Cellulose, Stem Strength and the Endo-(1,4)-β-Glucanase Gene Family in Barley and Maize.

By Margaret Buchanan

School of Agriculture Food and Wine

The University of Adelaide

October 2011

TABLE OF CONTENTS

Table of Figures……viii

Table of Tables……..xiii

Table of Equations……xv

Declaration……....xvii

Acknowledgements……..……..xix

Abstract……..xxi

Abbreviations……..……xxiii

Chapter 1 1

From Stalk Lodging to the Endo-(1,4)-β-glucanase Gene Family in Plants: An Overview

1.1. LODGING IN CEREAL CROPS 3

1.2. CELLULOSE 7

1.3. THE PLANT CELL WALL 12

1.3.1. Primary cell wall 17

1.3.2. Secondary cell walls 17

1.4. THE MONOCOTYLEDON STEM 19

1.5. CELLULOSE AND STEM STRENGTH 23

1.6. ENDO-(1,4)-B-GLUCANASES 24

1.6.1. Plant endo- (1,4)-β-glucanases 29

1.6.2. Endo-(1,4)-β-glucanases and CBDs 29

1.6.3. *In vitro* substrates 31

1.6.4. Endo-(1,4)-β-glucanases in cell wall breakdown 31

1.6.5. Endo-(1,4)-β-glucanases in plant cell wall synthesis 32

1.7. SUMMARY 36

Chapter 2 41

The Endo-(1,4)-β-glucanase Gene Families in Barley, Maize and Other Cereals

2.1. INTRODUCTION 43

2.2. METHODS....51
2.2.1. Database searches for GH9 endo-(1,4)-β-glucanase sequence data....51
2.2.1.1. Rice, sorghum and Arabidopsis....51
2.2.1.2. Barley....51
2.2.1.3. Maize....51
2.2.2. 3' SMART RACE PCR in barley....52
2.2.2.1. Preparation of RNA....53
2.2.2.2. 3' SMART RACE cDNA preparation....53
2.2.2.3. 3' SMART RACE PCR....54
2.2.3. Polymerase Chain Reaction (PCR) using Platinum Taq polymerase....54
2.2.4. Observation of nucleic acids by agarose gel electrophoresis 56
2.2.5. Cleaning of PCR products....56
2.2.6. DNA Sequencing....57
2.2.7. Cloning into the pGEM-T Easy vector....59
2.2.7.1. A-tailing PCR product....59
2.2.7.2. Ligation....59
2.2.7.3. Transformation of DH5α Escherichia coli....59
2.2.7.4. Alkaline lysis plasmid preparation....60
2.2.7.5. Restriction enzyme digest....61
2.2.8. Barley BAC library screen....61
2.2.8.1. Preparation of ^{32}P labeled DNA Probe....61
2.2.8.2. Probing BAC library filters....63
2.2.8.3. Identification/authentication of positive clones....64
2.2.8.4. Extraction of plasmid from BAC clones....64
2.2.9. Endo-(1,4)-β-glucanase sequence alignment and phylogenetic trees....65
2.2.10. Gene intron and exon structure....65
2.2.11. Mapping....66
2.2.12. Measuring pairwise codon-based evolutionary distance....66
2.2.12.1. Nei and Gojobori measurement of evolutionary distance 67
2.2.12.2. Phylogenetic analysis by maximum likelihood (PAML) codeml....68

2.2.12.3. Calculating substitution rates, time since separation and mutation rates 69

2.3. RESULTS 71

2.3.1. Sequence identification and gene structure 71

2.3.1.1. Rice, Sorghum and Arabidopsis endo-(1,4)-β-glucanase gene identification 71

2.3.1.2. Barley sequence identification 71

2.3.1.3. Maize endo-(1,4)-β-glucanase gene sequence identification 81

2.3.2. Endo-(1,4)-β-glucanase amino acid sequence alignment 83

2.3.2.1. Endo-(1,4)-β-glucanase sub-families 83

2.3.2.2. The endo-(1,4)-β-glucanase family tree 85

2.3.3. Intron and exon gene structure 89

2.3.4. In silico mapping of the endo-(1,4)-β-glucanase gene family in maize, sorghum and rice 95

2.3.4.1. In silico mapping of endo-(1,4)-β-glucanase genes in maize 96

2.3.4.2. In silico mapping of rice and sorghum endo-(1,4)-β-glucanase genes and gene synteny with maize 96

2.3.5. Synonymous substitution and evolutionary distance 104

2.3.5.1. Synonymous substitution rates for maize:sorghum orthologues 104

2.3.5.2. Synonymous substitution rates between barley and rice orthologues 105

2.3.5.3. Comparison of K_S of barley and rice with maize and sorghum orthologues 109

2.3.5.4. Synonymous substitution rates between some maize genes 113

2.3.5.5. Mutation rates since separation 113

2.4. DISCUSSION 115

2.5. SUMMARY 124

Chapter 3 127

A Transcriptional Analysis of Endo-(1,4)-β-glucanase Genes in Barley and Maize

3.1. INTRODUCTION 129

3.1.1. Development of the cereal stem 129

3.1.2. Transcriptional analysis....134
3.2. METHODS....139
3.2.1. Barley Base AffyChip data search....139
3.2.2. Maize MPSS database search....139
3.2.3. Plant tissue preparation for QPCR....139
3.2.3.1. Barley plant tissue series....139
3.2.3.2. Barley stem series....140
3.2.3.3. Maize stem series....143
3.2.4. Real-time Quantitative PCR (QPCR)....144
3.2.4.1. Production of cDNA....144
3.2.4.2. Primer design....145
3.2.4.3. QPCR....145
3.2.4.4. Co-transcription analysis....146
3.3. RESULTS....147
3.3.1. Barley....147
3.3.1.1. PLEXdb Affymetrix microchip data....147
3.3.1.2. QPCR transcription across a barley developmental tissue series....147
3.3.1.3. QPCR across a barley stem series....154
3.3.1.4. Correlation of endo-1,4-β-glucanase gene transcription in barley tissues....157
3.3.1.5. Correlation of transcript of the endo-1,4-β-glucanases with other cell wall genes....157
3.3.2. Maize....161
3.3.2.1. Maize MPSS....161
3.3.2.2. QPCR in a maize stem developmental tissue series....165
3.3.2.3. Correlation of gene transcription in maize tissues....173
3.4. DISCUSSION....177
3.5. SUMMARY....191
Chapter 4....195
Heterologous Expression of Endo-(1,4)-β-glucanases in Escherichia coli and Pichia pastoris
4.1. INTRODUCTION....197
4.2. METHODS....205

4.2.1. Expression of endo-(1,4)-β-glucanase genes in *E. coli* 205
4.2.1.1. Expression vectors and E. coli strains 205
4.2.1.2. pET-DEST42..... 205
4.2.1.3. pTOOL7 208
4.2.1.4. BL21 star 208
4.2.1.5. Origami (DE3) 208
4.2.1.6. Origami B (DE3)..... 209
4.2.1.7. Amplification and cloning of gene products..... 209
4.2.1.8. HvCEL1 209
4.2.1.9. HvCEL3 211
4.2.1.10. HvCEL11 211
4.2.1.11. Expression of endo-(1,4)-β-glucanases in E. coli..... 212
4.2.2. Expression of endo-(1,4)-β-glucanases in Pichia pastoris.. 213
4.2.2.1. pPICZα..... 213
4.2.2.2. HvCEL1 214
4.2.2.3. HvCEL3 215
4.2.2.4. Transformation of P. pastoris strain SMD1168H by electroporation..... 218
4.2.2.5. Heterologous expression in P. pastoris..... 220
4.2.2.6. Protein expression in P. pastoris..... 220
4.2.2.7. Sample analysis by dot blot 221
4.2.2.8. Protein purification 222
4.2.3. SDS PAGE 223
4.2.4. Staining gels with Coomassie Blue..... 225
4.2.5. Western Blot analysis 225
4.2.6. Viscometric cellulase assay..... 226
4.3. RESULTS 231
4.3.1. Heterologous expression in *E. coli* 231
4.3.1.1. Expression in the pET-DEST42 expression vector 231
4.3.1.2. Expression in the pTOOL7 vector 231
4.3.1.3. HvCEL1 in pTOOL7..... 234
4.3.1.4. HvCEL1, HvCEL3 and HvCEL11 in pTOOL7 234
4.3.2. Protein expression in *P. pastoris* 237
4.4. DISCUSSION..... 243

4.5. SUMMARY 250

Chapter 5 253

Transformation of Barley with Endo-(1,4)-β-glucanase cDNAs

5.1. INTRODUCTION 255

5.2. METHODS 259

5.2.1. Amplification and cloning into the pMDC32 vector 259

5.2.1.1. Amplification of the barley cDNA sequence 259

5.2.1.2. TA Cloning into the pCR8/GW/TOPO vector 261

5.2.1.3. Transformation of DH5α E. coli cells 263

5.2.1.4. Recombination from pCR8/GW/TOPO into pMDC32 ... 263

5.2.1.5. Transformation of barley embryos with the pMDC32 vector 265

5.2.2. Growth of the transformed plants 266

5.2.3. Genomic DNA preparation 266

5.2.3.1. Phenol/chloroform/iso-amyl alcohol extraction 266

5.2.3.2. Extract-N-Amp 267

5.2.4. Southern Blotting 267

5.2.5. Quantitative Real Time PCR 268

5.2.6. Stem strength analysis 269

5.2.7. Cellulose analysis 271

5.2.8. Microscopy 272

5.2.9. Statistical Analysis 273

5.3. RESULTS 275

5.3.1. Cloning into pMDC32 and transformation into barley 275

5.3.1.1. HvCEL1 275

5.3.1.2. HvCEL5 275

5.3.2. Analysis of T_0 transgenic barley plants 276

5.3.2.1. Phenotype 276

5.3.2.2. Southern Blotting 276

5.3.2.3. Transcript Analysis 276

5.3.2.4. Stem and leaf strength and cellulose analyses 282

5.3.3. Analysis of the *HvCEL1* T_1 barley plants 287

5.3.3.1. Selection of null segregants and QPCR 288

5.3.3.2. Phenotype of the T_1 plants 291

5.3.3.3. Stem strength and cellulose analysis of the T_1 plants .. 298

5.3.3.4. Bright field microscopy of stems of T_1 plants 303

5.3.3.5. Epidermal glue impressions 306

5.4. DISCUSSION 317

5.5. SUMMARY 329

Chapter 6 333
Summary and Future Directions

BIBLIOGRAPHY 343

APPENDIX A 367

APPENDIX B 370

APPENDIX C 373

APPENDIX D 393

APPENDIX E 397

APPENDIX F 400

APPENDIX G 403

APPENDIX H 405

TABLE OF FIGURES

Figure 1-1: Lodging in cereals. 4
Figure 1-2: Chemical composition of cellulose. 6
Figure 1-3: Cellulose synthesis rosettes in the plasma membrane..... 8
Figure 1-4: Proposed mechanism of cellulose production. 8
Figure 1-5: Models of type I and type II cell walls..... 10
Figure 1-6: Different levels of substitution of the glucuronoarabinoxylans. 11
Figure 1-7: Structural chemistry of a small section of a (1,3;1,4)-β-glucan molecule. 11
Figure 1-8: The phylogeny of the CESA and CSL families of genes. 14
Figure 1-9: Evidence of (1,3;1,4)-β-glucans in secondary cell walls in a barley leaf. 15
Figure 1-10: Shows two plant cells with intervening section of the plant cell wall at maturity. 16
Figure 1-11: Cellulose orientation in primary and secondary cell walls. 18
Figure 1-12: Transverse sections of a barley stem..... 20
Figure 1-13: Transverse section of a maize stem. 21
Figure 1-14: Two mechanisms of action for hydrolytic cleavage used by glycosyl hydrolase enzymes. 26
Figure 1-15: Endo-(1,4)-β-glucanases protein structure 27
Figure 1-16: Protein structure of the three GH9 endo-(1,4)-β-glucanase sub-families. 30
Figure 1-17: Phenotype associated with Arabidopsis *kor1-2* T-DNA insertion mutant. 34
Figure 1-18: Phenotype of the rice T-DNA insertion mutant of *OsGlu1*. 34
Figure 2-1: Synteny linkage analysis for cereals. 44
Figure 2-2: 0.7ug 1 kb Plus ladder. 55
Figure 2-3: A diagram of the pGEM-T Easy vector. 58
Figure 2-4: Results of barley sequence identification. 75
Figure 2-5: Barley 3’ SMART RACE. 77
Figure 2-6: Two examples of the use of row and column pooling to locate clones from positive PCR amplification. 80
Figure 2-7: A parsimonious phylogenetic tree of cereals and *Arabidopsis*. 86

Figure 2-8: A parsimonious tree showing maize homoeologues with sorghum orthologues. 88
Figure 2-9: The positions of the intron splice sites on an amino acid alignment of five representative maize endo-(1,4)-β-glucanases. 91
Figure 2-10: Intron/exon structure for GH9A sub-family and clade B1h 93
Figure 2-11: Intron/exon structure for CBD sub-family GH9C and clade B3. 94
Figure 2-12: The approximate chromosome regions of the endo-(1,4)-β-glucanase genes in maize. 99
Figure 2-13: Map positions of maize, rice and sorghum for the endo-(1,4)-β-glucanase orthologues. 101
Figure 2-14: Codon based evolutionary distances for maize and sorghum genes. 106
Figure 2-15: Codon based evolutionary distances for barley and rice genes. 110
Figure 2-16: Orthologous gene pairs for maize:maize, maize:sorghum and barley:rice. 112
Figure 2-17: The ratio of number of synonymous sites to total sites plotted against K_S 112
Figure 3-1: Anatomy and development of a monocotyledon plant stem. 130
Figure 3-2: Maize internode elongation over a 10 day period. 131
Figure 3-3: Flow chart to describe MPSS library production. 135
Figure 3-4: Sloop barley stem internodes with leaves removed to show the stages of development used for the barley stem series. 141
Figure 3-5: A B73 maize variety internode as prepared for the maize stem tissue series. 142
Figure 3-6: PLEXdb Affymetrix transcript data for endo-(1,4)-β-glucanase genes in barley variety Morex. 149
Figure 3-7: QPCR levels of 12 endo-(1,4)-β-glucanase genes across the developmental tissue series. 151
Figure 3-8: Transcript profiles for the endo-(1,4)-β-glucanase genes across the barley tissue developmental series. 152
Figure 3-9: Transcript profiles for endo-(1,4)-β-glucanase genes across the barley tissue developmental series. 153
Figure 3-10: Transcript profiles of endo-(1,4)-β-glucanase genes across a barley stem tissue series. 155

Figure 3-11: Transcript profiles of endo-(1,4)-β-glucanase genes across a barley stem tissue series. 156
Figure 3-12: Correlations between several genes analysed across the barley tissue and stem developmental series. 159
Figure 3-13: Graphs showing co-transcription patterns for the endo-1,4-β-glucanase gene family and other cell wall related families. 163
Figure 3-14: MPSS transcript data for stalk tissues in maize. 168
Figure 3-15: Transcript patterns for *ZmCEL3*, *ZmCEL11 ZmCEL14*. 170
Figure 3-16: QPCR transcripts for *ZmCEL12* in the stem series. 170
Figure 3-17: QPCR transcripts for *ZmCEL1* and *ZmCEL18*. 171
Figure 3-18: Maize stem tissue series QPCR profiles for for *ZmCEL2*, *ZmCEL7*, *ZmCEL10*, *ZmCEL19* and *ZmCEL20*. 172
Figure 3-19: Graphs showing relative transcript levels for maize genes in the stem series. 174
Figure 3-20: Shows a schematic representation of the life of a cell with the maize genes actively transcribed at each stage of development. 189
Figure 4-1: The amino acid alignment of the termite *Nasutitermes takasagoensis* and two barley HvCEL1 constructs. 202
Figure 4-2: The termite *Nasutitermes takasagoensis* GH9 cellulase enzyme. 203
Figure 4-3: The pET-DEST42 expression vector. 206
Figure 4-4: The pTOOL7 expression vector. 207
Figure 4-5: The pPICZαA eukaryote expression vector 216
Figure 4-6: The “transfer sandwich”. 228
Figure 4-7: Ostwald viscometer used to measure relative viscosity of liquids. 229
Figure 4-8: Coomassie stained gel and western blot for HvCEL1 in the pET-DEST42. 232
Figure 4-9: Coomassie stained gels and western blot analysis of HvCEL1 in the pTOOL7 vector in Origami (DE3) cells. 233
Figure 4-10: Coomassie stained gel and Western blot of HvCEL1, 3 and 11 in pTOOL7. 235
Figure 4-11: The Coomassie stain and Western blot of HvCEL11_CBD/pTOOL7 in *E. coli*. 236
Figure 4-12: Coomassie stain and Western blot expression of HvCEL1 and HvCEL3 in *Pichia pastoris* for six days. 239

Figure 4-13: Coomassie stain and western blot for expression of HvCEL1 in *Pichia pastoris* 240
Figure 5-1: The linearised pCR8/GW/TOPO vector prior to cloning. 262
Figure 5-2: The pMDC32 vector. 264
Figure 5-3: Measuring leaf and stem material using an Instron machine. 270
Figure 5-4: Southern blot for *HvCEL1* (G235) T0 transformed barley plants. 277
Figure 5-5: QPCR on *HvCEL1* (G235) transformed barley lines. 279
Figure 5-6: Relative levels of endogenous and transgene transcripts for *HvCEL5* (G234). 281
Figure 5-7: Agarose separation gel following amplification of the hygromycin gene to confirm the presence of the *HvCEL5* (G234) TDNA. 281
Figure 5-8: Combined results of the stem strength and stem cellulose content from *HvCEL1* and *HvCEL5* transgenic plants. 283
Figure 5-9: Correlation between flexural load and both stem dry weight and cellulose levels for *HvCEL1* (G235). 284
Figure 5-10: A graph of the correlation between flexural load and internode diameter. 285
Figure 5-11: The results of the average leaf pulls for HvCEL5 (G234) and HvCEL1 (G235). 286
Figure 5-12: QPCR results for *HvCEL1* T1 plants. 290
Figure 5-13: Photographs of HvCEL1 T1 plants at 2 months. 293
Figure 5-14: Height of the *HvCEL1* T1 plants at 2 months. 294
Figure 5-15: Average leaf height of *HvCEL1* T1 plants at 2 months. 294
Figure 5-16: T1 plants at 5 months. 295
Figure 5-17: Adventitious root production. 296
Figure 5-18: Graphs of HvCEL1 T1 plants showing flexural load to break, % cellulose, mg cellulose per cm stem and stem diameter. 299
Figure 5-19: Correlations between various stem parameters measured on the HvCEL1 T1 transgenic plants. 301
Figure 5-20: *HvCEL1* G235 T1 plants showing cellulose levels and internode length with respect to phenotype. 302
Figure 5-21: Bright field microscopy of Toluidine Blue stained cross-sections of the second internode below the peduncle. 305
Figure 5-22: Toluidine blue staining of a barley stem. 307

Figure 5-23: Toluidine blue staining of stem cross-sections highlighting sclerenchymal tissues in the T1 barley plants. 309
Figure 5-24: Examples of super-glue impressions of the external surface of stem internodes. 310
Figure 5-25: (opposite page) Graphed results of measured parameters for *HvCEL1* T1 plants that were selected for further analysis. 313
Figure 5-26: Relative cellulose and transcript data for the T1 HvCEL1 plants. 321

TABLE OF TABLES

Table 2-1: Orthologous genes between maize, barley, sorghum and rice. 72
Table 2-2: Pooling clones of a barley BAC library.....78
Table 2-3: Sub-families of the endo-(1,4)-β-glucanases in cereals and *Arabidopsis*.....82
Table 2-4: Specific introns associated with maize genes.....92
Table 2-5: Maize endo-(1,4)-β-glucanase gene map information.....97
Table 2-6: Maize endo-(1,4)-β-glucanase genes with the map position of their respective sorghum and rice orthologues.....102
Table 3-1: Barley transcript correlations for some genes in the tissue series and stem series.....160
Table 3-2: Correlation coefficients of the endo-1,4-β-glucanase family with other genes involved in synthesis of cell walls.....164
Table 3-3: MPSS tag sequences and approximate position of tag in the ORF. 166
Table 3-4: MPSS transcript data for maize genes in the 12 core tissues. 167
Table 3-5**:** Nomenclature used for the various tissues and stages of internode development for the maize stem tissue series.....169
Table 3-6: Transcript patterns for endo-(1,4)-β-glucanase genes in barley and maize.....179
Table 3-7: A comparison of transcript between barley and maize orthologues. 185
Table 4-1: Primer sequences used for amplification of cDNAs. 210
Table 4-2: Predicted total length and weight of expressed proteins. 210
Table 4-3: Primers used to amplify sequences for cloning into pPICZα.....216
Table 4-4: Digest rates for NEB and Super Duper buffers. 217
Table 5-1: Matrix for amplification of *HvCEL1* and *HvCEL5* cDNAs.....260
Table 5-2: A summary of the Southern blot analysis for *HvCEL1*.....278
Table 5-3: Primers for QPCR.....279
Table 5-4: Results of QPCR analysis on T_1 generation of the HvCEL1 plants. 289
Table 5-5: A summary table of the results of the analysis of the T1 HvCEL1 plants.....292

Table 5-6: Epidermal cell and internode parameters of selected *HvCEL1* G235 T_1 plants. 311

Table 5-7: A selected sample of *HvCEL1* transgenic plants demonstrating the relationship between measured parameters..... 315

TABLE OF EQUATIONS

Equation 2-1: $p_S = S_d/S$ 68

Equation 2-2: $p_N = N_d/N$ 68

Equation 2-3: $d_S = -3/4\log_e(1-4/3p_S)$ 68

Equation 2-4: $d_N = -3/4\log_e(1-4/3p_N)$. 68

Equation 2-5: $Ks = (d_{ZO} + d_{SO}) / (2 \times 2 \times 50 \times 10^6)$ 70

Equation 2-6: $Ks = (d_{HZ} + d_{SZ}) / (2 \times 2 \times 50 \times 10^6)$ 70

Equation 2-7: $\omega = d_N/d_S$ 70

DECLARATION

I, Margaret Buchanan, certify that this work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signed

ACKNOWLEDGEMENTS

Many people assisted me during the course of this project. Firstly, I must acknowledge my two supervisors, Professor Geoffrey Fincher and Dr Rachel Burton. Their enthusiasm was unequalled and their support well measured. From Geoff I learned about project design and the importance of skilled writing. He also encouraged me to focus on the importance of moving forward in a focussed manner. Rachel had an ever optimistic approach to all aspects of the project and day to day support. She also provided the barley developmental cDNA library used as part of the QPCR analysis.

The project was enabled by a collaboration between the University of Adelaide and the DuPont-Pioneer company. Firstly, my thanks go to Scott Tingey of the DuPont-Pioneer company for making my trips to the USA possible. Also, thankyou to Dr Antoni Rafalski for his help with mapping of the maize genes at the DuPont laboratories in Wilmington, Delaware, as well as his interest in the project and strategic comments. Antoni was also instrumental in arranging sequencing of the barley BAC clones. Dr Kanwarpal Dhugga provided invaluable support during my trip to the Pioneer laboratories in Des Moines, Iowa, where he helped with access to, and analysis of the proprietary maize cDNA library and MPSS transcript data.

The members of the Plant Cell Wall group provided practical support on a day to day basis as well as help with specific aspects of the project. More particularly, Jillian Taylor provided invaluable technical support with regard to measuring stem and leaf strength on the Instron machine. Jill also helped with growing maize in the glasshouse. Dr Filomena Pettolino provided me with help with the cellulose analysis protocol. Natalie Kibble helped with completion of the cellulose analyses of the T1 barley stems. Without Dr Qisen Zhang's help with expression and protein extraction, the heterologous expression Chapter would not have been possible. Also providing support with regard to the heterologous expression was Maria Lombardi. Dr Andrew Harvey provided me with early lessons in the use of BLAST tools for sequence searches, the production of phylogenetic trees and gene structure.

