martis MM, Graner A, Schweizer
 P, Scholz U, Mayer K et al. 2013. Conserved synteny-based anchoring of the barley genome physical map. *Funct Integr Genomics* 13: 339-350.
- Qiu Z, Wan L, Chen T, Wan Y, He X, Lu S, Wang Y, Lin J. 2013. The regulation of cambial activity in Chinese fir (Cunninghamia lanceolata) involves extensive transcriptome remodeling. *New Phytologist* **199**: 708-719.
- Qureshi N, Saha BC, Cotta MA, Singh V. 2013. An economic evaluation of biological conversion of wheat straw to butanol: A biofuel. *Energy Conversion and Management* **65**: 456-462.
- Ramegowda V, Mysore KS, Senthil-Kumar M. 2014. Virus-induced gene silencing is a versatile tool for unraveling the functional relevance of multiple abiotic-stress-responsive genes in crop plants. *Frontiers in plant science* **5**.
- Ranik M, Myburg AA. 2006. Six new cellulose synthase genes from Eucalyptus are associated with primary and secondary cell wall biosynthesis. *Tree physiology* **26**: 545-556.
- Ranocha P, Denancé N, Vanholme R, Freydier A, Martinez Y, Hoffmann L, Köhler L, Pouzet C,
 Renou JP, Sundberg B. 2010. Walls are thin 1 (WAT1), an Arabidopsis homolog of
 Medicago truncatula NODULIN21, is a tonoplast-localized protein required for secondary
 wall formation in fibers. *The Plant Journal* 63: 469-483.
- Rao G, Zeng Y, He C, Zhang J. 2016. Characterization and putative post-translational regulation of α -and β -tubulin gene families in Salix arbutifolia. *Scientific reports* **6**.
- Rayon C, Olek AT, Carpita NC. 2014. Towards Redesigning Cellulose Biosynthesis for Improved Bioenergy Feedstocks. In *Plants and BioEnergy*, pp. 183-193. Springer.

- Rejab NA, Nakano Y, Yoneda A, Ohtani M, Demura T. 2015. Possible contribution of TED6 and TED7, secondary cell wall-related membrane proteins, to evolution of tracheary element in angiosperm lineage. *Plant Biotechnology* **32**: 343-347.
- Richmond T. 2000. Higher plant cellulose synthases. *Genome biology* 1: REVIEWS3001-REVIEWS3001.
- Richmond TA, Somerville CR. 2000. The cellulose synthase superfamily. *Plant physiology* **124**: 495-498.
- Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.
- Ruiz HA, Cerqueira MA, Silva HD, Rodriguez-Jasso RM, Vicente AA, Teixeira JA. 2013. Biorefinery valorization of autohydrolysis wheat straw hemicellulose to be applied in a polymer-blend film. *Carbohydr Polym* **92**: 2154-2162.
- Ruiz MT, Voinnet O, Baulcombe DC. 1998. Initiation and maintenance of virus-induced gene silencing. *The Plant Cell* **10**: 937-946.
- Saini JK, Saini R, Tewari L. 2015. Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: concepts and recent developments. *3 Biotech* 5: 337-353.
- Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution* **4**: 406-425.
- Salgotra RK, Gupta BB, Stewart CN, Jr. 2014. From genomics to functional markers in the era of next-generation sequencing. *Biotechnol Lett* **36**: 417-426.
- Sánchez-Rodríguez C, Bauer S, Hématy K, Saxe F, Ibáñez AB, Vodermaier V, Konlechner C, Sampathkumar A, Rüggeberg M, Aichinger E. 2012. Chitinase-like1/pom-pom1 and its

homolog CTL2 are glucan-interacting proteins important for cellulose biosynthesis in Arabidopsis. *The Plant Cell* **24**: 589-607.