All of the QPCR analyses were performed by Neil Shirley to whom I am very grateful. Neil also provided the data for the co-transcription of the endo-(1,4)-β-glucanases with other cell wall genes.

Thanks must go to Dr Rohan Singh and other members of the tissue culture laboratory at the ACPFG for the barley transformation. Thanks must also be extended to Ursula Langridge for her help with growing the resultant transgenic barley plants in the glasshouse and for providing access to the Sloop barley plants used for the stem developmental series.

Dr Gwenda Mayo was invaluable in her assistance with the microscopy. Thanks are in order for her technical advice, patience and enthusiasm. Invaluable support was also provided by Jodie Kretschmer with regard to the Southern blot analysis. Dr John Toubia pointed me in the right direction with regard to the evolutionary analysis. My thanks go to Dr Bu-Jun Shi for the loan of the BAC membranes and clones. I am also indebted to Dr Delphine Fleury for technical support with regard to screening the BAC membranes.

ABSTRACT

Failure of the stem to support the grain head in cereals is a serious cause of economic loss to farmers worldwide and increasing the cellulose content of the cell wall has the potential to increase stem strength. The glycosyl hydrolase family 9 endo-(1,4)-β-glucanases are believed to have a role in cellulose synthesis. Plants containing mutants of the KORRIGAN-like endo-(1,4)-β-glucanase genes have perturbed cell walls and reduced cellulose. Also growth enhancement occurs in poplar when overexpressing a soluble endo-(1,4)-β-glucanase. However, little is known of the gene family in cereals. Examination of the GH9 endo-(1,4)-β-glucanase gene family in barley and maize was undertaken with exploration of the potential biological correlation between cellulase activity and stem strength.

A detailed phylogeny of the endo-(1,4)-β-glucanase family for barley, maize, sorghum, rice and *Arabidopsis* was performed using sequence data from various sources. Notwithstanding that the maize genome contains 29 endo-(1,4)-β-glucanase genes, including homoeologues and duplicates, the number of genes across the different species only varies between 23 and 25. The phylogenetic tree showed variations in clade structure between the cereals and *Arabidopsis* and indicated some differential gene loss and gain. Where information on map position was available, synteny between the cereals was examined and found to be strong. Along with comparative intron and exon gene structure studies, orthologous genes across barley, maize, sorghum and rice were identified. An evolutionary analysis provided evidence that sorghum was the donor genome in an allotetraploid event in maize that occurred 10-20 million years ago. It was also possible to differentiate between homoeologues resulting from the tetraploidisation of maize and gene duplications.

A transcript analysis for barley and maize was performed and indicated that some members of the endo-β-1,4-D-glucanase gene family are transcribed across a wide range of tissues while other genes are specific to one tissue. Furthermore, there are strong correlations between several members of the endo-(1,4)-β-glucanase family at the transcriptional level. Similar correlations were also shown to exist between members of the endo-(1,4)-β-glucanase family and other genes known to be involved in cell wall synthesis. This data also suggested that evolutionary conservation of transcription exists between orthologues in barley and maize.

Analysis of protein activity in barley was undertaken on plants that had been transformed with endo-(1,4)-β-glucanase genes using the 35s overexpression promoter. Stem strength of these transgenic plants was found to be positively correlated with cellulose levels. Plants of the T_1 generation showed strong phenotypic variation. A number of plants were dwarfed and had much reduced

levels of cellulose. It was speculated that gene inactivation of both transgene and the endogenous gene was producing 'knockdown" plants. On the other hand, some overexpressing plants had higher levels of cellulose than the control plants, however, this was not strongly correlated with stronger stems.

Heterologous expression of barley endo-(1,4)-β-glucanases in microbial systems was attempted. In *E.coli* large amounts of protein were produced, but it was largely insoluble and inactive.

ABBREVIATIONS

AGRF	Australian Genome Research Facility
aa	Amino acid(s)
At	*Arabidopsis*
AX	arabinoxylan
BAC	bacterial artificial chromosome
BAC	bacterial artificial chromosome
BLAST	basic local alignment search tool
BMGY	Buffered glycerol complex medium
BMMY	buffered methanol complex medium
bp	base pairs
CAZy	carbohydrate active enzyme database
CBD	cellulose binding domain
CD	catalytic domain
cDNA	complementary deoxyribonucleic acid
CesA	cellulose synthase
cpm	cycles per minute
Cps	counts per second
Csl	cellulose synthase-like
DAP	days after pollination
DE	developing embryo
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleic acids: deoxyadenosine triphosphate, deoxycytosine triphosphate, deoxythymidine triphosphate and deoxyguanidine triphosphate
ECB	European corn borer
EDTA	ethylenediaminetetraacetic acid
ER	endoplasmic reticulum
ESTs	expressed sequence tags
GAX	galacturonoarabinoxylan
GH9	glycosyl hydrolase family 9
GSP	gene specific primer
GUS	B-glucuronidase
Hv	*Hordeum vulgare*
IPTG	isopropyl-thio-galactopyranoside
Ks	synonymous changes per synonymous site per year
LB	Luria broth
MLG	mixed linked glucans
MPSS	massively parallel signature sequencing
mRNA	messenger ribonucleic acid
mRNA	messenger RNA
Mya	million years ago

NCBI	National Centre for Biotechnology Information
Ni-NTA	Nickel-nitrilotriacetic acid
Ns	non-synonymous changes per synonymous site per year
NUP	nested universal primer
ORF	open reading frame
Os	*Oryza sativa*
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PMSF	phenylmethylsulphonyl fluoride
ppm	parts per million
QPCR	real time quantitative polymerase chain reaction
RE	restriction enzyme
RNA	ribonucleic acid
RT	room temperature
Sb	*Sorghum bicolour*
SDS	sodium dodecyl sulphate
SIGNAL	Salk Institute Genomic analysis Laboratory
SMART RACE	switching mechanism at the 3' end of RNA transcript rapid amplification of cDNA ends
SSC	Sodium chloride/sodium citrate
TAE	Tris-Acetate-EDTA
TBS	Tris buffered saline
TE	Tris-EDTA
TIGR	The Institute of Genomic Research
tRNA	transfer RNA
UDP	uridine di-phosphate
UPM	universal primer mix
UTR	untranslated region
XET/XTH	xyloglucan endotransglycosylase/hydrolase
XG	xyloglucan
Zm	*Zea mays*

Chapter 1

From Stalk Lodging to the Endo-(1,4)-β-glucanase Gene Family in Plants: An Overview

The overall aim of this project was to examine the endo-(1,4)-β-glucanase gene family (glycosyl hydrolase family 9, GH9) in barley and maize and to explore the potential biological correlation between cellulase activity and stem strength in cereals. Failure of the stem to support the head in cereal crops is a serious cause of economic loss to farmers worldwide and current knowledge of the role of plant cellulases in cellulose production and breakdown in the plant will be reviewed. Some reference to cellulases in microorganisms will be necessary, because the properties of cellulases in cereals are not well documented. This introductory chapter begins with a discussion on stalk lodging in cereals. The role of cellulose in the plant cell wall is examined and its function, synthesis and metabolism are referenced to the plant stem where possible.

1.1. LODGING IN CEREAL CROPS

The plant stem is also known as the stalk or culm, and these terms will be used throughout the text. Plant lodging is the permanent failure of the stem to support the plant, and results in the displacement of the head-bearing stem from a perpendicular position (Lang, 1986). Plants may be root lodged or stem lodged. In root lodging, the failure occurs at the root stalk interface, that is, the plant leans from the crown with the roots failing to support the plant. This often occurs after periods of high rainfall. Stem lodging occurs when the stalk fails and usually happens during strong winds and/or when the stalk is weakened by disease (Figure 1-1).

Lodging of cereal crops reduces yield by impeding grain development and making harvesting of grain difficult. Stem lodging in barley (Figure 1-1 A) is reported to cause reductions in yield of between 13% and 60%, depending on the cultivar, with losses greater if lodging occurs either at heading or at the soft dough stage of grain development (Jedel, 1991; Jezowski, 2003). Reductions in stem diameter, particularly of the middle internodes, has been noted as an important precursor to lodging in barley (Berry et al., 2006). On the other hand, in maize, stalk lodging (Figure 1-1 B) usually occurs in the middle

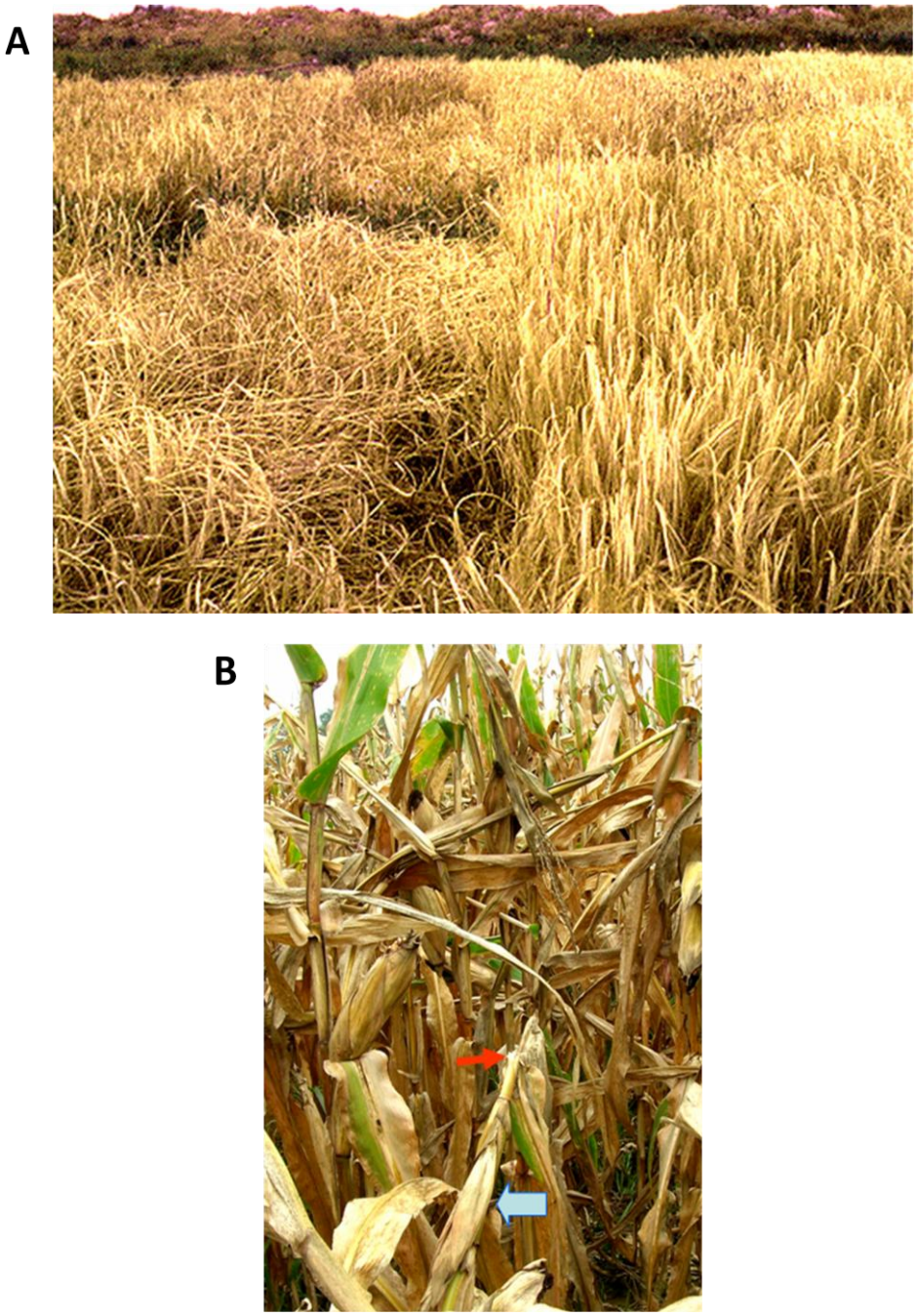


Figure 1-1: Lodging in cereals.

A: lodged experimental crop of barley showing varietal differences in resistance to lodging. Left side lodged, right side lodging resistant (Akar et al., 2007). B: Stem lodging in a crop of maize. Red arrow indicates where stem has broken at the internode immediately below the ear (blue arrow).

of the third internode below the ear, and is reported to cause annual yield losses of between 5% and 35% (Zuber, 1978; Esechie et al., 2004).

From the 1930s to the 1970s harvested yields of maize have increased in the order of 60-90% in the U.S. corn belt (Tollenaar, 1989). This is in part because of the "Green Revolution", that has resulted in dramatic increases in yield due to the application of increased fertiliser and irrigation, and the introduction of new, higher yielding cereal varieties. In maize, higher yielding hybrids also have increased biomass. Therefore harvest index, which is a measure of grain as a proportion of total plant biomass, is in the order of 60%, has remained unchanged. In rice and wheat, the yield increases were enabled with the help of dwarfing genes that interfere with the plant hormone gibberellic acid production (Peng et al., 1999; Hedden, 2003).

Maximum yields in maize are obtained by increasing planting density and growing maize hybrids that produce only one ear (Duvick, 1999). Higher planting density increases root competition for nutrients, and leaf competition for sunlight, thus placing a limit on the plants ability to produce sugars and may lead to an energy stress (Duvick, 1999). If the plant is stressed, the increase in translocation of materials from the stem to the ear during grain development may cause parenchymal breakdown of the stalk, thus weakening it and increasing the likelihood of lodging (Liebhardt et al., 1968). Furthermore, production of only one ear makes the plant asymmetrically top-heavy. Despite stem strength improving with the introduction of new hybrids, increased yield per stalk and increased planting density have correlated with increased stem lodging (Jampatong, 2000; Echezone, 2007). The result is a "Catch 22" situation, where attempts to increase stem strength may reduce crop yields because the plant stalk and grain are competing for limited available carbon (Jampatong, 2000; Martin, 2004). This competition for a limited supply of energy secondarily reduces stem strength because it makes the plant more susceptible to fungal diseases such

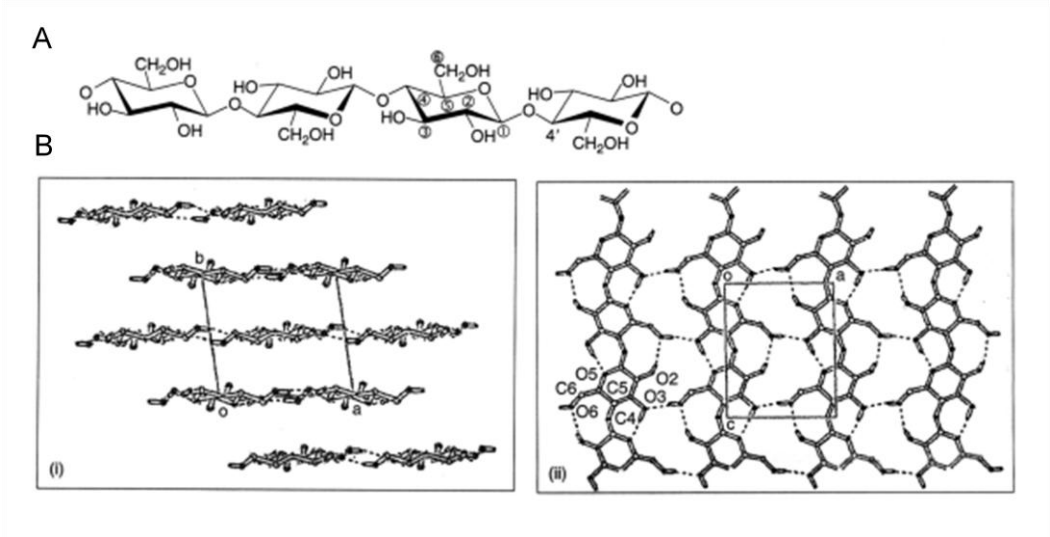


Figure 1-2: Chemical composition of cellulose.

A. illustrates 1,4-β-linked D-glucopyranosal residues orientated at 180° to each other, with C atoms numbered I to 6, B. i.) cross section of microfibril looking at cut surface. b-o indicates van der Waal's forces binding the layers together. and ii) looking at the top of a single layer of a cellulose microfibril. Dotted lines indicate H-bonds between and within glucose molecules producing flat layers. (Taken from Fincher and Stone, 2004)

as stem rot, and insect pests such as European corn borer (ECB). The challenge, therefore, is to produce a plant with a stronger stem that does not have a concomitant reduction in yield. The strength of the stem is correlated with the composition of the cell wall, which is a complex structure with cellulose as its major component.

1.2. CELLULOSE

Cellulose is the most abundant macromolecule on Earth, with an estimated 10^{11} tonnes produced annually (Brown, 2004). It is largely a product of higher plants, but is also produced by some algae, fungi and bacteria. In all cases it is contained in the walls surrounding individual cells. The structure of cellulose can be seen in Figure 1-2. Each cellulose molecule is a long chain polysaccharide of (1,4)-β-linked glucopyranosyl residues and it is this (1,4)-β linkage that results in each residue being oriented at 180° to adjacent residues. This produces a long ribbon-like structure. Hydrogen bonding within and between the molecules allows formation of flat sheets and Van der Waal's forces acting between the sheets produce a stable hydrophobic crystalline microfibril (Fincher and Stone, 2004).

Cellulose molecules contain approximately 8000 glucan residues in primary cell walls and up to 15,000 residues in secondary cell walls (Brown, 2004). In plants, cellulose is produced by cellulose synthase (CESA) enzymes that are located in the plasma membrane (Figure 1-3). It is postulated from electron microscopic evidence and early cellulose microfibril analysis that CESA proteins are organized into complexes of 36 molecules (Delmer, 1999; Doblin, 2002). Electron microscopy also suggests that CESA proteins form a hexamer, arranged with five other hexamers that combine into the cellulose producing module called a rosette, or terminal complex (Figure 1-4). UDP-glucose residues provide the substrate for the CESA enzymes, which are believed to extrude 36 individual (1,4)-β-glucopyranosyl (cellulose) chains into the cell wall, where they crystallize into a cellulose microfibril (Salnikov, 2001).

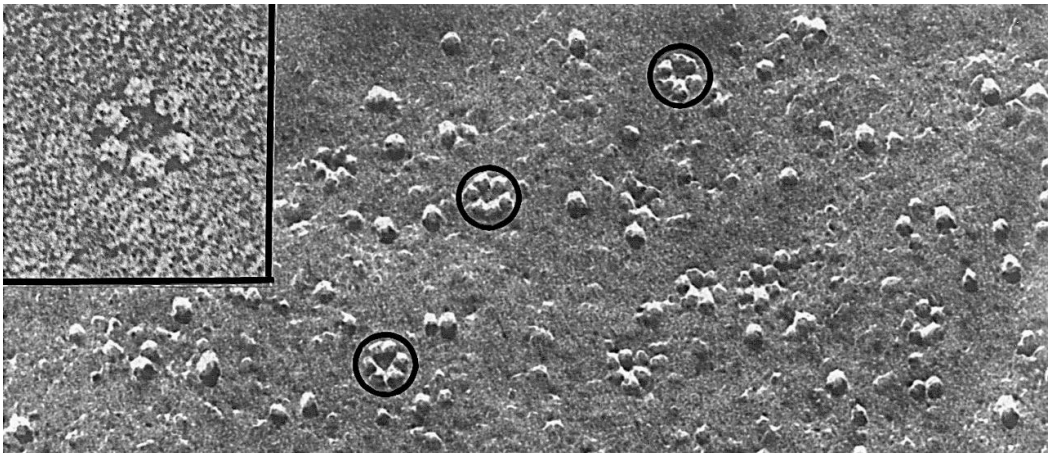


Figure 1-3: Cellulose synthesis rosettes in the plasma membrane.

Freeze fracture replica of rosettes (circled) in the plasma membrane of a *Zinnia elegans* tracheary element at 222,000x magnification. Inset of one rosette is at 504,545x magnification (Taken from Delmer (1999), Ann. Rev. Plant Physiol. Plant Mol. Biol., courtesy Mark J. Grimson and Candace H. Haigler, Department of Biological Sciences, Texas Tech University, Lubbock, Texas)

NOTE:
This figure/table/image has been removed to comply with copyright regulations.
It is included in the print copy of the thesis held by the University of Adelaide Library.

Figure 1-4: Proposed mechanism of cellulose production.

The cellulose synthesis rosette is situated in the plasma membrane. A single 1,4-β-glucan strand is produced by each CESA molecule so a total of 36 strands are extruded into the cell wall space from the rosette, with H-bonding enabling the formation of cellulose microfibrils (Somerville et al., 2004)

Recent studies into the genetics of cellulose synthesis indicate that plants may have up to 13 *CESA* gene paralogues and that three of these genes are co-ordinately transcribed to produce the complex. In *Arabidopsis*, 10 *CESA* genes have been identified. Immunoprecipitation studies show that in *Arabidopsis*, AtCESA4, AtCESA7 and AtCESA8 are co-ordinately expressed during secondary cell wall production (Taylor et al., 2003). In maize, 13 *CESA* genes have been found (Appenzeller et al., 2004) and massively parallel signature sequencing (MPSS) analysis indicates that *ZmCESA10, ZmCESA11* and *ZmCESA12* are co-ordinately transcribed during secondary cell wall production (Appenzeller et al., 2004). Eight *CESA* genes have been identified in barley (Burton et al., 2004). Quantitative real time polymerase chain reaction (QPCR) analysis shows that *HvCESA1, HvCESA2* and *HvCESA6* are co-transcribed in barley in tissues undergoing primary cell wall production such as root tip and coleoptile, whereas *HvCESA4, HvCESA7* and *HvCESA8* are transcribed almost exclusively in maturing root and stem, which suggests that they are chiefly involved in secondary cell wall production (Burton et al., 2004). That in each species examined, three *CESA* genes or proteins are being co-ordinately transcribed provides strong evidence for the proposal that three specific CESA enzymes are required for the formation of a functional cellulose producing enzyme complex. Studies involving *CesA* mutant genes in *Arabidopsis* show that loss of activity of a single CESA subunit will prevent rosette formation and increase production of non-crystalline cellulose (Arioli et al., 1998). Plants carrying mutations of *AtCesA4* produce a phenotype with irregular xylem cell walls, which collapse because the walls of these cells have a crystalline cellulose content of only 30% of the wildtype xylem cell walls (Taylor et al., 2003). Such mutations illustrate the importance of cellulose as a structural component of the cell wall, where it combines and interacts with other key non-cellulosic poly-saccharides to form a reinforced matrix, as summarised below.

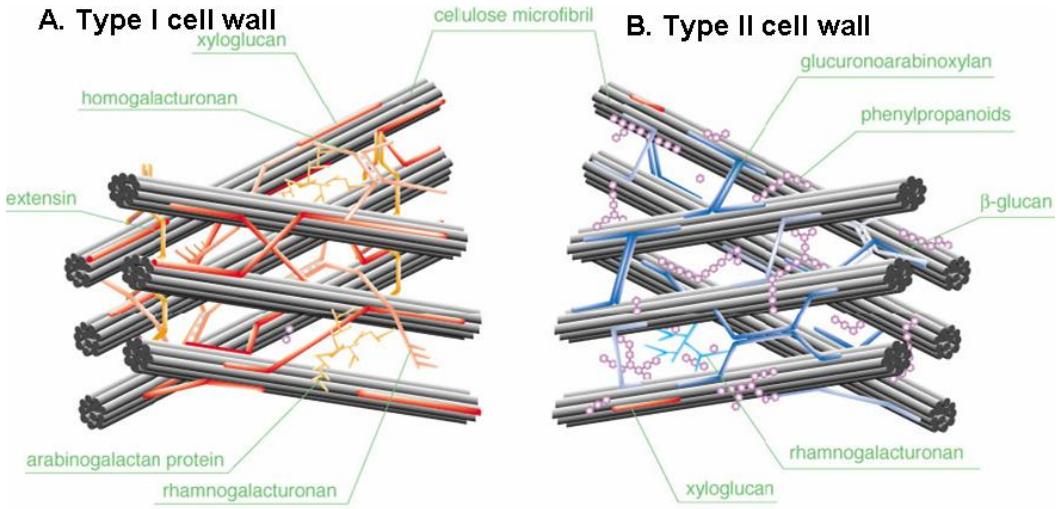


Figure 1-5: Models of type I and type II cell walls

These diagrams provide two models at each end of a cell wall spectrum as described in the text and illustrate some differences between them. Note the close association of xyloglucan (red) (Type I) and glucuronoarabinoxylan (dark blue) (Type II) to the cellulose microfibrils (grey) KEY: grey: cellulose microfibrils, red: xyloglucan, blue: glucuronoarabinoxylans, purple circles: phenylpropanoids (lignin), pale blue: β-glucans, yellow: structural proteins, aqua or yellow: rhamnogalacturonan (pectin). (Yokoyama and Nishitani, 2004)

NOTE:
This figure/table/image has been removed
to comply with copyright regulations.
It is included in the print copy of the thesis
held by the University of Adelaide Library.

Figure 1-6: Different levels of substitution of the glucuronoarabinoxylans.

Glucuronoarabinoxylans (GAX) consist of (1,4)-β-xylanose chains with variable numbers of arabinosyl or glucosyl groups attached. A shows a higher level of substitution than B. Less substitution is seen in secondary cell walls and this enables the GAX molecules to H-bond more efficiently to the cellulose microfibrils and to each other. In cereal grains arabinosyl substituents predominate (Taken from Verbruggen et al., 1998)

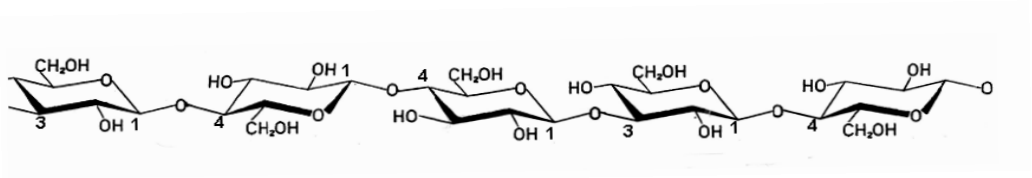


Figure 1-7: Structural chemistry of a small section of a (1,3;1,4)-β-glucan molecule.

It is a glucopyranosyl chain similar to cellulose, but with the mixed C linkages. Numbers indicate 1→3 or 1→4 carbon links. The 1→3 confers a kink in the structure and these linkages occur at irregular intervals. (Taken from Fincher and Stone, 2004).

1.3. THE PLANT CELL WALL

The plant cell wall is a chemically complex structure composed principally of a crystalline cellulose framework, embedded in a matrix that is composed predominantly of non-cellulosic polysaccharides, but also contains proteins and lignin. The proportions of each polysaccharide vary from plant to plant, between tissues and according to stage of maturation (Lerouxel et al., 2002; Somerville et al., 2004; Yokoyama and Nishitani, 2004). For convenience, plant cell walls have been divided into Type I and Type II cell walls that provide the two ends of a spectrum of cell wall composition (Figure 1-5) (Carpita and Gibeaut, 1993). In reality, however, a continuum exists between the two extremes. In dicots, cell walls comprise mainly xyloglucans (XG), pectins (heteroglycans with diverse side branches) and lignins as their non-cellulosic components (Carpita and McCann, 2002). Present in lesser amounts are arabinoxylans, glucomannans and galactoglucomannans.

The main non-cellulosic components of the cell walls of the Poaceae, which are typically found in cereals and grasses, are heteroxylans, principally glucuronoarabinoxylans (GAX) (Figure 1-6) (Fincher and Stone, 2004). Cell walls of cereals also contain the (1,3;1,4)-β-glucans (Figure 1-7). Compared to dicots, smaller amounts of pectins, lignins and XGs are present in the Poaceae (Fincher and Stone, 2004). Not surprisingly, it is estimated that some 15% of the plant genome is devoted to cell wall production and maintenance (Carpita et al., 2001a). Although not a great deal is known of these genes, their function or control, some knowledge is emerging.

As mentioned previously, the process of cell wall synthesis and assembly involves the extrusion of cellulose into the cell wall by the cellulose synthase rosettes, which are located in the plasma membrane. Cellulose forms the cell wall "scaffold". On the other hand, non-cellulosic components have been shown to be produced in the endoplasmic reticulum and Golgi apparatus, from where they are transported to the plasma membrane in vesicles, and subsequently secreted into the cell wall (Dhugga et al., 2004; Liepman et al., 2005). Once in

the cell wall, XG and GAX can form non-covalent interactions with each other. They also interact with, and form a sheath over, the cellulose microfibrils (Carpita et al., 2001b). As well as the formation of non-covalent bonds, non-cellulosic polysaccharides might also be linked through covalent bonding (Thompson and Fry, 2001; Hrmova et al., 2007). It has been suggested that (1,3;1,4)-β-glucans and XGs may be covalently linked with each other and with cellulose by transglycosylation, although *in vivo* evidence for such covalent cross-links has not yet been provided (Hrmova et al., 2007). The enzymes implicated in these possible cell wall modification activities are the xyloglucan endotransglycosylases/hydrolases (XET/XEH), members of the glycoside hydrolase family 16 (http://www.cazy.org/fam/acc_GH.html). Another group of enzymes important in the production of non-cellulosic polysaccharides are the polysaccharaide synthases. This group of enzymes includes the CESA enzymes, mentioned earlier, but also the cellulose synthase-like (CSL) family. The phylogenetic relationship between the different members of the polysaccharide synthases is presented in Figure 1-8.

The CSL enzymes are involved in the synthesis of some of the non-cellulosic polysaccharide components of the cell wall and are slowly being identified. The CSLs have been named from A to J (there is no CSLI) (Richmond and Somerville, 2001; Hazen et al., 2002; Farrokhi et al., 2006). Two of these sub-families, CSLF and H, are known to play a role in (1,3;1,4)-β-glucan production in barley (Burton et al., 2006; Doblin et al., 2009). Transcript data in barley, together with heterologous expression in *Arabidopsis* show that HvCSLF is responsible for (1,3;1,4)-β-glucan production in primary cell walls (Burton et al., 2006). More recently, transcript evidence indicates that the *HvCSLH1* genes are actively transcribed in tissues producing secondary cell walls (Doblin et al., 2009). Furthermore, (1,3;1,4)-β-glucans have been found in secondary tissues of barley (Figure 1-9) (Burton and Fincher, 2009), and transgenic expression of HvCSLH1 in *Arabidopsis* produced (1,3;1,4)-β-glucan (Doblin et al., 2009).

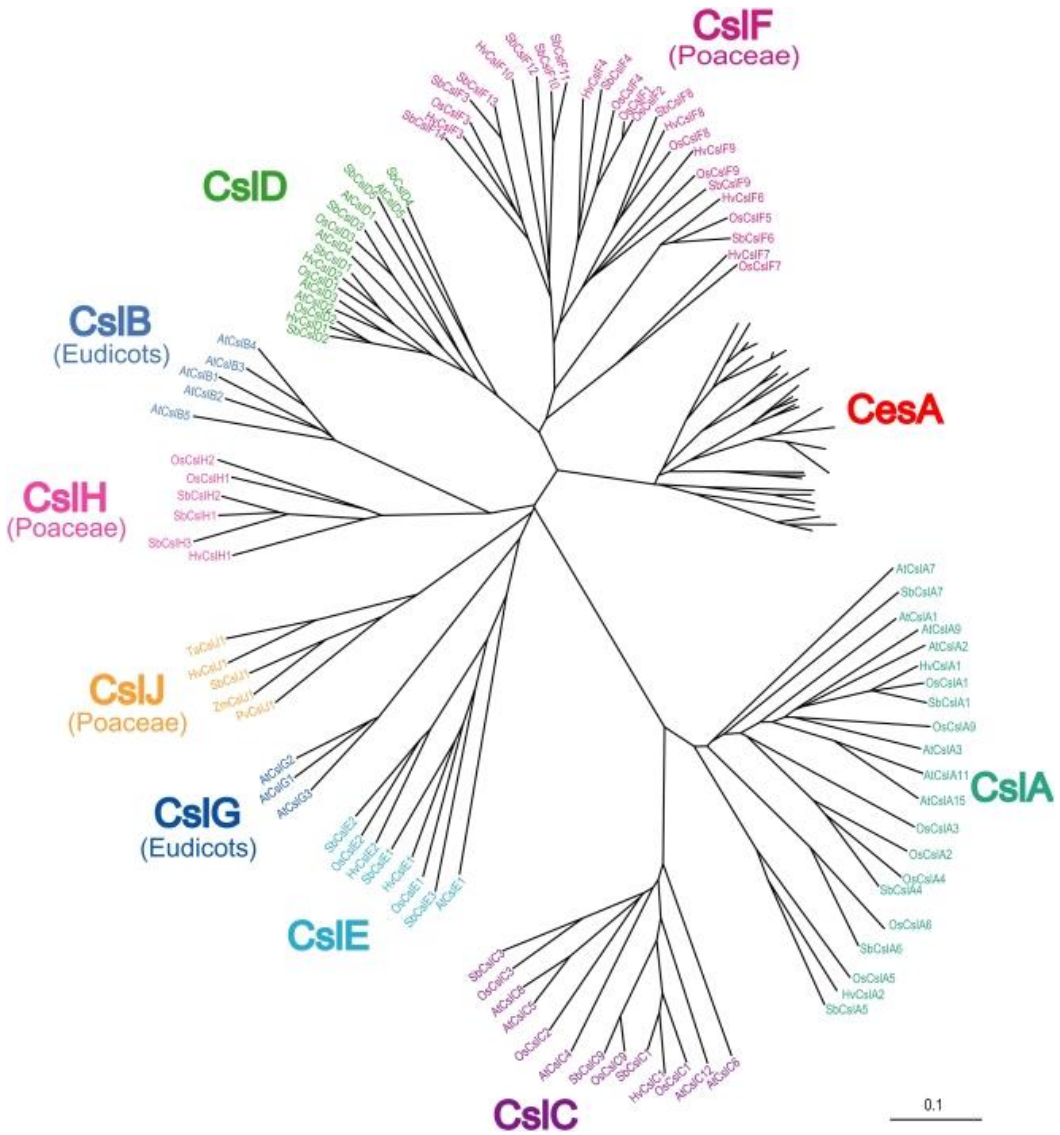


Figure 1-8: The phylogeny of the CESA and CSL families of genes.

This phylogenetic tree, presented as a cladogram, illustrates the relationship between the cellulose synthase and cellulose synthase-like gene families. It should be noted that this tree indicates that CslF, J and H contain no dicot species and have so far only been found in the Poaceae and CslB and G are in dicots but have not been found in the Poaceae (Fincher, 2009).

NOTE:
This figure/table/image has been removed
to comply with copyright regulations.
It is included in the print copy of the thesis
held by the University of Adelaide Library.

Figure 1-9: Evidence of (1,3;1,4)-β-glucans in secondary cell walls in a barley leaf.

Thin Section of a Young Barley Leaf probed with the BG1 Monoclonal Antibody and Detected with Goat Anti-Mouse IgG Alexa Fluor® 488. The presence of (1,3;1,4)-β-glucan in secondary cell walls of the vasculature and other cells can be seen. The image was taken with a Leica AS LMD Laser Dissection Microscope with DFC 480 camera using fluorescence filter I3 (excitation filter 450–490nm BP, barrier filter 515nm LP) and shows significant labeling in the vascular bundle (vb), the sclerenchyma fibres (sf), and the guard cells (gc), with lighter labeling of the epidermal cells (e) and no labeling of the mesophyll cells (m) (Taken from Burton and Fincher, 2009).

NOTE:
This figure/table/image has been removed
to comply with copyright regulations.
It is included in the print copy of the thesis
held by the University of Adelaide Library.

Figure 1-10: Shows two plant cells with intervening section of the plant cell wall at maturity.

After cytokinesis the middle lamella forms and is used as a scaffold for deposition of the primary cell wall. This is followed by deposition of several secondary cell wall layers. Each successive layer of wall is deposited on the cytosolic side of the previous layer. (http://219.221.200.61/ywwy/zbsw(E)/pic/ech4-20.jpg).

Coordination of gene transcription and expression for the synthesis of the cell wall components is important because the mature cell wall is produced sequentially as the cells divide, differentiate, grow and mature (Dickison, 2000). Several layers can sometimes be detected and these include the middle lamellar, the primary cell wall and the secondary cell wall (Figure 1-10).

1.3.1. Primary cell wall

The primary cell wall provides early support for the growing plant but is also designed to accommodate the processes of cell division and elongation. At cytokinesis (or cell division), Golgi-derived non-cellulosic polysaccharides are deposited at the cell plate, where they form the middle lamella (Dickison, 2000). The primary cell wall is subsequently deposited between the middle lamella and the plasma membrane. In primary cell walls, the cellulosic component is irregularly arranged in cellulose microfibrillar strands (Figure 1-11 A). These strands may be associated with XG, (1,3;1,4)-β-glucans and GAX in cell walls (Carpita and Gibeaut, 1993; Yokoyama and Nishitani, 2004). Growing structures such as meristematic regions, coleoptiles, cotyledons, the base of growing leaves and growing root tips comprise cells with primary walls.

During cell elongation the deposited primary wall must be sufficiently flexible and extensible to allow for increases in cell length, while still maintaining structural support and resistance to turgor pressure.

1.3.2. Secondary cell walls

Once a cell has stopped elongating, secondary cell wall deposition occurs between the plasma membrane and the primary cell wall. The secondary cell wall provides superior structural support for the plant and is usually many times thicker than the primary cell wall (Harche et al., 1990; Jung and Casler, 2006). Secondary deposition in cell walls is usually accompanied by an increase in the amount of cellulose in the cell wall and by the deposition of lignin throughout the wall (Jung and Casler, 2006). The architecture of the secondary wall is quite

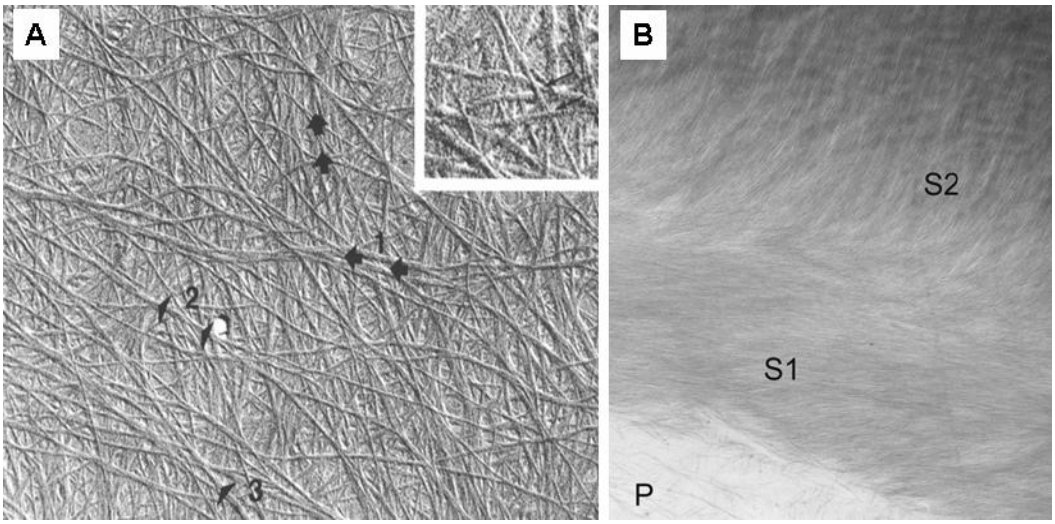


Figure 1-11: Cellulose orientation in primary and secondary cell walls.

A: primary cell wall microfibril orientations in the *Petunia hybrida* meristematic cell wall as indicated by the arrows. The difference in predominant orientations of the cellulose microfibrils is approximately 90°. Inset: arrow shows interwoven microfibrils (Wolters-Arts et al., 1993). B: Transmission electron micrograph showing secondary cell wall microfibril orientation in *Pinus radiata* tracheary elements: P (primary cell wall), S1 and S2 secondary cell wall depositions. Note approximately 90° difference in orientation of S1 to S2 and increased density of cellulose microfibrils compared to P. (Taken from Donaldson and Xu, 2005.)

different to that of the primary wall (Figure 1-11 B). Cellulose deposition in secondary walls is more ordered and usually consists of up to three main layers, (Donaldson and Xu, 2005). Within each layer the cellulose microfibrils are arranged parallel to each other, but each layer is oriented at a different angle. The S2 layer may be very thick compared to S1 and S3. In its role in providing support, the secondary cell wall maintains the shape of the plant cell, and hence the shape of, and support for, the plant more generally.

1.4. THE MONOCOTYLEDON STEM

The role of the stem in plants is to support the leaves, flowers and fruits, to act as a conduit for the transport of water, solutes and photosynthates around the plant and for storage (Dickison, 2000). The tissues of the cereal stem include the epidermis, vascular bundles, rind and pith (Figure 1-12 and Figure 1-13). On the surface of the plant is the epidermis, which provides an impervious, protective outer layer. Within the epidermis are pores called stomata, whose opening and closing are strictly controlled to allow water and gas exchange between the plant and the atmosphere (Esau, 1977). Situated beneath the epidermis is the collenchyma layer, which is usually several cells thick. Collenchyma cells are elongated (up to 2mm) and overlapping, but retain an active protoplast that allows them to respond to any need for modification of the cell wall. The walls of collenchyma cells are markedly thickened, contain high levels of cellulose and may (Hatfield et al., 1999) or may not (Esau, 1977; Dickison, 2000) contain lignin. In barley, there is a correlation ($r = 0.82$) between thickness of the walls of the collenchyma cells and strength of the stems (Baker et al., 1990). In maize, this tissue is referred to as the rind and contributes approximately 20% of the cross-sectional stem area and 80% of cell wall material (Kawamura et al., 2002; Jung and Casler, 2006). In sorghum, the rind and the outer vascular bundles together comprise approximately 15% of the cross-sectional stem area and 60% of the total cell wall material (Hatfield et al., 1999) and again stem strength is positively correlated with rind thickness (Prest et al., 1983). It has been

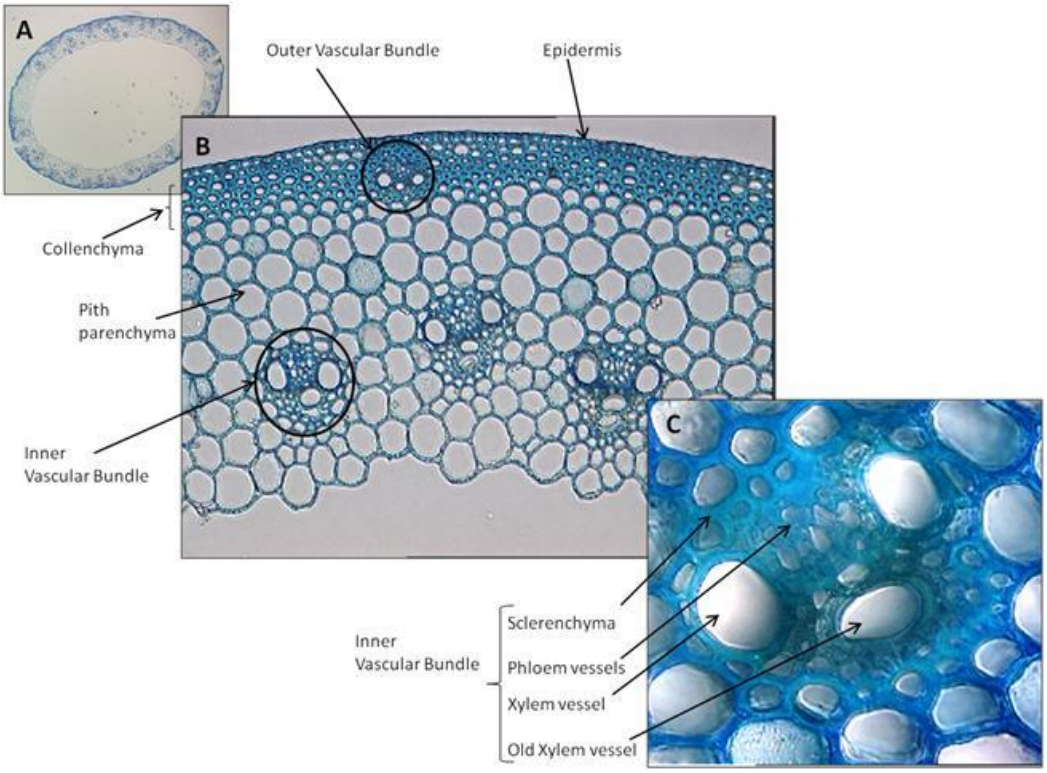


Figure 1-12: Transverse sections of a barley stem

A: illustrates the hollow nature of the barley stem (X40 magnification) (Image from Dr Rachel Burton, Australian Centre for Plant Functional Genomics), B: X200 magnification to illustrate the different tissues of a barley stem, C: X630 magnification illustrating the different cell types that comprise the vascular bundles. All sections are stained with Toluidine Blue.

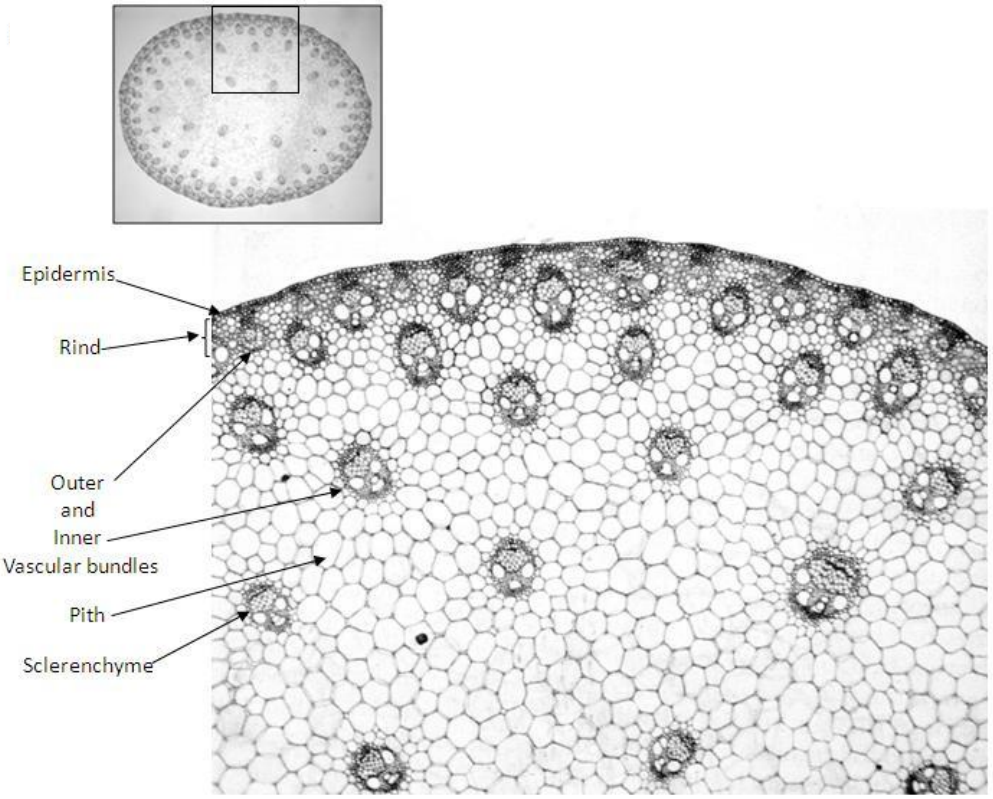


Figure 1-13: Transverse section of a maize stem.