- Sarkar P, Bosneaga E, Auer M. 2009. Plant cell walls throughout evolution: towards a molecular understanding of their design principles. *Journal of Experimental Botany* **60**: 3615-3635.
- Saxena IM, Brown Jr RM, Fevre M, Geremia RA, Henrissat B. 1995. Multidomain architecture of beta-glycosyl transferases: implications for mechanism of action. *Journal of Bacteriology* 177: 1419.
- Saxena IM, Brown R. 1995. Identification of a second cellulose synthase gene (acsAII) in Acetobacter xylinum. *Journal of bacteriology* **177**: 5276-5283.
- Saxena IM, Brown RM. 2005. Cellulose biosynthesis: current views and evolving concepts. Annals of botany **96**: 9-21.
- Scholey D, Burton E, Williams P. 2016. The bio refinery; producing feed and fuel from grain. *Food Chemistry* **197**: 937-942.
- Schreiber M, Wright F, MacKenzie K, Hedley PE, Schwerdt JG, Little A, Burton RA, Fincher GB, Marshall D, Waugh R. 2014a. 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.
- Schreiber M, Wright F, MacKenzie K, Hedley PE, Schwerdt JG, Little A, Burton RA, Fincher GB,
 Marshall D, Waugh R et al. 2014b. The barley genome sequence assembly reveals three additional members of the CslF (1,3;1,4)-beta-glucan synthase gene family. *PLoS One* 9: e90888.

- Schwerdt JG, MacKenzie K, Wright F, Oehme D, Wagner JM, Harvey AJ, Shirley NJ, Burton RA, Schreiber M, Halpin C. 2015. Evolutionary dynamics of the Cellulose synthase gene superfamily in grasses. *Plant physiology* **168**: 968-983.
- Scofield SR, Huang L, Brandt AS, Gill BS. 2005. Development of a virus-induced gene-silencing system for hexaploid wheat and its use in functional analysis of the Lr21-mediated leaf rust resistance pathway. *Plant Physiology* **138**: 2165-2173.
- Senthil-Kumar M, Mysore KS. 2011. New dimensions for VIGS in plant functional genomics. *Trends Plant Sci* 16: 656-665.
- Sethaphong L, Haigler CH, Kubicki JD, Zimmer J, Bonetta D, DeBolt S, Yingling YG. 2013. Tertiary model of a plant cellulose synthase. *Proceedings of the National Academy of Sciences* 110: 7512-7517.
- Shaveta, Bansal N, Singh P. 2014. F–/Cl– mediated microwave assisted breakdown of cellulose to glucose. *Tetrahedron Letters* **55**: 2467-2470.
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology* 7: 539.
- Simeao Resende RM, Casler MD, Vilela de Resende MD. 2014. Genomic Selection in Forage Breeding: Accuracy and Methods. *Crop Science* **54**: 143-156.
- Singh J, Zhang S, Chen C, Cooper L, Bregitzer P, Sturbaum A, Hayes PM, Lemaux PG. 2006. High-frequency Ds remobilization over multiple generations in barley facilitates gene tagging in large genome cereals. *Plant molecular biology* 62: 937-950.

- Singh M, Singh S, Randhawa H, Singh J. 2013. Polymorphic homoeolog of key gene of RdDM pathway, ARGONAUTE4_9 class is associated with pre-harvest sprouting in wheat (Triticum aestivum L.). *PLoS One* **8**: e77009.
- Singh S, Tan HQ, Singh J. 2012. Mutagenesis of barley malting quality QTLs with Ds transposons. *Functional & Integrative Genomics* **12**: 131-141.
- Singhania RR, Saini JK, Saini R, Adsul M, Mathur A, Gupta R, Tuli DK. 2014. Bioethanol production from wheat straw via enzymatic route employing Penicillium janthinellum cellulases. *Bioresource technology* **169**: 490-495.
- Slabaugh E, Davis JK, Haigler CH, Yingling YG, Zimmer J. 2014. Cellulose synthases: new insights from crystallography and modeling. *Trends in Plant Science* **19**: 99-106.
- Slavov GT, Nipper R, Robson P, Farrar K, Allison GG, Bosch M, Clifton-Brown JC, Donnison IS, Jensen E. 2014. Genome-wide association studies and prediction of 17 traits related to phenology, biomass and cell wall composition in the energy grass Miscanthus sinensis. *New Phytologist* 201: 1227-1239.
- Sokhansanj S, Mani S, Stumborg M, Samson R, Fenton J. 2006. Production and distribution of cereal straw on the Canadian Prairies. *Canadian Biosystems Engineering* **48**: 3.
- Sorek N, Yeats TH, Szemenyei H, Youngs H, Somerville CR. 2014. The Implications of Lignocellulosic Biomass Chemical Composition for the Production of Advanced Biofuels. *BioScience* 64: 192-201.
- Stothard P. 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **28**: 1102, 1104-1102, 1104.
- Stratmann JW, Hind SR. 2011. Gene silencing goes viral and uncovers the private life of plants. *Entomologia Experimentalis Et Applicata* **140**: 91-102.