Magnified cross sections of maize stem illustrating the distribution of the different tissues. Inset shows a cross section of the whole stem. Note the solid pith filled centre compared with the hollow stem of barley. (Esau, 1977)

suggested that increasing the number and density of vascular bundles in the sorghum stem reduces stalk lodging (Das et al., 2008). The pith, which is situated central to the collenchyma, is the most abundant tissue in the stem. In maize it comprises approximately 80% of the stem cross-sectional area but contributes less than 20% of the total cell wall material (Jung and Casler, 2006). The majority of the cells of the pith are parenchyma cells. Characterised by their relatively thin cell walls and hexagonal shape, they are important for storage of carbon-based energy, primarily as starch (Esau, 1977; Dickison, 2000).

Sclerenchyma cells may be scattered throughout the stem and can contain high levels of cellulose (Esau, 1977; Dickison, 2000). Two sclerenchyma cell shapes have been identified, fibres and sclereids. Fibres are elongated and up to 7cm long, while sclereids, which are rare in monocotyledons, exhibit quite variable cell shapes. Walls of sclerenchyma cells are variably thin or thickened, and in monocotyledons may be heavily lignified. At maturity, the cytoplasm is lost, as the secondary cell walls thicken and cells become non-living. Sclerenchyma cells provide an important mechanical support role for the stem and the fibres are important in paper making and for hemp and sisal production (Esau, 1977). In monocotyledons, fibres are evident immediately beneath the epidermis and are often seen around, or on either side of, the vascular bundles.

The vascular bundle comprises the xylem, phloem and associated cells (Esau, 1977; Dickison, 2000). A transverse section of a barley vascular bundle can be found in Figure 1-12 B. The xylem is responsible for transport of water and solutes from the roots to the rest of the plant. The tracheary vessels of the xylem are highly lignified, providing water impermeability for the tracheids and, more generally, rigid support for the stem. The phloem transports photosynthates and other organic solutes around the plant. The basic element of the phloem is the sieve tube, which, like xylem vessels, has associated sclerenchyma cells. In grasses, the vascular bundles of the stem are discrete elements. The barley stem is hollow, with vascular bundles arranged concentrically around the stem wall (Figure 1-12). One set of smaller vascular bundles is associated with the collenchyma and one or two sets of larger vascular bundles are located within the

pith. Maize, like barley, has concentrically arranged vascular bundles associated with the collenchyma or rind. However, unlike barley, maize stems are not hollow but are filled, principally, with pith cells and have large numbers of vascular bundles arranged concentrically throughout (Figure 1-13). Thus, the stem comprises a number of cell types that are important for protection, storage, transport and strength. Each cell type has a wall with a composition designed not only to support those specific functions but also to provide structural support for the plant more generally.

1.5. CELLULOSE AND STEM STRENGTH

The major provider of strength in the stem was historically thought to be lignin. Deposition of lignin commences at the time of secondary cell wall production and lignin can eventually spread throughout both primary and secondary walls (Boudet, 2000). Lignins are complex aromatic polymers that can form bridges between cellulose microfibrils and the rest of the cell wall matrix. They confer rigidity to the plant, decrease wall permeability and are thought to be involved in plant defence. Reduced lignin levels may induce xylem collapse (Grabber, 2005). Although lignin is a major provider of stem stability, recent evidence suggests that cellulose may be as important, if not more important, for the provision of stem strength. Stem strength in maize has been shown to be directly related to cellulose content of the stem (Appenzeller et al., 2004). There is a strong correlation (r^2 = 0.85) between internodal flexural stem strength and cellulose content of the internode. This, together with the observation that cellulose makes up approximately 50% of the total dry matter of the stem cell wall, suggests that increasing cellulose concentration in the cell wall may improve the mechanical strength of the stem (Appenzeller et al., 2004).

Additional evidence supporting the view that cellulose is a prime provider of stem strength comes from experiments in which tomato plants were subjected to periodic gyratory shaking (Heuchert et al., 1983). The stems of the shaken plants contained 15% more cellulose than control plants; importantly, other cell wall components such as non-cellulosic polysaccharides and lignin content remained

unchanged. Stems of the shaken plants had increased resistance to shear forces, increased rigidity and elasticity and reduced brittleness, compared with control plants. Further support for the importance of cellulose is found in maize data from the USA, where European corn borer (ECB) infestation of the maize stalk is a common cause of lodging (Martin, 2004). Cellulose content of the rind in maize is an important factor in resistance to ECB as shown by the significant positive correlation between rind penetrance, as measured by a rind penetrometer, stem strength, cellulose content of the stem and resistance to ECB. Unfortunately, when positively selecting for stem strength and increased cellulose content of the stems, there is a negative correlation with grain yield (Martin, 2004). Conversely, selecting against stem strength, in favour of yield, results in an even greater reduction in grain yield due to increased lodging and insect damage.

In any case, cellulose is an important determinant of stem strength and if it can be increased there may be significant improvement in crop yields by reducing lodging. However, manipulating the *CESA* family of genes to increase cellulose content of the cell wall may be difficult due to the complexity of expression and the formation of the multi-enzyme terminal complex. Although a seemingly contradictory principle, a possible alternative for increasing cellulose in the plant is through manipulation of the plant endo-(1,4)-β-glucanase gene family.

1.6. ENDO-(1,4)-B-GLUCANASES

Endo-(1,4)-β-glucanases, or cellulases, are documented on the Carbohydrate Active Enzyme database (CAZy) (http://www.cazy.org/fam/acc_GH.html) and are classified as glycoside hydrolases (GH). GH enzymes are involved in hydrolysis of O-glycosyl linkages between sugar residues of polysaccharides and have been classified into 107 families on the basis of their sequence and primary structures (Henrissat, 1991; Henrissat and Bairoch, 1993). GHs are also identified by the Enzyme Commission (EC) nomenclature EC3.2.1.x, where x denotes the enzyme substrate. Enzymes that hydrolyse cellulose are classified as EC3.2.1.4 and are found in at least 14 GH families.

The mechanism of action of glycoside hydrolases is one of general acid catalysis (Davies and Henrissat, 1995), whereby a pair of aspartate and/or glutamate residues act as the catalytic acid/base and the catalytic nucleophile layer of non-cellulosic XG polysaccharides and possibly lignin (Bauer et al., 1973; Hayashi and Maclachlan, 1984; Hayashi et al., 1987; Juy et al., 1992). The hydrolytic cleavage reactions produce either an inversion or retention of the original glucosidic linkage configuration (Figure 1-14) (Koshland Jr, 1953; Withers et al., 1986; Gebler et al., 1992; Davies and Henrissat, 1995).

A large body of work exists on cellulases of bacterial and fungal organisms. This reveals that cellulases can be divided into exo-glucanases and endo-(1,4)-β-glucanases on the basis of whether they hydrolyse cellulose from the non-reducing ends (exo) or from within the molecule (endo), respectively (Davies and Henrissat, 1995). The substrate binding site of endo-(1,4)-β-glucanases is an open cleft on the surface of the enzyme that binds the cellulose chain and hydrolyses internal glycosidic linkages, essentially at random (Figure 1-15 A). The active site of classical exo-glucanases, also known as cellobiohydrolases, on the other hand, is contained within a tunnel that runs through the enzyme (Figure 1-15 B). One end of the cellulose molecule binds to the enzyme, and is fed through the molecular "tunnel". As the enzyme progresses along the chain it removes cellobiosyl units from the non-reducing terminus of the polysaccharide.

Whatever the mode of hydrolysis, enzymatic degradation of the cellulose crystalline microfibril is problematic because enzymes have difficulty gaining access to the individual chains, due to the tightly bound nature of the crystalline microfibril structure. The microfibrils might also be protected by, and H-bonded to, a layer of non-cellulosic xyloglucan polysaccharides and possibly lignin (Bauer et al., 1973; Hayashi and Maclachlan, 1984).

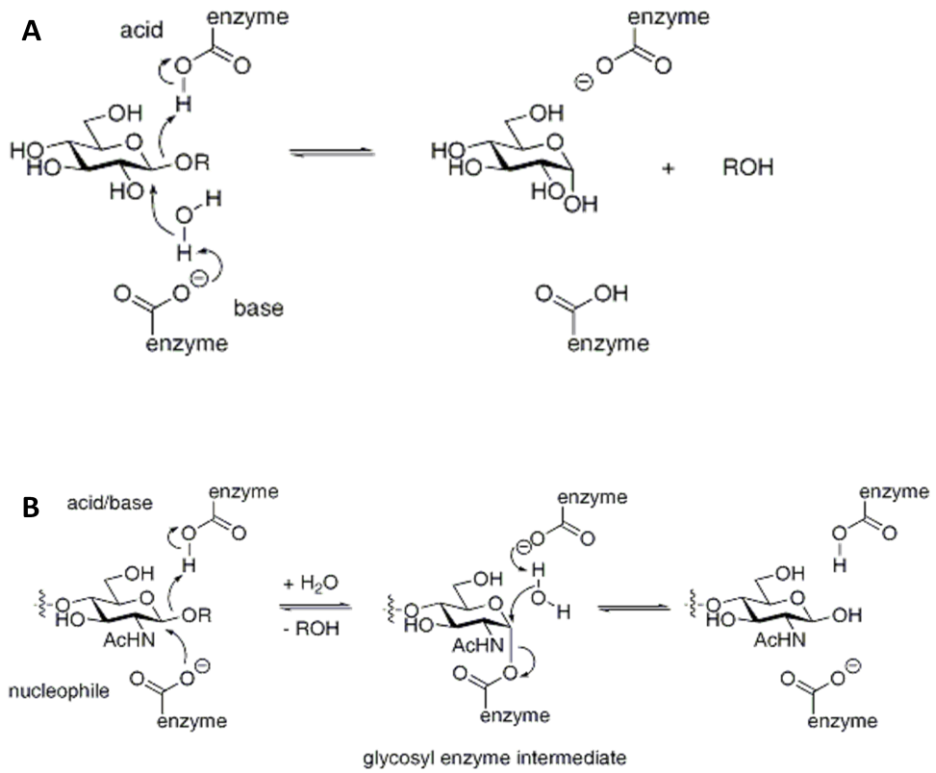


Figure 1-14: Two mechanisms of action for hydrolytic cleavage used by glycosyl hydrolase enzymes.

A: the process of inversion and **B**: retention of the original configuration (Williams, 2006) (http://en.wikipedia.org/wiki/Glycoside_hydrolase).

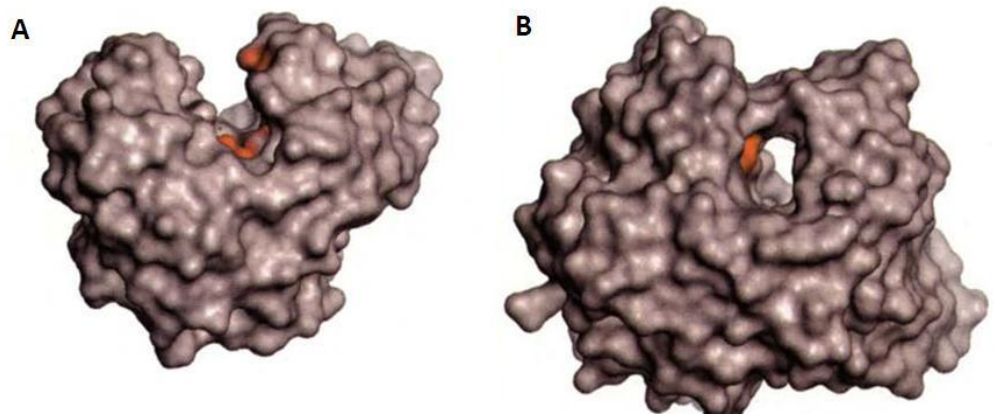


Figure 1-15: Endo-(1,4)-β-glucanases protein structure

Models of A: an endoglucanase protein illustrating the open enzymatic cleft and B: an exoglucanase with a closed tunnel through which the polysaccharide chain progresses. The red dots indicate the active catalytic residues (Davies and Henrissat, 1995).

Notwithstanding the problem of access, many bacteria and fungi, which rely heavily on cellulose for their energy requirements, have evolved very efficient mechanisms for the hydrolysis of cellulose. Electron microscopic examination of crystalline cellulose hydrolysis by *Trichoderma reesei* cellulase complexes revealed that the cellulose microfibrils were initially split longitudinally into bundles, which became thinner and finally disappeared within 20-30 minutes (White and Brown, 1981). Attempts to hydrolyse the cellulose using individual cellulases were generally unsuccessful, indicating that digestion of cellulose microfibrils, in this case, required the action of a suite of enzymes. Thus as part of their cellulose hydrolytic machinery, microorganisms have evolved cellulolytic multienzyme complexes known as cellulosomes (Levy et al., 2002). Cellulosomes contain scaffolds that support associated cellulase hydrolysing enzymes and cellulose binding domains (CBD). CBDs have aromatic residues that bind the glucosyl residues in cellulose and allow for more intimate and stable contact with the microfibril (Mattinen et al., 1997). CBDs may be non-covalently attached to the enzymes or may be covalently connected to the catalytic domain by linker sequences (Solingen et al., 2001). Loss of the CBD considerably reduces enzyme activity and so CBDs are considered essential for the digestion of crystalline cellulose, however the truncated enzymes are still usually able to hydrolyse soluble cellulose (Gilkes et al., 1991; Juy et al., 1992; Gilad et al., 2003). The role and mechanism of action of the CBD is uncertain; it may act as a "grip" on the insoluble microfibril and thus disrupt its crystalline structure to improve access to individual chains, and/or it may stabilise the interaction of the catalytic domain of the enzyme with the substrate (Linder and Teeri, 1997; Davies et al., 2005).

Thus, bacteria and fungi are well supplied with cellulose hydrolytic mechanisms, where they are essential for the survival of many organisms. However, in plants, where cellulose is of great structural importance, and inadvertent or widespread hydrolysis could prove catastrophic, how and where are exo- and endo-(1,4)-β-glucanases utilised?

1.6.1. Plant endo- (1,4)-β-glucanases

Based on sequence analysis, plant endo-(1,4)-β-glucanases are categorized as members of the glycoside hydrolase family 9 (GH9) group of genes (http://www.cazy.org/fam/acc_GH.html). Experimental data suggest that the GH9 endo-(1,4)-β-glucanases hydrolyse using the inverting mechanism (http://www.cazy.org/CAZY/; Withers et al., 1986; Gilkes et al., 1991; Gebler et al., 1992). Plant GH9 endo-(1,4)-β-glucanases have been divided into three sub-families, namely GH9A, GH9B and GH9C, on the basis of protein structure (Figure 1-16) (Urbanowicz et al., 2007a). The GH9A sub-family are membrane bound enzymes with a short NH_2-terminal cytoplasmic domain, a putative single spanning transmembrane domain and an extracellular COOH-terminal catalytic domain (CD) (Brummell et al., 1997). The putative transmembrane domain is identified by a hydrophobic region of 39 amino acids flanked on each side by four or five charged amino acids (Boyd and Beckwith, 1990; Brummell et al., 1997). A proline-rich region situated at the end of the COOH-terminal, suggests the possibility of protein-protein interactions (Kay et al., 2000). The GH9B sub-family are soluble enzymes with a CD and an NH_2-terminal signal peptide that may target the protein to the cell wall, via the Golgi apparatus or endoplasmic reticulum. The enzymes of the GH9C sub-family are also soluble endo-(1,4)-β-glucanases with an NH_2-terminal signal peptide and CD as described for the GH9B sub-family, but they also have an elongated COOH-terminal sequence containing a putative CBD.

1.6.2. Endo-(1,4)-β-glucanases and CBDs

Hydrolysis of crystalline cellulose by bacterial endo-(1,4)-β-glucanases requires a suite of enzymes and CBDs. Until recently it was thought that plant endo-(1,4)-β-glucanases did not contain active CBDs, but evidence to the contrary has now emerged. Early work in ripening strawberries revealed an active endo-(1,4)-β-glucanase with a CBD that showed an expression pattern that overlapped temporally and spatially with an endo-(1,4)-β-glucanase lacking a CBD

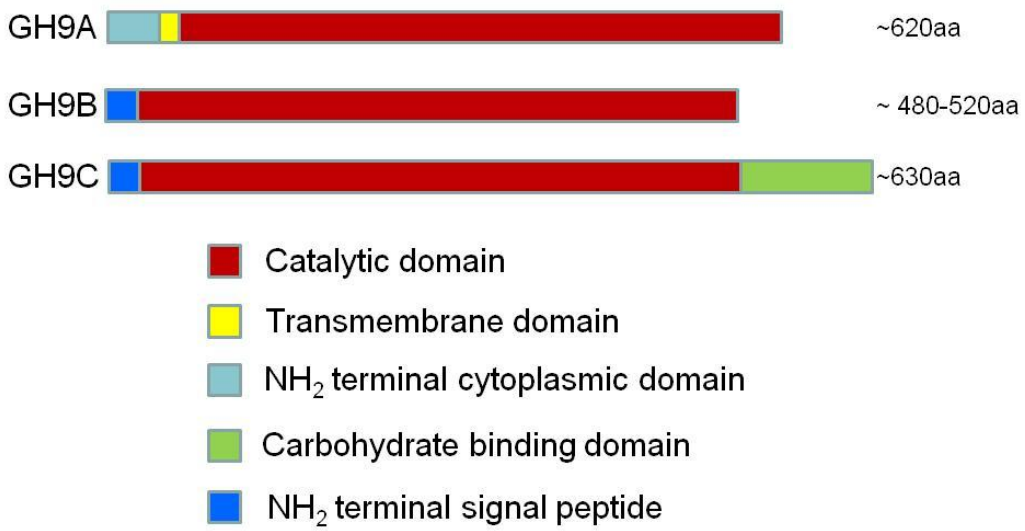


Figure 1-16: Protein structure of the three GH9 endo-(1,4)-β-glucanase sub-families.

GH9A has a cytoplasmic domain, transmembrane domain and catalytic domain; GH9B contains a catalytic domain and NH2-terminal signal peptide; GH9C has a catalytic domain, NH2-terminal signal peptide and a COOH-terminal carbohydrate binding domain.

(Trainotti et al., 1999). Recent whole genome sequencing of *Arabidopsis* and rice indicate that both species have at least three GH9C genes; each contain regions encoding putative CBDs (http://www.cazy.org/fam/acc_GH.html). Isolation of intact protein (Yoshida et al., 2006) or cloning and expressing these full length genes (Urbanowicz et al., 2007b) has proved problematic, because proteolysis occurs in the amino acid linker sequence between the CBD and CD. However, recent efforts to hydrolyse crystalline cellulose, using a fusion protein comprising a CBD from *Solanum lycopersicum* (tomato) and a bacterial catalytic domain from *Thermobifida fusca*, were rewarded with hydrolytic activity of the enzyme against bacterial microcrystalline cellulose and the partially acid hydrolysed Avicel (Urbanowicz et al., 2007b). Activity of the fusion protein, however, was less than that for the intact *T. fusca* endo-(1,4)-β-glucanase.

1.6.3. *In vitro* substrates

The different endo-(1,4)-β-glucanase sub-families with various structures suggest that these enzymes may be active against a variety of substrates. *In vitro* activity of several endo-(1,4)-β-glucanase enzymes shows hydrolysis of celluloses of varying crystallinity, length and methoxycarbon substitution levels, as well as lichenan (a (1,3;1,4)-β-glucan with a higher ratio of 1,3 linkages found in lower plants such as hornworts), (1,3;1,4)-β-glucans, arabinoxylans, glucomannans and XG (Molhoj et al., 2001; Hayashi et al., 2005; Yoshida and Komae, 2006; Urbanowicz et al., 2007b). Most of the endo-(1,4)-β-glucanases studied are active on more than one substrate. Activity against a variety of non-cellulosic polysaccharides would indicate that these enzymes could potentially have a role in cell wall matrix modification and breakdown.

1.6.4. Endo-(1,4)-β-glucanases in cell wall breakdown

In plants, endo-(1,4)-β-glucanases are implicated in the breakdown of cell walls, with ample reference in the literature to these enzymes being expressed during fruit and leaf abscission, and senescence. Several endo-(1,4)-β-glucanases have been identified in ripening peach, pepper and strawberry, but their role is unclear. Their presence is associated with cell wall breakdown but acid-stable

(crystalline) cellulose concentrations are unchanged after fruit ripening and it is suggested that they might be involved in wall loosening by hydrolysis of non-cellulosic glucans, such as XGs, within the matrix. On the other hand, endo-(1,4)-β-glucanases were found in early grape berry formation but disappeared as ripening progressed (Nunan et al., 2001). Endo-(1,4)-β-glucanases have not been found to play a role in ripening of tomato fruits but are present at the fruit abscission zone (Brummell et al., 1999). Abscission of the leaves of peppers (Ferrarese et al., 1996) and oranges (Burns et al., 1998) has also been found to require an endo-(1,4)-β-glucanase. In soybean, endo-(1,4)-β-glucanases are associated not only with leaf, flower and pod abscission zones but also with growing roots and shoots (Kemmerer and Tucker, 1994). Endo-(1,4)-β-glucanase expression is also observed to increase as the anther develops (Sexton et al., 1990). In *Arabidopsis*, an endo-(1,4)-β-glucanase associated with root cap sloughing has been identified and loss of activity results in retarded root growth (del Campillo et al., 2004). Although endo-(1,4)-β-glucanases are clearly implicated in the breakdown of cell walls, they also have a potential role in cellulose synthesis during plant growth.

1.6.5. Endo-(1,4)-β-glucanases in plant cell wall synthesis

In *Arabidopsis*, rice, tomato and *Populus tremuloides*, endo-(1,4)-β-glucanase gene orthologues are known to affect plant phenotype and cellulose content in the cell wall (Brummell et al., 1997; Zuo et al., 2000; Shani et al., 2004; Bhandari et al., 2006; Shani et al., 2006; Zhou et al., 2006). An endo-(1,4)-β-glucanase, commonly known as KORRIGAN, has been characterised and shown to have an important role in the production of cellulose in the cell wall of plants. Several *korrigan* mutants have been reported in the literature (Nicol et al., 1998; Zuo et al., 2000; His et al., 2001; Lane et al., 2001; Sato et al., 2001; Szyjanowicz et al., 2004; Zhou et al., 2006). These reports indicate that endo-(1,4)-β-glucanase genes play a role in cell wall formation and cellulose production and are discussed in the following section.

A T-DNA insertion mutation in the promoter region of the *Arabidopsis* gene *AtKOR*, designated *Atkor1-2*, produced cells with incomplete cell walls that were multinucleated and misshapen, indicating incomplete cytokinesis (Figure 1-17) (Zuo et al., 2000). The AtKOR1-GFP fusion protein located to the cell plate during cytokinesis. These plants also showed a dwarf phenotype with radial root swelling and swollen epidermal cells. Embryonic development was not perturbed, as effects of the mutant were confined to vegetative development.

Nicol et al. (1998), using a different mutant allele of the same *AtKOR* gene, established that KORRIGAN was important in cell elongation and cellulose synthesis in primary cell walls. This group noted that the cells of the mutant were enlarged and walls were significantly thicker with loss of wall architecture compared with the wildtype. The mutant also exhibited wall separation between cells. Using a heat inducible *AtKor1* mutant, *acw1*, Sato et al. (2001) reported similar findings.

Further mutational studies reveal that *AtKOR* is probably involved in both primary and secondary cell wall cellulose production (Lane et al., 2001; Szyjanowicz et al., 2004), with wall crystalline cellulose levels in the mutant plants up to 30% lower than those of the wildtype, but with increased levels of pectin and soluble cellulose (His et al., 2001; Sato et al., 2001). This observation is central in implicating endo-(1,4)-β-glucanases as participants in cell wall biosynthesis.

The rice endo-(1,4)-β-glucanase, OsGLU1, is a transmembrane endo-(1,4)-β-glucanase with 75% sequence similarity to AtKOR (Zhou et al., 2006). A T-DNA insertion in the second intron of the *OsGlu1* gene produced a dwarf rice plant with shortened leaves and internode length (Figure 1-18). Microscopically, the cells of the stem were shorter and narrower than those of the wildtype as were the cells in the leaf tips, indicating that this enzyme may play a role in cell elongation. However the morphology of the cells appeared normal. Analysis of

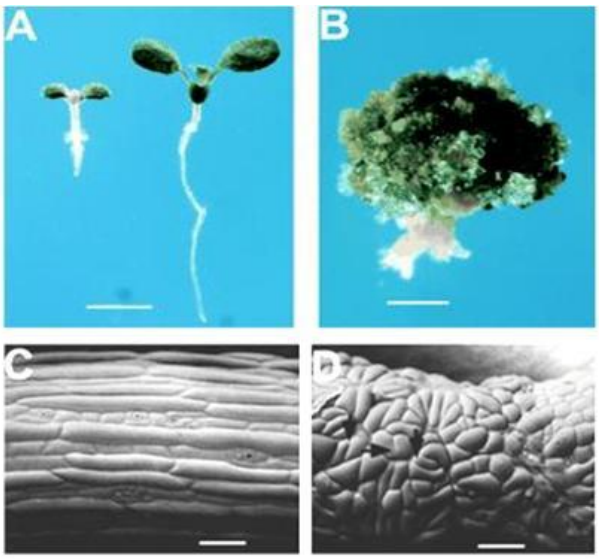


Figure 1-17: Phenotype associated with Arabidopsis *kor1-2* T-DNA insertion mutant.

A: cotyledon, left: mutant, right: wild type. B: phenotype of *kor1-2* plant. C and D: electron micrograph of epidermal cells: left: wild type, right: mutant. In D, arrows point to incomplete cell walls. Taken from (Zuo et al., 2000)

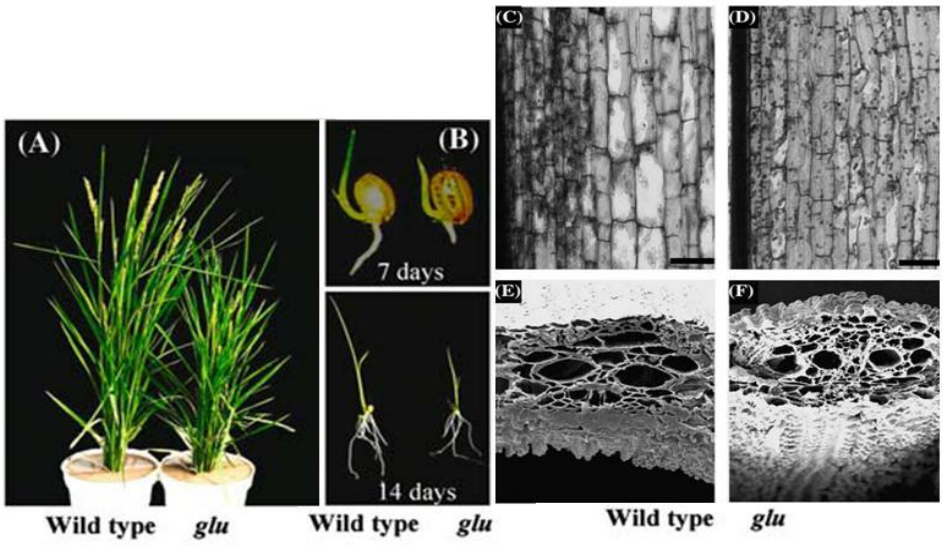


Figure 1-18: Phenotype of the rice T-DNA insertion mutant of *OsGlu1.*

Wild type on the left and mutant on the right. A: rice plants, B: germinating seedlings, C and D: endothelial cells and E and F: cross section of vascular bundles. In each case the *Osglu* insertional mutant is smaller, but otherwise appears to have normal development (Zhou et al., 2006). This contrasts with the *Arabidopsis* insertional mutant in Zuo et al., (2000), where cells have incomplete cell walls.

the cell wall revealed that cellulose levels were reduced by 10% and pectin was doubled.

Endo-(1,4)-β-glucanases and cellulose synthesis

In aspen (*Populus tremuloides*), *PtrKOR*, a putative orthologue of *AtKOR*, is transcribed in the phloem and xylem of developing internodes, and in the xylem of more mature internodes (Bhandari et al., 2006). In these experiments, an aspen sapling was bent for 7 days to produce a tension stress on the upper side of the stem. Tension wood is produced on the upperside of branches. It resists the forces of gravity and produces wood with much higher levels of cellulose than the underside or compression wood. Both *PtrKOR* mRNA and PtrKOR enzyme were significantly increased in the xylem cells of the tension or upperside wood, and xylem tissues had increased cellulose in the cell walls, compared to the underside or compression wood (Bhandari et al., 2006). Moreover, *PtrKOR* was shown to be co-transcribing with three the cellulose synthase genes, *PtrCESA1*, *PtrCESA2* and *PtrCESA3*, reported to be associated with cellulose deposition in secondary cell walls (Joshi et al., 2004).

That transmembrane endo-(1,4)-β-glucanases may be specific to cellulose synthesis rather than being involved in hydrolysis of non-cellulosic polysaccharides is implied where an endo-(1,4)-β-glucanase of *Brassica napis* was modified by truncation of the first 90 amino acids, which contain the transmembrane helix and cytosolic domains (Molhoj et al., 2001). This construct was cloned and protein heterologously expressed in *Pichia pastoris*. Importantly, and unlike other endo- (1,4)-β-glucanases, this enzyme did not hydrolyse XG, (1,3;1,4)-β-glucans, or AX, but only exhibited *in vitro* activity against cellulose derivatives that had low substitution levels, such as carboxymethyl cellulose 4 and 8 (CMC4, CMC8), or displayed low crystallinity, such as acid swollen cellulose and colloidal Avicel. This lent support to the possibility that the GH9A subfamily is not involved in cell wall breakdown of non-cellulosic polysaccharides but, rather, has a role in cellulose synthesis.

The actual role of KOR in cellulose synthesis is not yet known, although several hypotheses have been developed. The ends of the COOH-terminal domain of the sub-family GH9A enzymes contain a proline rich region (Williamson, 1994; Kay et al., 2000; Master et al., 2004). Potentially, this region may aid in subcellular localisation of the enzyme or association with either a carbohydrate moiety or protein (Master et al., 2004). Accumulation of lipid-linked (1,4)-β-glucans in cells of the *Arabidopsis AtKor* mutants (Sato et al., 2001) suggests that sitosterol-β-glucoside (SG) may act as a primer in cellulose synthesis (Peng et al., 2002) and that KOR may cleave the SG primer from the elongating cellulose molecule. Sterol biosynthesis mutants have reduced cellulose content (Schrick et al., 2004) providing further evidence that sterols may be part of the cellulose synthesis machinery. Further analysis of an endo-(1,4)-β-glucanase mutant line in *Arabidopsis* did not show differences in the quantity of SG and sitosterol-cellodextrins (SCD) compared to wild-type suggesting another role for KOR in cellulose production (Robert et al., 2004). Another hypothesis is that KOR may cleave the growing cellulose chains as they crystallise into microfibrils and thus control the length of the chains and/or total length of the microfibril or facilitate crystallisation in some other way (Brummell et al., 1997; Molhoj et al., 2002). It is also suggested that KOR1 may play a role in removing inactive CESA complexes or malformed cellulose microfibrils (Robert et al., 2004).

1.7. SUMMARY

In cereals, failure of the stem to support the head results in stem lodging and is an important cause of economic loss to farmers. Increasing stem strength, however, may result in yield penalties. The strength of the stem is reliant on the plant cell wall, a diverse and complex structure. The wall contains cellulose as a main component and exists in plants in the form of hydrophobic crystalline microfibrils. It is clear that increasing the cellulose content of the cell wall has the potential to increase stem strength, however, CESA enzymes, the primary components of cellulose synthesis, form complexes that may make manipulation of their expression problematic. Although the GH9 endo-(1,4)-β-glucanases are involved in cell wall breakdown, they are also believed to be involved in wall

formation and cellulose synthesis. This is illustrated by the role of the membrane associated KORRIGAN-like enzymes, mutants of which have perturbed walls and reduced cell wall cellulose. Furthermore, KORRIGAN expression increases when cellulose synthesis is increased.

Members of the endo-(1,4)-β-glucanase enzyme family are active *in vitro* against a wide range of substrates, including cellulose, xyloglucans, (1,3;1,4)-β-glucans and other substituted and branched 1,4-β-glucan chains, making them putative candidates for hydrolysing non-cellulosic constituents of the wall. While some of these genes produce soluble proteins, and some have CBDs, others are putatively located in the plasma membrane and it is these family members that have provided the main evidence for the role of endo-(1,4)-β-glucanases in cellulose production. However, little is known of their activity in cereals or their potential role in cellulose synthesis.

In the initial stages of this project, available sequence information was gathered from various databases for barley and maize endo-(1,4)-β-glucanase gene information and expanded on by contiguous assembly. This work is described in Chapter 2 and includes sequence information further developed in barley using amplification of cDNA ends (RACE) technology and bacterial artificial chromosome (BAC) sequencing. Further analyses included gene structure, phylogeny and evolutionary development of the endo-(1,4)-β-glucanase gene family in cereals. Chapter 3 includes a transcriptional analysis of some individual endo-(1,4)-β-glucanase genes in barley and maize in a variety of tissues with the focus on stem tissues. In Chapter 4, the results of recombinant expression in heterologous protein expression systems are described. Finally, the results of transformation of endo-(1,4)-β-glucanase genes into barley are contained in Chapter 5. Here an overexpression vector containing an endo-(1,4)-β-glucanase cDNA was inserted into barley and the resultant transgenic plants analysed for stem strength and cellulose content. Currently, our understanding of endo-(1,4)-β-glucanases in cereals is, at best, minimal, and this project was designed to at least partly redress this situation.

Chapter 2

The Endo-(1,4)-β-glucanase Gene Families in Barley, Maize and other Cereals

2.1. INTRODUCTION

In the introductory chapter, the endo-(1,4)-β-glucanases were described as enzymes that hydrolyse cell wall components. They have a diverse range of activities as noted in *Arabidopsis*, poplar and rice. However, current knowledge is restricted to just a few members of endo-(1,4)-β-glucanase families in a few species, little is known of the other genes in the family and information regarding the cereals, in particular, is lacking. The first part of this project was designed to explore the diversity of the GH9 family of endo-(1,4)-β-glucanase genes across several cereal species, and to further our current knowledge of the phylogeny and evolution of this gene family in cereals, with the focus on barley and maize.

Duplication, allotetraploidy and diploidisation of cereals

A whole genome duplication event estimated to have occurred approximately 90 million years ago (mya), preceded divergence of the cereals from other plant species (Paterson et al., 2004; Salse et al., 2008; Paterson et al., 2009). The earliest fossil evidence of the Poaceae is from the Paleocene, 65 – 23 mya (Linder, 1987; Crepet and Feldman, 1991) . Lack of fossil records prevents a more specific time for their emergence from the dicotyleons. However, the patterns of spread of cereals across the globe suggest that their evolution preceded complete separation of the mega-continents Panagaea and Gondwana and the evolution of hypsodont mammals. This places their origin at more than 65 mya. These authors suggest that the evolution and diversification of the Poaceae occurred as a result of an increase in global aridity and accounts for the lack of fossil records associated with their evolution. This time also coincides with the global catastrophe associated with the extinction of dinosaurs (Lambert, 2007). It is estimated that at around 50 mya, the antecedents of cereals such as Pooideae (barley and wheat) and Oryzoideae (rice) separated from the Panicoideae (maize and sorghum) (Kellogg, 1998; Salse et al., 2008). The Oryzoideae and Pooideae separated from each other soon after, at an estimated 46 mya. The ancestors of maize and sorghum separated from each other at

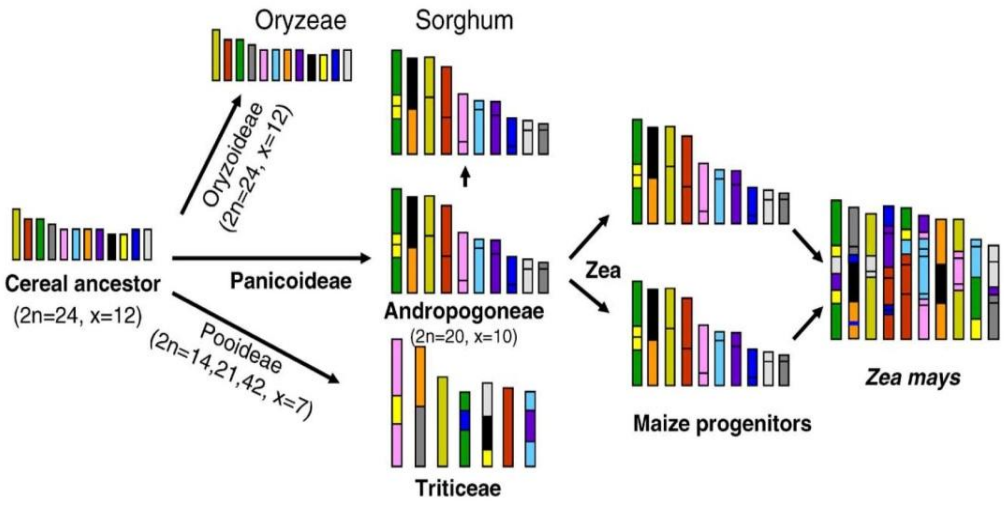


Figure 2-1: Synteny linkage analysis for cereals.

Using the rice colour coded syntenic blocks, progenitor chromosomes have been assembled based on comparisons of synteny linkage analysis blocks for maize, sorghum, rice and wheat. The cereal ancestor of the Oryzoideae, Panicoideae and the Pooideae had 12 chromosomes, but over time this number has been reduced to 10 in the Andropogoneae, which includes maize, sorghum and switchgrass, and seven in the Pooideae (barley and wheat). The maize genome increased in size after tetraploidy and rediploidisation, but chromosome number remained the same (Copied from Wei et al., 2007).

between 11.9 and 20.4 mya, but at an estimated 10 to 4.8 mya, the maize genome underwent a doubling of chromosome number by introduction of a foreign genome, or allotetraploid event followed by non-homologous recombination of the genome and subsequent reversion to the diploid state (Gaut and Doebley, 1997; Paterson et al., 2004; Swigoňová et al., 2004). It is hypothesised that a sorghum ancestor provided the introduced genome, though this is yet to be substantiated and recent sequencing of the sorghum genome may provide the required evidence (Paterson et al., 2009).

Since allotetraploidisation, the maize genome has experienced gene losses of at least 50% from one or other of the homoeologues, as well as substantial genome rearrangements that have comprised translocations and/or segmental inversions (Wei et al., 2007). For cereals in general, the original ancestral genome is believed to have comprised n = 5 chromosomes (Salse et al., 2008). At the time of separation of the ancestors of the Oryzoideae and Pooideae from the Panicoideae the genome comprised n = 12 (Figure 2-1). Further chromosomal rearrangements within each family since this time have produced chromosomal configurations as follows; Oryzoideae: n = 12, Pooideae: n = 7 and Panicoideae: n = 10 (Salse et al., 2008). All genomes have also experienced varying degrees of insertion of retrotransposable elements. However, despite the passage of over 70 million years and large discrepancies in genome size between these species; barley: 5,300Mb (International Barley Genome Sequencing Consortium), rice: 466.3Mb (Yu et al., 2005), maize: 2,700Mb (Wei et al., 2007), sorghum: 730Mb (Paterson et al., 2009); gene numbers and their order along segments of the chromosome have been more or less conserved (Paterson et al., 2004).

Synteny between cereals

Gene synteny was originally defined by John H. Renwick in 1971 and referred to gene loci on the same chromosome (Passarge et al., 1999). More recently, the term has come to mean the preserved order of genes along the chromosome between related species and is often used as a means of identifying orthologues between species (Heger and Ponting, 2007). Identification of gene synteny

between species by the use of linkage block analysis is a technique employed before significant genome sequence was available (Moore et al., 1995; Sankoff et al., 2007). From a rice linkage map these authors produced a single theoretical ancestral chromosome containing a series of linkage segments. Using scant sequence data, this linkage map was extrapolated from rice to maize, sorghum and wheat (Moore et al., 1995). Recent completion of sequencing of the rice and sorghum genomes, and near completion of the maize genome has provided more accurate and detailed synteny information (Figure 2-1) (Wei et al., 2007).

Whole genome sequencing

In recent years, advances in DNA sequencing techniques have resulted in the completion of many genome sequences. In plants, genomes that are completely sequenced, or close to completion, and are available for public use, include *Arabidopsis*, rice, poplar, sorghum, *Brachypodium distachyon* and maize. The International Rice Genome Sequencing Consortium published the completed genome assembly of the rice cultivar Nipponbare in 2005 (www.gramene.org). More recently, the *Sorghum bicolour*, cultivar Moench, assembly was published (Paterson et al., 2009). Sequencing of a maize B73 bacterial artificial chromosome (BAC) library, although described as incomplete with many BAC sequences still not fully assembled, has progressed to the point where most BACs are anchored to a physical map of the chromosomes and a large amount of sequence is available for public use (www.maizesequence.org). Barley genome sequencing is organised through the International Barley Genome Sequencing Consortium and is due for completion in 2012. The Triticeae, including wheat and barley, have very large genomes but low gene density, which has hampered sequencing progress (Schulte et al., 2009). For this reason, *Brachypodium* has been selected to be used as a model species for the Triticeae, and for which an 8X genome assembly has been recently published (Vogel et al., 2010).

Availability of sequenced genomes has provided the opportunity to perform evolutionary analyses of genes and gene families and to undertake comparative investigations of plant genome evolution.

Evolutionary analyses

Sequenced genomes have also provided the means to perform a more accurate investigation of gene phylogeny, the formulation of phylogenetic trees and measurement of evolutionary distances. Models used to estimate evolutionary distance are based on maximum parsimony or maximum likelihood (ML) (Ina, 1996). Parsimony examines the shortest possible pathway between sequences and then groups organisms according to the hypothesis that the pathway with the least number of changes is the most probable. The ML models consider all possible pathways between sequences and uses statistical probabilities to produce the most likely outcome. Estimation using ML models requires a large amount of computer time and storage space to perform the analyses, which makes them less conducive to processing large numbers of sequences. Evolutionary analyses use amino acid, nucleotide and codon based sequence similarities. Codon-based distance analyses measure the number of nucleotide differences between two species by calculating rates of synonymous and non-synonymous substitution of codons. Synonymous substitutions refer to those nucleotide changes at codon positions that do not change the encoded amino acid sequence whereas non-synonymous substitutions are those that change the amino acid of the protein (Kimura and Ohta, 1974; Nei and Gojobori, 1986; Goldman and Yang, 1994; Muse and Gaut, 1994; Ina, 1996). Because synonymous nucleotide substitutions do not alter the amino acid, it is assumed that they are under no evolutionary pressure and can therefore be used as a relative measure of time since separation of two genes from their common ancestor. If the rate of nucleotide substitution is known, this relative measure of time can be converted to an estimation of actual time since separation of the two genes. On the other hand, non-synonymous substitutions are assumed to be under the influence of evolutionary pressure. Because they produce mutations that alter the amino acid in the protein, they may be used to measure the degree of evolutionary pressure placed on the gene (Kimura and Ohta, 1974). In situations where there is maintenance of the *status quo* by selection against nucleotide non-synonymous substitutions, the mutation rate, or ω, will be $<<1$. This is known as purifying selection. Conversely, where

positive evolutionary pressure exists, ω will be >1. In the case of pseudogenes, no selection pressure exists in either direction, so ω ≈1.

As a first step in analysing the endo-(1,4)-β-glucanase family of genes, sequence information was sought for the family in barley, maize, rice, sorghum and *Arabidopsis* from databases and other sources. At the time that this work was done, the *Brachypodium* genome sequencing project was in its infancy and useful sequence data was not readily available. Although the endo-(1,4)-β-glucanase sequences from maize, *Arabidopsis*, rice and sorghum could be extracted or deduced from sequence databases, few sequences were available for barley endo-(1,4)-β-glucanases, and there was no information on the size of the family in barley. Thus cDNAs encoding endo-(1,4)-β-glucanases from barley were amplified, cloned, and gene sequences were determined from barley genomic BAC clones. Procedures used to obtain sequence information for the barley genes and enzymes are therefore described in this Chapter. It should also be noted that at the time of this analysis the maize genome was not publically available so many of the maize sequences were obtained from DuPont-Pioneer, under a formal collaboration between DuPont-Pioneer and the University of Adelaide.

Once gathered, the sequence data was used for a comparative analysis of gene structure, phylogeny, synteny and evolution across all, or some, of the above plant species. The phylogenetic analysis enabled the identification of orthologues between the cereals analysed. The analysis included the production of phylogenetic trees by neighbour joining bootstrap analysis based on amino acid sequences. Where genomic sequences were available, an intron/exon gene analysis was performed, adding further depth to the information on orthologous genes. Information on map positions for maize, sorghum and rice was gathered and the degree of map synteny between these species was identified. Map synteny provided further support, or otherwise, regarding orthology of genes between maize, sorghum and rice. Phylogenetic analysis was further developed through codon-based evolutionary distance analyses using the Nei and Gojobori

(1986) model as well as phylogenetic analysis by maximum likelihood (PAML) (Yang and Nielsen, 2000). These analyses, based on synonymous and non-synonymous substitutions, provided information regarding the evolution of the endo-(1,4)-β-glucanase gene orthologues in rice, barley, maize and sorghum.

2.2. METHODS

2.2.1. Database searches for GH9 endo-(1,4)-β-glucanase sequence data

2.2.1.1. Rice, sorghum and Arabidopsis

The endo-(1,4)-β-glucanases of plants are members of the glycosyl hydrolase family 9 (GH9). The Carbohydrate Active Enzyme database website (http://www.cazy.org/fam/acc_GH.html) was searched for GH9 hydrolases and all protein sequences on the website for members of the GH9 family belonging to *Arabidopsis thaliana* (*Arabidopsis*) and *Oryza sativa* (rice) were identified. For rice and sorghum, complementary deoxyribonucleic acid (cDNA) and genomic sequences, map positions and intron/exon data were found from searches of the Gramene website (http://www.gramene.org/). *Arabidopsis* cDNA, map and genomic information were sourced from the Salk Institute Genomic Analysis Laboratory (SIGNAL) database (http://signal.salk.edu/cgi-bin/tdnaexpress).

2.2.1.2. Barley

Rice and *Arabidopsis* amino acid sequences were used as queries to perform a basic local alignment search tool (BLAST) of *Hordeum vulgare* (barley) databases for expressed sequence tags (ESTs) and contiguous sequences (contigs), using the tBLASTn program which searches the translated nucleotide sequence database using an amino acid sequence query. This search was performed using The Institute of Genomic Research (TIGR) (http://www.tigr.org/), now a division of the J. Craig Venter Institute databases. The National Centre for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov/) was also searched using the same parameters. To check for duplicate sequence and EST and contig overlap, nucleotide sequences were aligned using ContigExpress (Vector NTI) and amino acid sequence alignment was performed using ClustalX and ClustalX2 (Larkin et al., 2007).