Strezov V, Evans TJ. 2014. Biomass Processing Technologies. CRC Press.

- Su Z, Hao C, Wang L, Dong Y, Zhang X. 2011. Identification and development of a functional marker of TaGW2 associated with grain weight in bread wheat (Triticum aestivum L.). *Theor Appl Genet* 122: 211-223.
- Sukumaran S, Dreisigacker S, Lopes M, Chavez P, Reynolds MP. 2015. Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments. *Theoretical and Applied Genetics* **128**: 353-363.
- Suzuki S, Li L, Sun Y-H, 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.
- Szyjanowicz PM, McKinnon I, Taylor NG, Gardiner J, Jarvis MC, Turner SR. 2004. The irregular xylem 2 mutant is an allele of korrigan that affects the secondary cell wall of Arabidopsis thaliana. *The Plant Journal* **37**: 730-740.
- Tai Y, Bragg J, Edwards MC. 2005. Virus vector for gene silencing in wheat. *Biotechniques* 39: 310.
- Takata N, Taniguchi T. 2015. Expression divergence of cellulose synthase (CesA) genes after a recent whole genome duplication event in Populus. *Planta* **241**: 29-42.
- Taketa S, Yuo T, Tonooka T, Tsumuraya Y, Inagaki Y, Haruyama N, Larroque O, Jobling SA. 2012. Functional characterization of barley betaglucanless mutants demonstrates a unique role for CslF6 in (1,3;1,4)-beta-D-glucan biosynthesis. *J Exp Bot* 63: 381-392.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution* **30**: 2725-2729.

- Tanaka K, Murata K, Yamazaki M, Onosato K, Miyao A, Hirochika H. 2003. Three distinct rice cellulose synthase catalytic subunit genes required for cellulose synthesis in the secondary wall. *Plant Physiology* 133: 73-83.
- Tanaka M, Tanaka H, Shitsukawa N, Kitagawa S, Takumi S, Murai K. 2015. Homoeologous copyspecific expression patterns of MADS-box genes for floral formation in allopolyploid wheat. *Genes & genetic systems*.
- Tang Y, Liu X, Wang J, Li M, Wang Q, Tian F, Su Z, Pan Y, Liu D, Lipka AE. 2016. GAPIT Version 2: an enhanced integrated tool for genomic association and prediction. *The Plant Genome* 9.
- Taylor-Teeples M, Lin L, de Lucas M, Turco G, Toal T, Gaudinier A, Young N, Trabucco G, Veling M, Lamothe R. 2015. An Arabidopsis gene regulatory network for secondary cell wall synthesis. *Nature* 517: 571-575.
- Taylor NG, Howells RM, Huttly AK, Vickers K, Turner SR. 2003. Interactions among three distinct CesA proteins essential for cellulose synthesis. *Proc Natl Acad Sci U S A* 100: 1450-1455.
- Taylor NG, Scheible W-R, Cutler S, Somerville CR, Turner SR. 1999. The irregular xylem3 locus of Arabidopsis encodes a cellulose synthase required for secondary cell wall synthesis. *The plant cell* **11**: 769-779.
- Team RC. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013.
- Thomas LH, Forsyth VT, Šturcová A, Kennedy CJ, May RP, Altaner CM, Apperley DC, Wess TJ, Jarvis MC. 2013. Structure of cellulose microfibrils in primary cell walls from collenchyma. *Plant physiology* 161: 465-476.

- Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES. 2011. Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat Genet* 43: 159-162.
- Tian J, Chang M, Du Q, Xu B, Zhang D. 2014. Single-nucleotide polymorphisms in PtoCesA7 and their association with growth and wood properties in Populus tomentosa. *Molecular Genetics and Genomics* 289: 439-455.
- Trapp SC, Croteau RB. 2001. Genomic organization of plant terpene synthases and molecular evolutionary implications. *Genetics* **158**: 811-832.
- Tumuluru JS, Tabil LG, Song Y, Iroba KL, Meda V. 2014. Grinding energy and physical properties of chopped and hammer-milled barley, wheat, oat, and canola straws. *Biomass and Bioenergy* **60**: 58-67.
- Turner SR, Somerville CR. 1997. Collapsed xylem phenotype of Arabidopsis identifies mutants deficient in cellulose deposition in the secondary cell wall. *The plant cell* **9**: 689-701.
- Vain T, Crowell EF, Timpano H, Biot E, Desprez T, Mansoori N, Trindade LM, Pagant S, Robert S, Höfte H. 2014. The cellulase KORRIGAN is part of the cellulose synthase complex.
 Plant physiology 165: 1521-1532.
- Varshney RK, Mahendar T, Aggarwal RK, Börner A. 2007. Genic molecular markers in plants: development and applications. In *Genomics-assisted crop improvement*, pp. 13-29. Springer.
- Verhertbruggen Y, Yin L, Oikawa A, Scheller HV. 2011. Mannan synthase activity in the CSLD family. *Plant signaling & behavior* **6**: 1620.
- Vogel J. 2008. Unique aspects of the grass cell wall. *Current opinion in plant biology* **11**: 301-307.

- Wang D, Yuan S, Yin L, Zhao J, Guo B, Lan J, Li X. 2012a. A missense mutation in the transmembrane domain of CESA9 affects cell wall biosynthesis and plant growth in rice. *Plant Science* 196: 117-124.
- Wang H, Smith KP, Combs E, Blake T, Horsley RD, Muehlbauer GJ. 2012b. Effect of population size and unbalanced data sets on QTL detection using genome-wide association mapping in barley breeding germplasm. *Theoretical and Applied Genetics* 124: 111-124.
- Wang L, Guo K, Li Y, Tu Y, Hu H, Wang B, Cui X, Peng L. 2010a. Expression profiling and integrative analysis of the CESA/CSL superfamily in rice. *BMC plant biology* **10**: 282.
- Wang L, Guo K, Li Y, Tu Y, Hu H, Wang B, Cui X, Peng L. 2010b. Expression profiling and integrative analysis of the CESA/CSL superfamily in rice. *Bmc Plant Biology* **10**.
- Wang W, Wang L, Chen C, Xiong G, Tan X-Y, Yang K-Z, Wang Z-C, Zhou Y, Ye D, Chen L-Q. 2011. Arabidopsis CSLD1 and CSLD4 are required for cellulose deposition and normal growth of pollen tubes. *Journal of experimental botany*: err221.
- Wang Y, You FM, Lazo GR, Luo M-C, Thilmony R, Gordon S, Kianian SF, Gu YQ. 2013. PIECE:
 a database for plant gene structure comparison and evolution. *Nucleic acids research* 41: D1159-D1166.
- *Wasteneys* GO. 2004. Progress in understanding the role of microtubules in plant cells. *Current opinion in plant biology* **7**: 651-660.
- Wightman R, Turner SR. 2008. The roles of the cytoskeleton during cellulose deposition at the secondary cell wall. *The Plant Journal* **54**: 794-805.
- Wong C, Bernardo R. 2008. Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. *Theoretical and Applied Genetics* **116**: 815-824.