2.2.1.3. Maize

A tBLASTn was performed using the sequences from rice, *Arabidopsis*, wheat

and barley against the NCBI database to find *Zea mays* (maize) sequences. This search was then repeated against the Dupont-Pioneer B73 EST database at Des Moines, Iowa, USA. As with the barley sequences, proprietary alignment using the ContigExpress program and ClustalX were performed to check for duplicates and sequence overlap. To obtain maize genomic sequence, B73 BAC library sequences available in the Dupont-Pioneer proprietary collection and at the maizesequence.org website were searched using maize and sorghum cDNA query sequences. Syntenic regions between the maize endo-(1,4)-β-glucanase sequences and those of rice were determined using proprietary information from DuPont-Pioneer.

2.2.2. 3' SMART RACE PCR in barley

As most of the barley endo-(1,4)-β-glucanase sequences obtained from the database were incomplete, DNA walking using the switching mechanism at the 3' end of RNA transcript rapid amplification of cDNA ends (SMART RACE^{TM}) (Clontech) PCR sequencing techniques were used to extend the barley ESTs for which a 3' UTR sequence was not available. SMART RACE PCR removes the need for adaptor ligation and second strand DNA transcription by using a reverse transcriptase with terminal transferase activity for first strand cDNA production.

The 3' SMART RACE (Clontech) procedure utilises an oligo dT primer prefaced by a SMART CDS (Clontech) oligonucleotide for first strand cDNA production. A universal primer mix (UPM) anneals to the CDS oligonucleotide to prime the transcription of the cDNA for the reverse PCR reaction and a gene specific primer (GSP) is used to prime the forward reaction. To reduce non-specific primer annealing, ideally, the melting temperatures (Tm) for GSPs need to be >65°C, with a GC content of at least 60% if possible, to allow high annealing temperatures (68°C). For most genes at least two GSPs were designed, but not all EST sequences were long enough to enable design of more than one primer with a sufficiently high Tm. Primer sequences are listed in Table 1, Appendix A.

2.2.2.1. Preparation of RNA

Total ribonucleic acid (RNA) was extracted from plant tissue using the phenol/chloroform extraction method. Approximately 100 mg of tissue was ground in 1 ml of Trizol-like reagent containing 38% phenol equilibrated to pH 4.3, 1 M guanidine thiocyanate, 1 M ammonium thiocyanate, 0.1 M sodium acetate, 5% (v/v) glycerol and sterile filtered water. The ground tissue/Trizol-like reagent mixture was incubated at room temperature for 5 mins, 200 µl of chloroform was added, the mixture was shaken vigorously and left to stand at room temperature for 3 mins. This was centrifuged at 11,000 x g for 10 mins at 4°C. The upper aqueous phase was carefully transferred to a clean 1.5 ml tube and 500 µl of isopropanol was added to precipitate the RNA. The contents of the tube were mixed and left at room temperature for 10 mins. The tube was centrifuged at 11,000 x g for 10 mins at 4°C and the supernatant removed by pipette and discarded, leaving the RNA pellet in the tube. The pellet was washed in 75% (v/v) ethanol and spun at 7,500 x g for 5 mins at 4°C, the supernatant was removed. The resultant pellet was air dried and resuspended in 30 to 50 µl sterile filtered water. A 2 µl aliquot of resuspended RNA was separated on a 1% agarose gel to confirm that the RNA was not degraded, and the preparation was also measured for quantity and purity on the NanoDrop ND-1000 Spectrometer (NanoDrop Technologies). The RNA suspension was stored at -80°C.

2.2.2.2. 3' SMART RACE cDNA preparation

The 3' RACE Ready cDNA (Clontech) was prepared by adding 1 to 3 µl RNA to 1 µl 12 µM 3' CDS primer A and sterile filtered water to 5µl, mixed and spun briefly then incubated at 70°C for two minutes. Tubes were immediately cooled on ice for 2 mins and the contents briefly spun before adding 2 µl 5X 1st strand buffer (Clontech), 1 µl 20 mM dithiothreitol (DTT), 1 µl 10 mM dNTP mix and 0.5 µl Powerscript reverse transcriptase (Clontech). The contents were mixed by gentle pipetting and incubated at 42°C for 90 mins in a hot lid thermal cycler, after which 100 µl of TE buffer was added. The 3' RACE Ready cDNA was tested by PCR using a primer pair for the enzyme glyceraldehyde-3-phosphate dehydrogenase (GAPDH), and stored in 30 µl aliquots at -20°C.

2.2.2.3. 3' SMART RACE PCR

A 25 µl 3' SMART RACE PCR reaction mix was prepared with 16 µl sterile filtered water, 2.5 µl 10X buffer, 1 µl deoxyribonucleic acids comprising 10 mM each deoxyribo-adenosine triphosphate, deoxyribo-cytosine triphosphate, deoxyribo-thymidine triphosphate and deoxyribo-guanidine triphosphate (dNTPs), 1.5 µl 3'RACE Ready cDNA, 2.5 µl universal primer mix (UPM), 1 µl gene specific primer (GSP) (10 µM) and 0.5 µl BDTaq Advantage II polymerase (Clontech). Cycling conditions were as follows, Step 1: 94°C for 20 secs, Step 2: 68°C for 30 secs and Step 3: 72°C for 3 mins, with steps 1 to 3 repeated for 40 cycles.

If no distinct bands were seen on a 1% agarose gel, nested reactions were performed under the same conditions except a 20 to 50X dilution of the product from the first round PCR was used as the template. The nested GSP was located 3' of the first GSP and a nested universal primer (NUP) replaced the UPM. To improve specificity, "touchdown" cycling conditions were sometimes employed as follows: 5 cycles of 94°C for 30 secs and 72°C for 3 mins, followed by 5 cycles of 94°C for 30 secs, 70°C for 30 secs and 72°C for 3 mins, and finally 30 cycles of 94°C for 30 secs, 68°C for 30 secs and 72°C for 3 mins. A more stringent touchdown was also used: 5 cycles of 94°C for 30 secs and 72°C for 3 mins followed by 30 cycles of 94°C for 30 secs, 70°C for 30 secs and 72°C for 3 mins.

2.2.3. Polymerase Chain Reaction (PCR) using Platinum Taq polymerase

A PCR reaction mix using 0.1 µl Platinum Taq DNA Polymerase (Invitrogen) contained 2.5 µl 10X reaction buffer, 1 µl dNTPs (10 mM), 1.5 mM magnesium chloride ($MgCl_2$), 0.2 µM each forward and reverse primers, 1 µl template and sterile filtered water to a total volume of 25 µl. Where amplification proved difficult, 10% dimethylsulphoxide was added to the PCR mix with a concomitant reduction in the volume of sterile filtered water. Cycling conditions for the amplification were Step 1: denaturation at 95°C for 3-5 mins, Step 2:

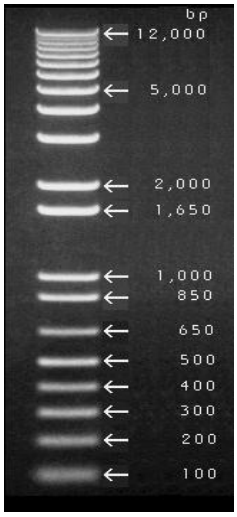


Figure 2-2: 0.7ug 1 kb Plus ladder.

The ladder was run on a 0.9% agarose gel and stained with ethidium bromide. The 1650 bp band is generated from pUC. The bands smaller than 1000 bp are derived from lambda DNA (Invitrogen).

denaturation at 95°C for 30 secs, Step 3: annealing at between 45 and 60°C for 30 secs and Step 4: elongation at 72°C for 1 min per kilobase (kb). Steps 2 to 4 were repeated for 25 to 40 cycles.

2.2.4. Observation of nucleic acids by agarose gel electrophoresis

RNA and DNA fragments were size separated using a voltage of 120 V. Between 2 and 25 µl of RNA or DNA were loaded into wells in a gel containing 1 - 2% (w/v) agarose dissolved in a Tris-Acetate-EDTA (TAE) buffer containing 0.04 M Tris-HCl, 1 mM NaEDTA, 0.02 M acetic acid and stabilized at pH 8.0 using NaOH, with addition of 0.5 µg per ml ethidium bromide. The resultant DNA products were separated by agarose gel electrophoresis and observed under UV light. The 2% gel was used for separating small molecular weight DNA (100-300 bp) while the 1% gel was used for RNA and larger DNA products (>300bp). All nucleic acid separations were performed with the Invitrogen 1kb Plus DNA ladder to verify fragment sizes (Figure 2-2).

2.2.5. Cleaning of PCR products

Non-DNA, short DNA sequence (oligonucleotides) and other contaminants such as salts contained in the PCR reaction mixture were removed using the clean-up gel extraction column, NucleoSpin Extract II (Machery-Nagel). If multiple products were obtained, the PCR products were separated on a gel and the DNA fragment of interest was cut out of the gel and dissolved at 50°C for 10 mins in NucleoSpin Extract II Buffer NT at the rate of 200 µl per 100 mg gel. In instances where a single product was seen on the gel, the sample was made up to 50 µl with Tris-EDTA (TE) buffer and one volume of sample was mixed with two volumes of NucleoSpin Extract II Buffer NT. Primer dimers and small molecular weight fragments were removed by diluting NucleoSpin Extract II Buffer NT with water, according to the manufacturer's instructions. Increasing the dilution of the buffer allowed removal of higher molecular weight DNA fragments. NucleoSpin Extract II Buffer NT samples were placed on spin columns and centrifuged at11,000 x g for 1 min. The DNA retained on the

column was washed with 650 µl NucleoSpin Extract II Buffer NT3 containing 70% ethanol and again spun at11,000 x g for 1 min. The "flow through" was removed and the column re-spun for 2 min to ensure that no ethanol residue was retained in the column. The spin column was placed in a clean microcentrifuge tube and elution was effected by adding 15 - 30 µl of NucleoSpin Extract II Buffer NE (5 mM Tris/HCl, pH 8.5) and left at room temperature for approximately 1 min. The column was spun for 1 min at 11,000 x g, the "flow through" placed back on the column and the process repeated. This second incubation and spin was to ensure a higher yield without a dilution penalty. To ensure that the cleaned product had not been lost and to ascertain its concentration, 5 µl was separated on a 1% gel.

2.2.6. DNA Sequencing

The concentration of the DNA fragment was checked using the NanoDrop ND-1000 Spectrometer (NanoDrop Technologies) and confirmed by gel electrophoresis. Sanger sequencing was performed at the Australian Genome Research Facility (AGRF) by capillary separation using ABI Prism BigDye Terminator Sequencing Reaction Kits (Applied Biosystems)) and the AB 3730*xl* sequencer. The sequencing reaction mix contained 1.35 µl BD Buffer, 1 µl primer (4 µM), 0.5 µl BD enzyme, approximately 100 ng of DNA template and sterile filtered water to a final volume of 12 µl. The sequencing PCR cycle used was Step 1: 96°C for 5 mins, Step 2: 96°C for 10 secs, Step 3: 50°C for 5 secs, Step 4: 60°C for 4 mins with steps 2 to 4 repeated 25 times. The amplified sequencing reaction mixture was cleaned by adding 75 µl of a wash mixture containing 75% (v/v) ethanol and 0.2 mM $MgSO_4$, incubated at room temperature for 20 min and centrifuged at 16,000 x g for 17 min. The supernatant was removed by tapping the microfuge tube upsidedown on absorbent paper until minimal visible liquid remained in the tube. Removing as much of the supernatant as possible from the tube was important to reduce the amount of contamination by dye reagents. Drying was completed at 37°C for 15 to 30 mins and the samples sent to AGRF for analysis.

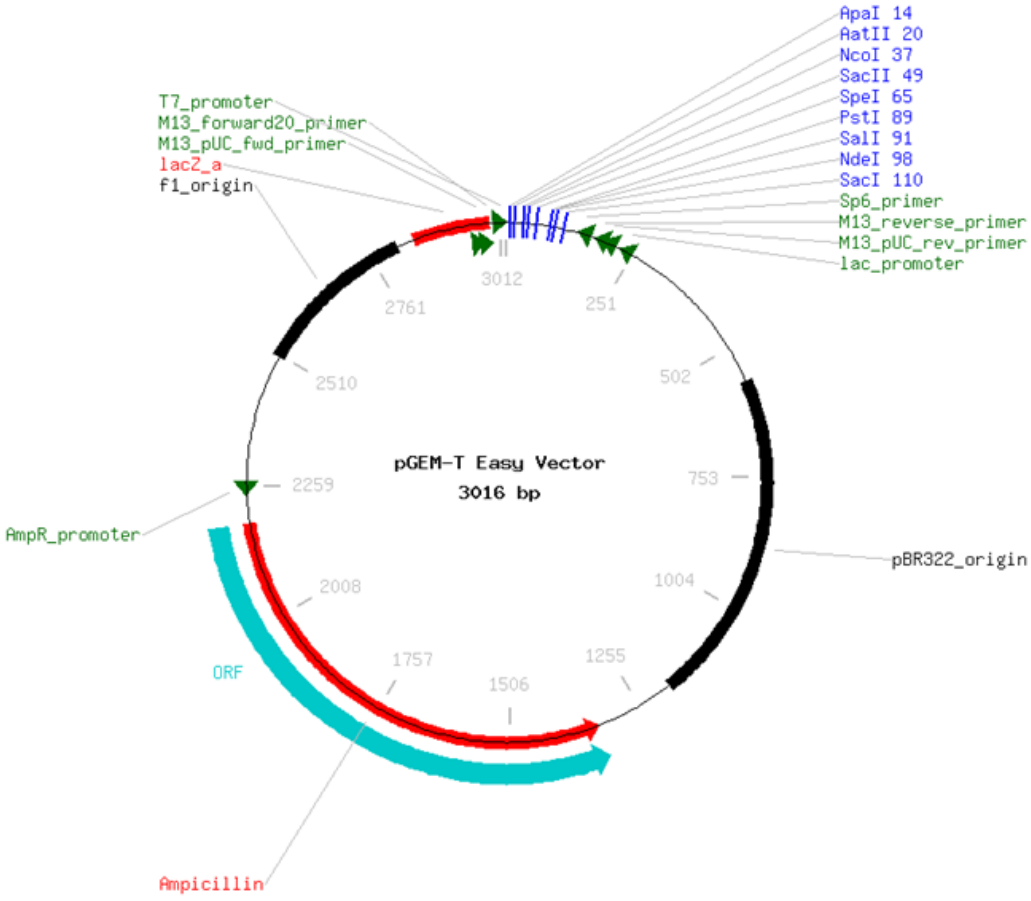


Figure 2-3: A diagram of the pGEM-T Easy vector.

The Amp_R gene, M13 forward and reverse primer sets and a multiple restriction enzyme site and AT cloning, as well as LacZ_a gene allowing for blue/white selection are present in this vector. Cloning into the vector using AT cloning or the restriction enzyme cloning sites disrupts LacZ transcription and hence expression and prevents the production of blue colonies.

2.2.7. Cloning into the pGEM-T Easy vector

The pGEM-T Easy vector (Figure 2-3) in the pGEM-T Easy Vector System I (Promega) has 3' thymidine overhangs to enable more efficient cloning. This overhang takes advantage of the deoxyribo-adenosine base that is added to the end of the amplified product by several polymerase enzymes. The vector contains an ampicillin resistance gene. The β-galactosidase gene that spans the cloning site creates a convenient blue/white bacterial colony selection method. Blue/white selection occurs when the bacterial cell colonies are grown on media containing 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal); successful cloning disrupts the β-galactosidase gene situated across the cloning site, thus preventing the hydrolysis of Xgal to an insoluble blue product. Successful cloning prevents production of blue colonies, so all white colonies should contain cloned inserts.

2.2.7.1. A-tailing PCR product

The 10 µl PCR reaction mix was added to 2.5 µl 10X buffer, 1.5 mM $MgCl_2$, 1 µl dNTPs or dATPs (10 mM), 0.1 µl Platinum Taq polymerase (Invitrogen) and sterile filtered water to 25 µl. This mixture was incubated at 95°C for 3 min and 72°C for 45 min.

2.2.7.2. Ligation

Cloning into the pGEM-T vector was carried out by adding 2 µl freshly amplified, cleaned PCR product or A-tailed PCR mix to 1 µl pGEM-T Easy vector (diluted 1/5 with sterile filtered water), 0.5 µl Invitrogen or Promega T4 Ligase, 5 µl 2X T4 ligase buffer (Promega) or 2 µl of 5X T4 ligase buffer (Invitrogen) and sterile filtered water to a total volume of 10 µl. This mixture was incubated at room temperature for 1 hour and subsequently at 4°C for 36-72 hours.

2.2.7.3. Transformation of DH5α Escherichia coli

Heat shock competent DH5α *E.coli* cells were thawed on ice for 10 mins and the

total ligation mix added to the cells and left on ice for 20 mins. The cell mixture was incubated (heat shocked) at 42°C for 30-45 secs before replacing on ice for 2 mins. Media comprising 200 µl Luria broth (LB) containing 10% (w/v) Tryptone, 5% (w/v) Yeast Extract and 5% (w/v) NaCl was added to the heat shocked cells and the mixture incubated at 37°C for 1 hour. The total cells were spread onto LB agar plates containing 100 µg per ml ampicillin and 20 µl per ml 98% (v/v) Xgal (Sigma) and incubated for approximately 16 h at 37°C. A number of white colonies were selected from the plates and restreaked onto fresh LB agar plates with 100 µg per ml ampicillin and incubated at 37°C overnight. The remainder of each colony was used as template for colony PCR, using the M13F and M13R primer pair and Platinum Taq polymerase to check insert size.

2.2.7.4. Alkaline lysis plasmid preparation

Colonies from selected DH5α cells were inoculated into 5ml LB containing 100 µg per ml ampicillin and incubated overnight on a shaker at 120 to 200 cycles per minute at 37°C. Plasmid was extracted from 2 ml cell culture as follows. The cell culture was centrifuged at 16,000 x g for 30 secs, the supernatant removed and the pellet containing the cells resuspended in 100 µl of GTE buffer containing 50 mM glucose, 25 mM Tris-HCl and 10 mM EDTA and placed on ice. A fresh solution of cell breaking buffer containing 0.2 M NaOH and 1% sodium dodecyl suphate (SDS) was prepared and 200 µl was added to the tube, which was gently rolled 4 to 6 times and placed back on ice. Immediately, 150 µl 3 M potassium acetate was added and the tube rolled as before. All tubes were placed at -20°C for 10 to 60 mins. The tubes were removed from the -20°C freezer, centrifuged at 16,000 x g for 15 mins and the supernatant containing the plasmid was removed to a fresh tube containing 1 ml 100% ethanol. The mixture was incubated at room temperature for 10 mins and centrifuged at 16,000 x g for 10 mins. The supernatant was carefully removed and discarded and the pellet washed in 75% (v/v) ethanol. The tube was re-centrifuged at 16,000 x g, the supernatant discarded and the pellet air or vacuum dried in the Savant Speed Vac SC110. Depending on the size of the pellet, it was resuspended in 20-50 µl TE

buffer containing RNase40, the plasmid concentration was measured on the NanoDrop (NanoDrop Technologies) and subsequently stored at -20°C.

2.2.7.5. Restriction enzyme digest

Insert length cloned into the extracted plasmid was checked by performing an *Eco*RI restriction enzyme (RE) digest using 1-2 µl of prepared plasmid, 0.15 µl *Eco*RI (20 U per µl) RE (New England Biolabs), 1 µl 10X *Eco*RI RE digest buffer (New England Biolabs) and sterile filtered water to 10 µl. The mixture was incubated at 37°C for 2-3 hours and separated on a 1% agarose gel.

2.2.8. Barley BAC library screen

Barley genomic gene sequences were obtained by screening a Morex barley BAC library. The BAC library, Morex "HVVMRXALLhC" *Hind*III, is cloned into the 7.5 kb pIndigoBAC-5 *Hind* III vector and transformed in the DH10B *E.coli* strain (Bu Jun Shi, Australian Centre for Plant Functional Genomics). The library contains ten filters, giving a 3.5X coverage with an average clone size of 115 kb.

2.2.8.1. Preparation of ^{32}P labeled DNA Probe

Primers were prepared for genes representing the different subfamilies of the endo-(1,4)-β-glucanases (Urbanowicz et al., 2007a). Where there was insufficient sequence in barley to prepare a probe of at least 500 nucleotide bases, sequence information from rice or maize was used for primer preparation on the assumption that nucleic acid sequence differences between species were small for areas of high identity. Primers used for probe amplification are shown in Table 2, Appendix A.

Probe sequences were amplified by PCR in a reaction mix using the PlatinumTaq DNA polymerase (Section 2.2.3). Template was provided by 1 µl barley cDNA (*HvCEL3* and *14*) or 50X dilution of the gene open reading frame cloned into the pCR8/GW/TOPO vector (Invitrogen) (*HvCEL5* and *11*) and sterile filtered water to a total volume of 25 µl. The amplified products were separated on a 1%

agarose gel and cloned into the pGEM-T Easy vector and sequenced by AGRF for authentication. Amplified sequences used to probe the BAC library can be found in Appendix B. The lengths of the probes were 859 bp for *HvCEL3*, 660 bp for *HvCEL5*, 670 bp for *HvCEL11* and 1300 bp for *HvCEL14*.

Radioactive probes were prepared by adding 4 µl probe DNA to 1 µl sterile filtered water and 3 µl random 9mer primer mix. This was heated at 100°C for 5 mins and placed on ice for 5 mins. To this was added 12.5 µl labeling buffer containing 60 µM each primer (dATP, dTTP and dGTP), 150 mM Tris-HCl pH 7.6, 150 mM NaCl, 30 mM $MgCl_2$ and 300 µg per ml bovine serum albumin (BSA) with 1.5 µl Klenow fragment and the mixture was briefly centrifuged. In the radiation laboratory, 3.5 µl CT^{32}P was added and the mixture incubated at 37°C for 1 h. The probe was cleaned on a Sephadex-100 column. The Sephadex-100 column was prepared by combining 300 ml TE buffer and 10 g Sephadex-100 and carefully adding to a Pasteur pipette plugged at the shoulder with glass wool. The Sephadex-100 was allowed to settle as the TE eluted and the column was topped up until the top of settled Sephadex-100 was at the neck of the pipette. To prevent the column from drying out, the pipette was filled with TE buffer and sealed with Parafilm^{TM} until required. To clean the probe, three microfuge tubes were placed beneath the pipette to collect the "flowthrough". The Parafilm^{TM} was removed from the pipette and the excess TE buffer allowed to flow into microfuge tube #1. As the last of the TE buffer flowed into the Sephadex column, the labeled probe was added to the pipette and allowed to flow into the column. TE buffer was added and kept above the level of the Sephadex as the probe flowed through. Radioactivity of the "flowthrough" was continuously measured at the base of the pipette and when it reached 100 counts per second (cps), the pipette was allowed to drain into tube #2 to collect the cleaned probe. Radioactivity levels peaked at over 1000 cps and when they had declined to 500cps the remaining "flowthrough" was collected in tube #3. The tube #2 containing the labeled probe, was boiled for 5 mins with 500 µl carrier DNA (5 mg per ml salmon sperm DNA) and placed on ice for 5 mins.

2.2.8.2. Probing BAC library filters

Because the filters were new, they were prewashed in 5X sodium chloride/sodium citrate (SSC) solution containing 0.75 M NaCl and 0.075 M tri-sodium citrate for 5 mins before being rolled into cylinders and placed in two bottles with five membranes in each. Prehybridisation solution containing 15 ml 10X SSC, 10.5 ml sterile filtered water, 3 ml Denhardt's III (2% (w/v) BSA, 2% (w/v) Ficoll 400, 2% (w/v) polyvinyl-pyrrolidone 360 and 10% (w/v) SDS) and 1.5 ml carrier DNA (5mg per ml), was added to each container and gently rotated in an oven overnight at 65°C. The next day the prehybridisation solution was replaced with the hybridization solution containing the prepared probe. The hybridization solution contained 0.5 ml sterile filtered water, 3 ml 5X hybridisation buffer (3 M NaCl, 100 mM Pipes and 25 mM Na_2EDTA at pH 6.8), 3 ml Denhardt's III, 3 ml 25% (w/v) Dextran sulphate and 0.5 ml carrier DNA (5mg per ml). After incubating at 65°C for 15 mins, 10 ml of hybridisation solution was added to each bottle and rotated at 65°C for 5 min to equilibrate with the membranes before adding the labelled probe taking care not to allow it to touch the membranes before dilution with the hybridisation solution. The membranes were rotated with the labelled probe in hybridisation solution overnight in the 65°C oven.

After removal of the hybridisation solution, 40 ml 2X SSC with 0.1% SDS was added to the bottles and incubated as before for 20 mins. The membranes were removed from the bottles and laid flat in a rectangular plastic container to which 200 ml washes of successively lower concentrations of SSC (2nd wash: 1X SSC, 0.1% SDS, 3rd wash: 0.5X SSC, 0.1% SDS, 4th wash: 0.2X SSC, 0.1% SDS) were performed in a shaking water bath at 65°C. After the second and successive washes, the radioactivity of the filters was measured by Geiger counter and washes ceased when counts were 10-20 cps.

Filters were blotted dry, covered in plastic and placed DNA side up in cassettes with intensifying screens. In the dark room, X-ray film was placed over the filters and the cassettes left at -80°C until developed film showed clear images

such that the position on the membrane of clones to which the probe had annealed could be identified.

2.2.8.3. Identification/authentication of positive clones

Positive clones were grown in 100 µl LB in 96 well plates and were also plated onto LB agar, both containing 12.5 µg per ml chloramphenicol (Sigma). These were incubated overnight at 37°C and stored at 4°C. PCR, using Platinum Taq polymerase (Section 2.2.3) was used to authenticate that the positive BACs from the screen contained cellulase genes. DNA template for PCR was prepared by taking clones grown in the 96-well plates, which were pooled by rows (A-H) and columns (1-12); 10 µl of culture from the wells in each row and each column was pooled, the cells centrifuged and the supernatant replaced with sterile filtered water. Products amplified in the PCR reaction were separated on a 2% agarose gel and sequenced by AGRF. Primers used in the PCR reactions are presented in Table 3, Appendix A. Because the cellulase genes have areas of high similarity, where possible, primers were designed from 3'UTR regions of the genes so that gene-specific products were generated. This was not possible in all cases, so where no 3'UTR region was available, areas of lowest sequence identity between genes were used for primer design.

2.2.8.4. Extraction of plasmid from BAC clones

Clones were inoculated into 5 ml LB with 12.5 µg per ml choramphenicol in 50 ml tubes and shaken at 120 to 200 rpm overnight at 37°C. Cultures were centrifuged at 3000 x g for 10 mins, the supernatant discarded, the cells resuspended by vortexing with 200 µl Qiagen P1 containing RNase A and the mixture transferred to a 1.5 ml microfuge tube. A 400 µl solution containing 0.2 M NaOH and 1% SDS was added, the tube gently inverted six times and left at room temperature for 5 mins before 300 µl Qiagen P3 was added and the tube again gently inverted 6 times. The tubes were left on ice for at least 7 mins and up to 1 h and centrifuged for 25 mins at 16,000 x g at 4°C. The supernatant was removed to a fresh 1.5 ml tube to which 550 µl isopropanol was added, mixed by

inversion 4 to 6 times and centrifuged for 5 mins at 12,000 x g. Care was taken to leave as little supernatant as possible in the tube. This was achieved by re-centrifuging the tube for 10 secs to dry the sides, with subsequent removal of the remainder of the supernatant. The pellet was washed with 400 µl 75% (v/v) ethanol, the tube spun at 12,000 x g for 2 mins, the supernatant discarded and the pellet dried and resuspended in 15 to 20 µl TE buffer. The presence of the extracted plasmid was checked by separation on a 1% agarose gel and DNA concentration measured using the NanoDrop (NanoDrop Technologies).

2.2.9. Endo-(1,4)-β-glucanase sequence alignment and phylogenetic trees

The multiple sequence alignment of genes encoding endo-(1,4)-β-glucanases from barley, maize, rice, sorghum and *Arabidopsis* was performed using amino acid sequences in the ClustalX and updated ClustalX2 software program (Larkin et al., 2007). The parsimonious phylogenetic tree was produced using a 1000 repetition neighbour joining (NJ) bootstrap analysis in Clustal X and ClustalX2 and viewed in TreeView (Page, 1996). This analysis was performed to more easily classify the genes in accordance with the structural sub-families as identified by Urbanowicz et al. (2007). Identification of orthologues was also more easily achieved.

2.2.10. Gene intron and exon structure

For maize and barley, intron/exon structures were deduced from the genomic sequence. In the first instance, the amino acid sequences from cDNA obtained from available genes, ESTs and contigs were used as templates. The genomic sequence was translated in six reading frames using internet resources (http://bio.lundberg.gu.se/edu/translat.html), and the cDNA sequence templates were employed to aid identification of the GT…AG intron borders (Rogozin et al., 2005). For genes that did not already have a full length cDNA sequence, similarity of protein sequence with orthologous rice genes was assumed, and these were used, instead, as the template.

Rice and sorghum intron/exon structures were obtained from the Gramene website (http://www.gramene.org/), confirmed by amino acid sequence alignment using ClustalX and ClustalX2 (Larkin et al., 2007) and adjusted where discrepancies existed.

2.2.11. Mapping

In silico mapping of maize endo-(1,4)-β-glucanase genes was performed by searching the Dupont–Pioneer B73 BAC and chromosomal supercontig database comprising public and proprietary sequences and maizegenome.org website using B73 BAC identification numbers. The positions of these BACs on the maize Chromosomes were determined using proprietary software. The Gramene website was used for determining the map locations of sorghum and rice sequences using gene identification numbers as the query.

2.2.12. Measuring pairwise codon-based evolutionary distance

Measurement of evolutionary distances between orthologues was performed as a means of estimating time since separation of the plant species or genes, to verify the trees obtained from the ClustalX2 program and to provide a measure of rates of mutation for the genes. The number of different methods available for measuring evolutionary distances is a reflection of the difficulty of performing this task with any degree of accuracy (Nei and Gojobori, 1986; Muse and Gaut, 1994; Ina, 1996; Yang and Nielsen, 2000). Many of the available programs contain assumptions that create some degree of uncertainty in the estimations. The uncertainty results from having to assess the base changes that have occurred between two sequences since their separation, which includes transitions and transversions. A transition is defined as a base substitution of purine or pyrimidine to another purine or pyrimidine and a transversion is a base substitution of a purine to pyrimidine or vice versa. Furthermore, when comparing two sequences, if two nucleotide differences occur in one codon, there

are two possible pathways and if all three nucleotides in the codon are different, six different pathways are possible.

In order to perform an evolutionary analysis using nucleotide sequences, the sequence must first be aligned codon by codon with the protein sequence. This was done using MAGNOLIA software, and saved in ClustalW format. In two cases MAGNOLIA did not correctly align the codons with the amino acid and also inserted extra nucleotides into the sequence so some manual adjustment was required. Sequences were analysed for rates of synonymous (S) and non-synonymous (N) substitution with the following modifications: the nucleotide ORF was manually truncated to remove signal peptides that have low levels of sequence identity between genes (personal communication: Ute Baumann and John Toubia, Australian Centre for Plant Functional Genomics).

2.2.12.1. Nei and Gojobori measurement of evolutionary distance

The phylogenetic analysis program, Molecular Evolution Genetics Analysis software version 4.0 (MEGA4), was used to compare evolutionary distances between gene orthologues (Tamura et al., 2007) with the Nei and Gojobori model (Nei and Gojobori, 1986). This program was selected because it is readily available, easy to use, provides several different analysis programs and is used by many published authors (Gaut et al., 1996; Salse et al., 2008; Xu and Messing, 2008b; Xu and Messing, 2008a; Yang et al., 2008). The Nei-Gojobori Model was implemented, firstly calculating the p-distance which is the ratio of substitutions at synonymous or non-synonymous sites as a ratio of total synonymous or non-synonymous sites. (Nei and Gojobori, 1986). Pairwise distance analyses and standard error were calculated using the bootstrap method with 500 replicates.

The Nei-Gojobori model makes several assumptions. Firstly, it is assumed that substitution rates for all synonymous sites are equal, including rates for transitions and transversions. This model assumes equal probability for all

possible pathways when there is more than one nucleotide substitution at a codon position. It also assumes equal codon usage (Nei and Gojobori, 1986).

The proportion of synonymous (p_S) or non-synonymous substitutions (p_N) per total synonymous or non-synonymous sites is calculated as follows:

Equation 2-1: $p_S = S_d/S$

Equation 2-2: $p_N = N_d/N$

where S_d and N_d represent the number of synonymous and non-synonymous differences respectively, and S and N the total number of synonymous and non-synonymous sites.

To correct for multiple substitutions per site, the Jukes Cantor formula is used. However, silent sites with two-fold and three-fold degeneracy are not taken into account (Jukes and Cantor, 1969):

Equation 2-3: $d_S = -3/4\log_e(1-4/3p_S)$

and

Equation 2-4: $d_N = -3/4\log_e(1-4/3p_N)$.

2.2.12.2. Phylogenetic analysis by maximum likelihood (PAML) codeml

The phylogenetic analysis by maximum likelihood (PAML) using the codeml program estimates synonymous and non-synonymous changes by including differences in codon usage and rate ratios of transition/transversion substitutions (κ) (Goldman and Yang, 1994; Yang and Nielsen, 2000). Several assumptions are also applied with this method. Estimation of the transition/transversion rate ratio is based on the fourfold degenerate sites and non-degenerate sites. Non-synonymous substitutions are assumed to have the same rate as synonymous substitutions. This may not be true because non-synonymous substitutions are weighted by the fact that changes in amino acids are more likely to be to another

similar class of amino acid (Yang and Nielsen, 2000). Furthermore, no consideration is given for insertions and deletions. The PAML codeml method involves the implementation of several steps (Yang and Nielsen, 2000).

1. A weighted average for κ is estimated by firstly identifying the nondegenerate and fourfold degenerate sites of sequence codons, and is then calculated according to the number of respective sites. A transition probability matrix is produced from these estimates.

2. Numbers of synonymous and non-synonymous sites are calculated using the estimated κ and nucleotide base frequencies estimated from the data.

3. The starting position for distance and mutation rates (ω) is estimated using the methods of Nei and Gojobori (1986).

4. The numbers of synonymous and non-synonymous differences are counted using κ, codon frequencies and current distance and ω values. When the two codons differ by more than one site the transitional probability matrix is used to weight the different possible pathways.

5. Calculation of d_S and d_N is performed with corrections for multiple hits and base frequency differences.

Steps 4. and 5. are repeated until convergence occurs. Usually only two or three iterations are required to achieve convergence (Yang and Nielsen, 2000). The webserver PAL2NAL (www.bork.embl.de/pal2nal) was used to perform the PAML codeml estimations (Goldman and Yang, 1994; Suyama et al., 2006).

2.2.12.3. Calculating substitution rates, time since separation and mutation rates

In order to gain an insight into time of separation of maize from sorghum, and barley from rice, estimates of the number of substitutions per synonymous site per year (Ks) were performed. For maize (Z) and sorghum (S) and for the maize homoeologues, the rice (O) orthologue was used as the outgroup. In the case of

rice and barley (H), the maize orthologue provided the outgroup. The average d_S of the two genes with the outgroup was divided by the estimated time since separation (T) of those genes from the outgroup (Gaut et al., 1996). In this case, an estimate of 50 mya was used as the separation time (Kellogg, 1998; Salse et al., 2008).

Rate of synonymous substitution for maize and sorghum

Equation 2-5: $Ks = (d_{ZO} + d_{SO}) / (2 \times 2 \times 50 \times 10^6)$

and for barley and rice

Equation 2-6: $Ks = (d_{HZ} + d_{SZ}) / (2 \times 2 \times 50 \times 10^6)$

Information on the direction of evolutionary pressure can be measured by calculating the ratio of non-synonymous to synonymous substitutions:

Equation 2-7: $\omega = d_N/d_S$

2.3. RESULTS

2.3.1. Sequence identification and gene structure

Barley, rice, sorghum and maize cDNA sequences were obtained by database searches, using all of the rice and *Arabidopsis* endo-(1,4)-β-glucanases as query sequences. In barley, 3' SMART RACE was also employed to obtain cDNA sequence. Genomic sequence for maize, rice and sorghum were acquired by sequence searches of databases using gene identification tags or cDNA as the query. No barley genomic sequence was available on public databases so a Morex barley *Hind*III BAC library was screened for genomic endo-(1,4)-β-glucanase gene sequence. A summary of the results of all cereal searches is presented in Table 2-1.

2.3.1.1. Rice, Sorghum and Arabidopsis endo-(1,4)-β-glucanase gene identification

Searches for GH9 family genes on the CAZy and Gramene databases found 25 endo-(1,4)-β-glucanase genes in rice and 23 in sorghum. An *Arabidopsis* sequence search was performed using the CAZy and SIGNAL websites and 25 genes were found. One of these genes, *AtGH9A16*, is annotated on the SIGNAL website as a pseudogene.

2.3.1.2. Barley sequence identification

A summary of the results of the barley sequence search, 3' SMART RACE and the Morex barley BAC screen is presented in Figure 2-4. A search on the CAZy, TIGR and NCBI databases found a total of 28 barley endo-(1,4)-β-glucanase contigs and ESTs. After alignment using the ContigExpress, and ClustalX programs this was reduced to 24 unique sequences, of which two appeared to be full-length, two were almost full-length and the remaining 20 were contigs and single ESTs. BI956376 and BI779097 overlapped by nine identical amino acids, but investigation of the overlapping nucleotide sequence showed three

Table 2-1: Orthologous genes between maize, barley, sorghum and rice.

The maize gene names were arbitrarily assigned. The maize gene annotation is a recent addition to the maizesequence.org website. The barley TIGR contig denotes ESTs and contigs that were annotated on the TIGR database and the barley gene name was arbitrarily assigned. The sorghum and rice annotations are from the gramene.org website. These orthologous genes are derived from protein sequence, intron splice sites, codon-based evolutionary distances and, where applicable, genome locations. It was not possible to assign HvCEL4 and HvCEL8 definitively to either ZmCEL1 or ZmCEL2 and their respective sorghum and rice orthologues. Several short barley ESTs have not been assigned gene names.

		Barley				
Gene name	**Gene annotation**	**TIGR contig**	**Gene name**	**Sorghum**	**Rice**	
ZmCEL12	N/A	TC147741	HvCEL5	Sb06g034260	Os04g0497200	LOC_Os04g41970
ZmCEL7	GRMZM2G003379_T01	TC149562		Sb01g036480.1	Os03g0329500	LOC_Os03g21210
ZmCEL32	GRMZM2G 167669_T01	No orthologue		Sb07g019470.1	Os08g0387400	LOC_Os08g29770
ZmCEL30	GRMZM2G 147849_t01	Hvkor1	HvCEL1	Sb01g008860.1	Os03g0736300	LOC_Os03g52630
ZmCEL12	GRMZM2G 477603_01	TC147741	HvCEL5	Sb06g034260	Os04g0497200	LOC_Os04g41970
ZmCEL8/29	GRMZM2G 703265_T01	TC134447	HvCEL12	Sb02g024050.1	Os09g0394300	LOC_Os09g23084
ZmCEL26	N/A	TC144321	HvCEL14	Sb04g028520.1	Os02g0738600	LOC_Os02g50490
ZmCEL14	GRMZM2G 110735_T01	Hvkor1	HvCEL1	Sb01g008860.1	Os03g0736300	LOC_Os03g52630
ZmCEL9	GRMZM2G 153987_T01	No orthologue		Sb10g030140.1	Os06g0715300	LOC_Os06g50140
ZmCEL17	GRMZM2G 141911_T01	BM442502		Sb04g001960.1	Os02g0123700	LOC_Os02g03120
ZmCEL24	N/A	TC130995	HvCEL18	Sb04g003690.1	Os02g0151300	LOC_Os02g05744
ZmCEL1	GRMZM2G 76049_T01	TC153055+ TC143214	HvCEL4	Sb04g028790.1	Os02g0733300	LOC_Os02g50040
		BI779097+ TC151646	HvCEL8			
ZmCEL25	GRMZM2G 151257_T01	TC144321	HvCEL14	Sb04g028520.1	Os02g0738600	LOC_Os02g50490
ZmCEL13	N/A	No orthologue		Sb04g034920.1	Os02g0778600	LOC_Os02g53820
ZmCEL21	GRMZM2G 009025_T01	No orthologue		Sb07g001480.1	Os08g0114200	LOC_Os08g02220
ZmCEL15	203408.4 FGT003	No orthologue		Sb09g002490.1	Os05g0129200	LOC_Os05g03840
ZmCEL2	GRMZM2G 455642_T01	TC153055+ TC143214	HvCEL4	Sb10g009270.1	Os06g0256900	LOC_Os06g14540
		BI779097+ TC151646	HvCEL8			
ZmCEL8/29	N/A	TC134447	HvCEL12	Sb02g024050.1	Os09g0394300	
ZmCEL18	N/A	TC136133	HvCEL2		Os09g0530200	LOC_Os09g36060
ZmCEL11	GRMZM2G 147422_T01	BM369748+	HvCEL3	Sb02g030990.1	Os09g0533900	LOC_Os09g36350
ZmCEL20	GRMZM2G 453565_T01	AL406224+ CK025855+ TC132904	HvCEL11	Sb03g001560.1	Os01g0219600 Os01g0220100	LOC_Os01g12030 LOC_Os01g12070
ZmCEL10	GRMZM2G 331566_T01	TC149885	HvCEL7	Sb03g012840.1	Os01g0312800	LOC_Os01g21070
ZmCEL19/28	GRMZM2G 143747_T01	BM377061		Sb09g006670.1	Os05g0212300	LOC_Os05g12150
ZmCEL3	GRMZM2G 154678_T01	TC153576+ TC150512	HvCEL10	Sb10g008880.1	Os06g0247900	LOC_Os06g13830
ZmCEL7	GRMZM2G 099101_T01	TC149562		Sb01g036480.1	Os03g0329500	LOC_Os03g21210
ZmCEL16	GRMZM2G 125436_T01	No orthologue		Sb06g017600.1	Os04g0443300	LOC_Os04g36610
ZmCEL6	GRMZM2G 066162_T01	TC147741	HvCEL5	Sb06g021440.1	Os04g0497200	LOC_Os04g41970
ZmCEL4	GRMZM2G 343144_T01	No orthologue		Sb06g032760.1	Os04g0674800	LOC_Os04g57860
ZmCEL12	GRMZM2G 178025_T01	TC147741	HvCEL5	Sb06g069910.01	Os04g0497200	LOC_Os04g41970
No orthologue				Sb07g020700.1	Os08g0425300	LOC_Os08g32940
No orthologue					Os12g0428200	LOC_Os12g24040

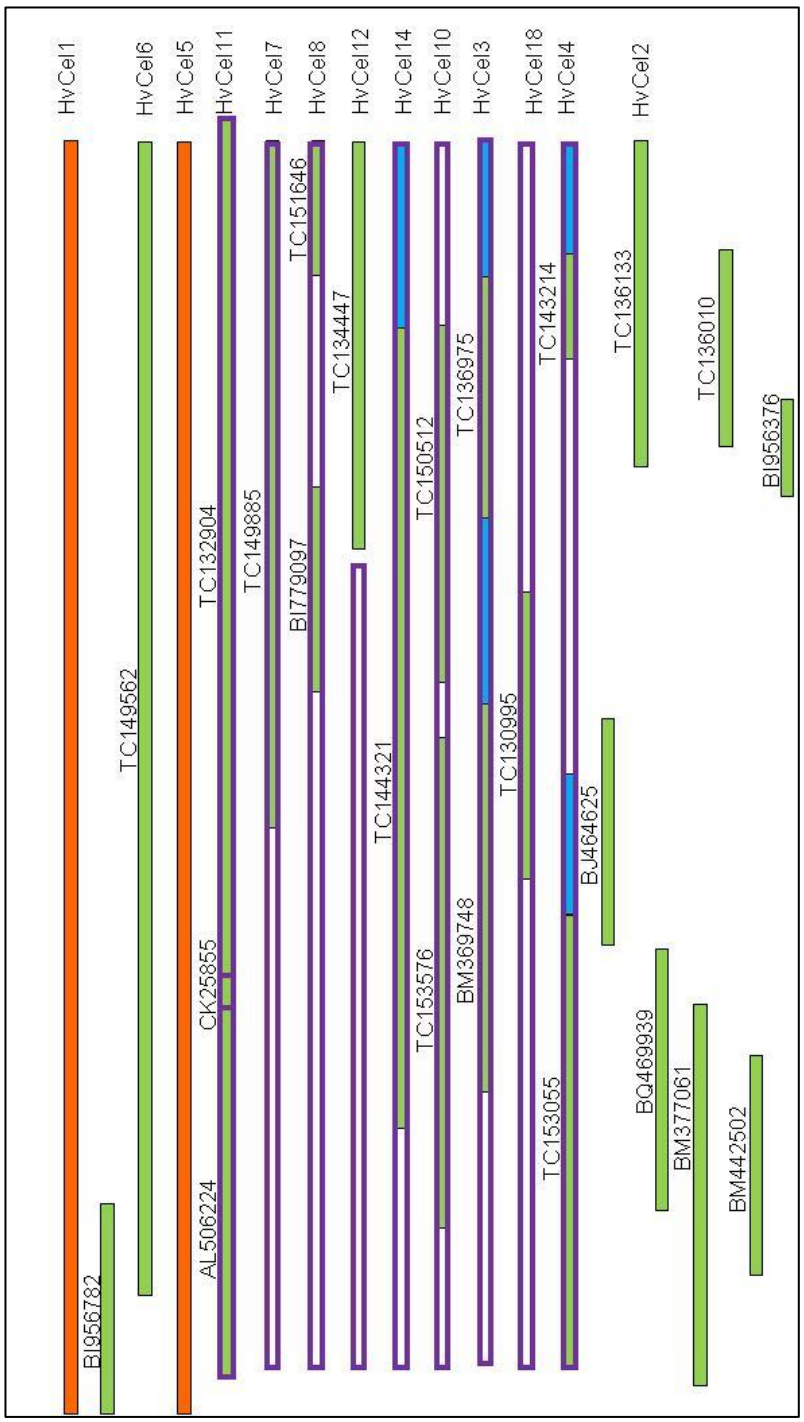

HvCel1
HvCel6
HvCel5
HvCel11
HvCel7
HvCel8
HvCel12
HvCel14
HvCel10
HvCel3
HvCel18
HvCel4
HvCel2
TC151646
TC134447
TC150512
TC136975
TC143214
TC136133
TC136010
BI956376
TC149885
BI779097
TC144321
TC153576
BM369748
TC130995
BJ464625
TC149562
TC132904
CK25855
AL506224
BI956782
BQ469939
BM377061
BM442502
TC153055

Figure 2-4: Results of barley sequence identification.