- Xu Y, Xie C, Wan J, He Z, Prasanna BM. 2013. Marker-Assisted Selection in Cereals: Platforms, Strategies and Examples. doi:10.1007/978-94-007-6401-9_14: 375-411.
- Yin L, Verhertbruggen Y, Oikawa A, Manisseri C, Knierim B, Prak L, Jensen JK, Knox JP, Auer M, Willats WG. 2011. The cooperative activities of CSLD2, CSLD3, and CSLD5 are required for normal Arabidopsis development. *Molecular plant* 4: 1024-1037.
- Yin Y, Huang J, Xu Y. 2009. The cellulose synthase superfamily in fully sequenced plants and algae. *BMC Plant Biol* **9**: 99.
- Yin Y, Johns MA, Cao H, 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.
- Yong W, Link B, O'Malley R, Tewari J, Hunter CT, Lu C-A, Li X, Bleecker AB, Koch KE, McCann MC. 2005. Genomics of plant cell wall biogenesis. *Planta* 221: 747-751.
- Yoshikawa T, Eiguchi M, Hibara K-I, Ito J-I, Nagato Y. 2013. Rice SLENDER LEAF 1 gene encodes cellulose synthase-like D4 and is specifically expressed in M-phase cells to regulate cell proliferation. *Journal of experimental botany* **64**: 2049-2061.
- Yuo T, Shiotani K, Shitsukawa N, Miyao A, Hirochika H, Ichii M, Taketa S. 2011. Root hairless
 2 (rth2) mutant represents a loss-of-function allele of the cellulose synthase-like gene
 OsCSLD1 in rice (Oryza sativa L.). *Breeding science* 61: 225-233.
- Zhang B, Deng L, Qian Q, Xiong G, Zeng D, Li R, Guo L, Li J, Zhou Y. 2009. A missense mutation in the transmembrane domain of CESA4 affects protein abundance in the plasma membrane and results in abnormal cell wall biosynthesis in rice. *Plant Molecular Biology* 71: 509-524.

- Zhang H, Fangel JU, Willats WGT, Selig MJ, Lindedam J, Jørgensen H, Felby C. 2014a. Assessment of leaf/stem ratio in wheat straw feedstock and impact on enzymatic conversion. *GCB Bioenergy* **6**: 90-96.
- Zhang N, Huo W, Zhang L, Chen F, Cui D. 2016. Identification of Winter-Responsive Proteins in Bread Wheat Using Proteomics Analysis and Virus-Induced Gene Silencing (VIGS). *Molecular & Cellular Proteomics* 15: 2954-2969.
- Zhang Q, Cheetamun R, Dhugga KS, Rafalski JA, Tingey SV, Shirley NJ, Taylor J, Hayes K, Beatty M, Bacic A. 2014b. Spatial gradients in cell wall composition and transcriptional profiles along elongating maize internodes. *BMC plant biology* 14: 27.
- Zhang Y, Ghaly A, Li B. 2012. PHYSICAL PROPERTIES OF RICE RESIDUES AS AFFECTED BY VARIETY AND CLIMATIC AND CULTIVATION ONDITIONS IN THREE CONTINENTS. American Journal of Applied Sciences **9**: 1757.
- Zhong R, Ye Z-H. 2014. Secondary cell walls: biosynthesis, patterned deposition and transcriptional regulation. *Plant and Cell Physiology*: pcu140.
- Zhou Y, Zhou C, Ye L, Dong J, Xu H, Cai L, Zhang L, Wei L. 2003. Database and analyses of known alternatively spliced genes in plants. *Genomics* 82: 584-595.
- Zuckerkandl E, Pauling L. 1965. Evolutionary divergence and convergence in proteins. *Evolving genes and proteins* **97**: 97-166.