Each bar represents a gene, EST or contig. The numbers above the bars refer to the TIGR contig number, as retrieved from the TIGR database and names on the right of the bars are arbitrarily assigned gene names. Orange bars represent full length ORFs obtained from public database searches; green bars represent the contigs and ESTs from public database searches; blue bars represent sequence obtained using 3' SMART RACE amplification; purple outlines show sequence obtained from Morex barley BAC library.

codons exhibiting differences in the "3rd base wobble" position, indicating that the two sequences were representing two different genes. 3'SMART RACE was used as a means of increasing the length of available sequence in the 3' direction. It was performed using GSP forward primers and UPM reverse primers (Clontech) as presented in Table 1, Appendix A. Total ribonucleic acid (RNA) was prepared from tissues in which the TIGR database had indicated the ESTs and contigs had been found. These included maturing barley caryopsis at 4, 6, 8 and 12 days after pollination (DAP), rachis at 4 and 12 DAP, stem, developing embryo (DE), scutellum, pistil and roots of 4 day old germinating seeds. An RNA "cocktail" comprising 1 µl caryopsis, 2 µl DE, 5 µl stem and 5 µl rachis was prepared as well as using separate tissue RNA preparations for cDNA production. In most cases, several primers and cDNA preparations were used in attempts to amplify gene sequence. This method proved successful for ESTs BM369748 and TC144321. The successful PCR using the BQ469939 primers when sequenced proved to be TC153055 (Figure 2-4).

After cloning into the pGEM-T Easy vector and transformation into DH5α *E. coli* cells, all three successfully amplified products were sequenced with the M13F and M13R primers. As an example, Figure 2-5 shows the results for BQ469939 as visualised on agarose gels.

Because the number of full length genes obtained from the public databases was limited, a Morex barley BAC library was screened. A total of 90 clones produced positive results from the screening. Due to the difficulty in determining the exact location of each clone on the membrane, 225 clones were selected for further analysis.

The 225 selected clones were grown on solid agar plates with chloramphenicol. From the three 96-well plates, row and column pools were produced which were pooled to make a megapool for each plate (Table 2-2). These megapools were tested for the presence of endo-(1,4)-β-glucanase genes by PCR using the Platinum Taq polymerase. A full list of primers used for the BAC screening can

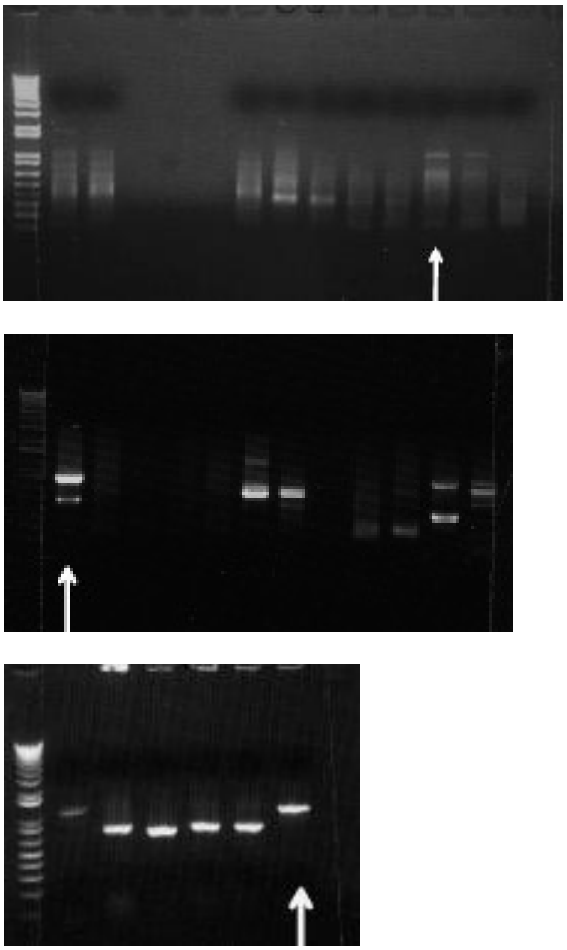


Figure 2-5: Barley 3' SMART RACE.

1% agarose gels showing a representative example of results for 3' SMART RACE PCR for EST BQ469939. White arrows indicate the relevant lane . Top Panel: the first round of PCR used a developing embryo derived cDNA template with the primer set 469939QF and UPM. This produced two main but faint bands seen on the agarose gel. Middle Panel: A second round of touchdown PCR was performed using "nested" primers, 469939F2 and NUP, situated 3' of the first round primers. A 1:50 dilution of the round one PCR was used as template for this second PCR. Products from the second round of PCR were cloned into the pGEM-T vector (Invitrogen) Bottom Panel: Colony PCR was performed on six different colonies produced after transformation of DH5α *E. coli* cell. The clone selected for sequencing is indicated by the white arrow.

Table 2-2: Pooling clones of a barley BAC library.

The combining of clones from the 96-well plate into row and column pools is illustrated in this diagram. Row and column pools containing 10ul from each row well, A-H, and 10ul fom each column well, 1-12, were made and one megapool, comprising 5ul from each row and column pool was produced.

COLUMNS 1-12 → / ROWS A-H ↓	1	2	3	4	5	6	7	8	9	10	11	12	ROW POOLS ↓
A	1	2	3	4	5	6	7	8	9	10	11	12	A
B	13	14	15	16	17	18	19	20	21	22	23	24	B
C	25	26	27	28	29	30	31	32	33	34	35	36	C
D	37	38	39	40	41	42	43	44	45	46	47	48	D
E	49	50	51	52	53	54	55	56	57	58	59	60	E
F	61	62	63	64	65	66	67	68	69	70	71	72	F
G	73	74	75	76	77	78	79	80	81	82	83	84	G
H	85	86	87	88	89	90	91	92	93	94	95	96	H
COLUMN POOLS →	1	2	3	4	5	6	7	8	9	10	11	12	MEGAPOOL

be found in Table 3, Appendix A. For initial screening, primer pairs amplifying products from 3' UTR regions of the genes and gene probes were tested. These primer pairs had been checked for the amplification of a single product using cDNA, and the 3' UTR primers had been used successfully for quantitative real time PCR (QPCR). Some ESTs did not have 3' UTR sequences, so primers were designed for regions of low identity to other cellulase genes. Initially, the megapools were used as a template to indicate whether the BAC screen had highlighted actual endo-(1,4)-β-glucanase genes. Following successful megapool amplification, clones were tested by PCR using the column and row pools as template (Table 2-2). Each primer pair was tested on the 12 column pools and 8 row pools and the results of the two sets of PCR amplification enabled the clone containing the sequence to be identified. Two representative genes identified by using these pools have been described in Figure 2-6. A total of 19 primer pairs from genes, contigs or ESTS were used to test the clones for the presence of endo-(1,4)-β-glucanase genes. Clones for which a positive amplification product was obtained were grown and plasmid isolated. The product was reamplified from the BAC clone and subsequently sequenced by AGRF.

There were 10 positive BAC clones sent for sequencing for the following genes, contigs and ESTs: *HvCEL3*, *HvCEL4*, *HvCEL5 HvCEL7*, *HvCEL8*, *HvCEL10*, *HvCEL11*, *HvCEL12*, *HvCEL14*, *TC153576*, *BI779097*, *BQ469939* and *TC153055*. *TC153576* was on the same BAC as *HvCEL10*, and *BI779097*, *BQ469939* and *TC153055* were on the same BAC as *HvCEL8*. Once the BAC validation had been performed, stab cultures were sent by post to the DuPont-Pioneer Field Laboratories in Wilmington, Delaware, USA for Sanger sequencing.

The sequencing resulted in full length gene sequence being obtained for eight genes. The *HvCEL12* sequence is missing the intron and exon located at the 3' end of the gene. Many attempts to amplify this section for sequencing proved fruitless. Results from sequencing of the *HvCEL5* (BAC 42) came back as

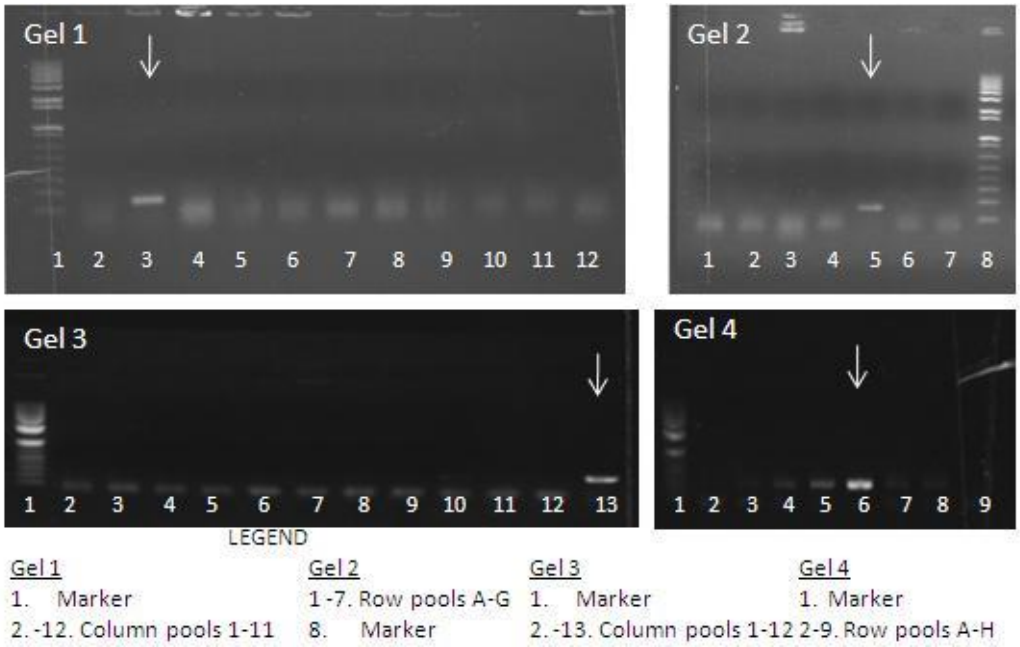


Figure 2-6: Two examples of the use of row and column pooling to locate clones from positive PCR amplification.

Gels1 and 2 are results of PCR using primers for HvCel4 from 96-well plate No. 3; Gel1 shows column pools with positive band in column 2 and Gel2 shows row pools with a positive band in row E. Thus a positive amplification product from Column 2 and Row E indicates, from Table 2.3, that clone 50 contains HvCel4. Gels 3 and 4 are results of PCR for primers for TC130995 from 96-well plate No. 1. This primer set was from a contig without 3' UTR. Positive amplification products were obtained from column 12 and row E indicating from Table 2.3 that clone 60 should contain the contig TC130995.

TC130995 (BAC60). This may be attributed to sample mix-up rather than incorrect BAC selection as all amplified products were sequenced and found to be as expected. The alignment of the *BQ469939* EST with the BAC sequence, as with the 3' SMART RACE, was similar to, but not identical with the BAC sequence with which it aligned. The resulting barley endo-(1,4)-β-glucanase gene sequences can be found in Appendix C and includes BAC sequence, derived ORFs and protein amino acid sequence. Figure 2-4 illustrates the progression of sequence improvement for barley, from initial database search to BAC sequencing. The final count for barley genes was 10 full length and two almost full-length gene sequences, and eight contigs and ESTs. Two of the full length sequences, HvCel1 and 5 are cDNA only.

2.3.1.3. Maize endo-(1,4)-β-glucanase gene sequence identification

All of the maize sequence data was found on the Dupont-Pioneer databases, which contain a large collection of cDNA proprietary sequences and a collection of public and proprietary, partially sequenced and unordered, B73 maize BACs.

Using the rice, barley and *Arabidopsis* sequence already at hand, the Dupont-Pioneer cDNA database was searched. A total of 35 maize full length ORF, contig and EST cDNAs were found; these were reduced to 28, since there were five duplicated sequences and one triplicated sequence found by alignment in ContigExpress and ClustalX2.

A BLAST of the maize B73 BAC library was performed using cDNA from maize ESTs and contigs and each BAC containing a cellulase gene was identified. This search found a total of 29 genes (Table 2-1). Evidence of homoeologues from the allotetraploidy event or segmental duplications was found for five pairs of genes. The following pairs of genes had sequence identity of more than 90% and were found on at least two map locations: *ZmCEL7* and *ZmCEL7B*, *ZmCEL8* and *ZmCEL29*, *ZmCEL14* and *ZmCEL30*, *ZmCEL25* and *ZmCEL26* and *ZmCEL12*, the latter being found in three different locations on the genome.

Table 2-3: Sub-families of the endo-(1,4)-β-glucanases in cereals and *Arabidopsis*.

Total gene numbers for barley, maize, sorghum, rice and *Arabidopsis* showing sub-families as annotated by Urbanowicz et al. (2007). One of the *Arabidopsis* genes, AtGh9B16, has been designated a pseudogene, and the barley gene numbers are incomplete.

	Barley	**Maize**	**Rice**	**Sorghum**	***Arabidopsis***
GH9A	4	7	3	3	3
GH9B	14	19	18	17	19
GH9C	2	3	4	3	3
TOTAL	**20**	**29**	**25**	**23**	**25**

2.3.2. Endo-(1,4)-β-glucanase amino acid sequence alignment

The rice, sorghum and *Arabidopsis* sequences were available on public databases as amino acid, cDNA and genomic sequences. The amino acid sequences for maize and barley were identified from cDNA translations or by removing the introns from BAC DNA sequences and translating the remaining nucleotide sequences. An alignment of the endo-(1,4)-β-glucanases from amino acid sequences was carried out using the Clustal X2 program (Larkin et al., 2007). The endo-(1,4)-β-glucanase family has been divided into sub-families on the basis of variations in protein sequences (Urbanowicz et al., 2007a). These are GH9A, or proteins with a NH_2-terminal single spanning transmembrane and catalytic domain, GH9B includes proteins with a catalytic domain, and GH9C has proteins with a catalytic domain and COOH-terminal carbohydrate binding domain. The amino acid alignment enabled the proteins to be grouped into their sub-families according to Urbanowicz et al. (2007). Moreover, an alignment of the available protein sequences from barley, maize, rice, sorghum and *Arabidopsis* provided a means of determining patterns of orthology by way of a phylogenetic tree.

2.3.2.1. Endo-(1,4)-β-glucanase sub-families

The allocation of genes to their respective sub-families is tabulated in Table 2-3. For several of the barley ESTs a significant amount of sequence is not available, which makes it difficult to place them correctly into sub-family groups.

Apart from enabling a sub-family head count, the alignment also made possible some further observations. The endo-(1,4)-β-glucanase catalytic site is recognized by a DAGD sequence as outlined in Chapter 1 and several unexpected sequence divergences were found at the DAGD active site.

1. DSGD- ZmCEL15, Sb09g002490, Os05g0129200; orthologues of sub family GH9B

2. DSGD- ZmCEL4, Sb06g032760, Os04g0674800; orthologues of sub family GH9C
3. DGGD- ZmCEL16, Sb06g017600, Os04g0443300; orthologues of sub-family GH9B
4. DAGG- Sb01g036480
5. DGGS- AtGH9A2.
6. NASD- TC136010
7. DGGG- HvCEL6

The DSGD and DGGD variations are not found in *Arabidopsis* and these variations are in orthologous sets of genes in the cereals, so it is likely that the amino acid substitution occurred after separation of monocots from dicots, but before separation of the various cereals.

The AtGH9A4 is classified as an enzyme with a transmembrane domain (Urbanowicz et al., 2007a), however it aligns with GH9B sub-family of genes. These authors report that this gene contains a transmembrane domain without a cytosolic tail. The SignalP website (Bendtsen et al., 2004) was used to predict subcellular localization of all of these genes, however, results for AtGH9A4 were uncertain as it predicted (probability = 0.5) that the first 30 aa contained a signal peptide. Further investigation showed that although the first 30 aa were hydrophobic, they were not of a hydrophobicity to warrant classification as a transmembrane domain (Personal communication: Maria Hrmova, University of Adelaide). The maize and rice orthologues of AtGH94A, when analysed with the SignalP program, contain NH_2-terminal signal peptides with a certainty probability of > 0.9.

Os12g0428200 has no DAGD motif and is significantly larger by at least 54 aa than other GH9 endo-(1,4)-β-glucanases. This sequence aligns only partially at the 3' end with endo-(1,4)-β-glucanases but does not BLAST to any other gene family. Despite a number of attempts to align this sequence with other GH9 endo-(1,4)-β-glucanases it is likely that this gene has been misannotated and it is the opinion of the author that it is a pseudogene.

2.3.2.2. The endo-(1,4)-β-glucanase family tree

The unrooted parsimonious phylogenetic tree generated from the amino acid sequence alignment of barley, maize, sorghum and *Arabidopsis* endo-(1,4)-β-glucanase genes is presented in Figure 2-7. This tree does not include the short barley ESTs, maize duplicated genes or Os12g0428200. The tree generated clearly does not strictly follow the structural sub-family grouping devised by Urbanowicz, et al. (2007).

On this tree, the GH9A subfamily includes clade B1 (Figure 2-7). The structure of the genes from clade B1 indicates that they are from sub-famil GH9B, but from a phylogenetic perspective they are derived from the GH9A sub-family.
The GH9C sub-family forms two clades and associated with the GH9C branch is a large clade, clade B3 (Figure 2-7), of the GH9B structural sub-family of genes. Clade B3 has just one *Arabidopsis* gene but four (maize) or five (rice and sorghum) cereal orthologues.

So, although protein structure clearly defines three sub-families, this phylogenetic tree suggests a more complex arrangement than that suggested by Urbanowicz et al. (2007). Sub-family structure is further developed in association with the intron-exon structures of the endo-(1,4)-β-glucanase genes in section 2.3.3.

Although the total number of *Arabidopsis* and rice genes only differs by one, the clade structure indicates a number of differences. Clade B3 in Figure 2-7 contains one *Arabidopsis* gene but has four barley and maize genes and five rice and sorghum genes. These genes are located on different chromosomes, and tree branch lengths indicate they are not the result of recent duplications. Clade B4 has two maize, rice and sorghum genes each, without any *Arabidopsis* representatives. Conversely, Clade B2 contains four *Arabidopsis* genes found as a tandem set on the *Arabidopsis* genome on Chromosome 2 (SIGNAL database), with only one orthologue each from maize, sorghum and rice. Also, Clades B5, B6 and B7 each contain two *Arabidopsis* genes and only one cereal gene. Notwithstanding the tandem duplications of *Arabidopsis* endo-(1,4)-β-glucanase

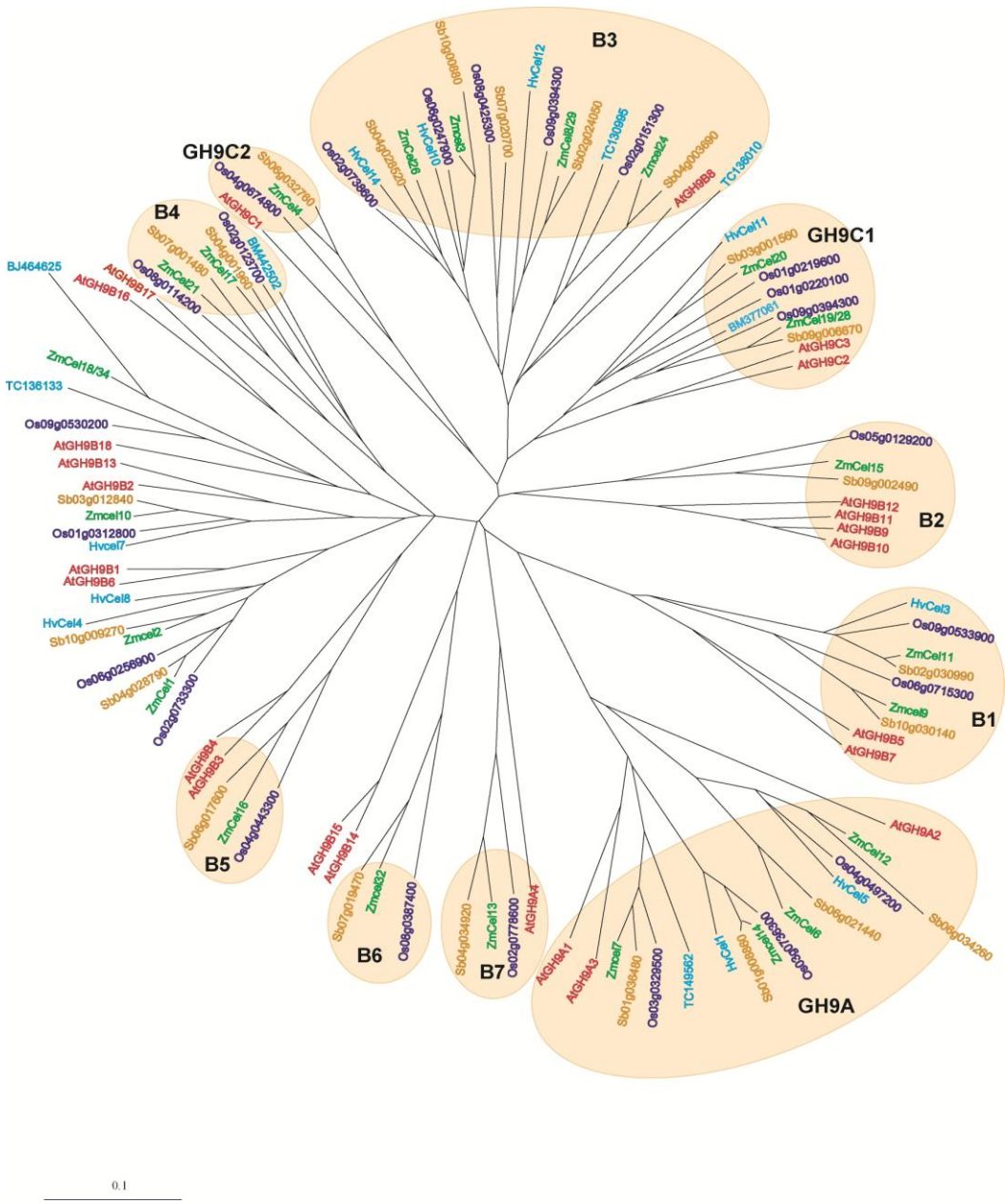


Figure 2-7: A parsimonious phylogenetic tree of cereals and *Arabidopsis*.

Included on this tree are barley (blue), maize (green), rice (purple), sorghum (tan) and Arabidopsis (red). Clades of interest as specified in the text are circled in orange. The bar in the left bottom corner provides a relative measure of branch length. It was produced in the Clustal X2 program (Larkin et al., 2007) and viewed in Treeview (Page, 1996) using 1000 repetition neighbour joining bootstrap analysis

genes just mentioned and the homoeologues and duplications in maize, there would appear to be a differential loss and gain of genes between the cereals and *Arabidopsis* that is not reflected in the numbers of genes in each sub-family. Most of the variation between cereals relates to tandem or segmental duplications that may have occurred since separation of cereals from dicots. Notwithstanding the maize homoeologues, both sorghum and maize have an additional orthologous GH9A gene compared to rice.

In order to examine more closely the relationship between the duplications and homoeologues in maize and sorghum, a second tree, again based on amino acid sequence, was produced to assess this relationship using rice as the outgroup (Figure 2-8). *ZmCEL12Chr1* was not included as the sequence is identical to that of *ZmCEL12Chr2*. The tree indicates that *ZmCEL7B*, *ZmCEL12Chr2* and *ZmCEL25* are more closely related to their sorghum orthologues than to their respective homoeologues. This suggests they may be derivatives of the allotetraploidy event and that a sorghum ancestor was the donor. On the other hand, *ZmCEL14* and *ZmCEL30* are more closely related to each other than their sorghum orthologue, which suggests a duplication rather than homoeology event. The orthologous pair, *ZmCEL8* and *ZmCEL29*, was not included as they are not complete sequence.

The use of amino acid sequences provided a basis for selection of orthologues in barley, maize, rice and sorghum. Alignment of amino acid sequence also enabled the identification of several maize sequences that may have been homoeologues or duplications. Identification was made of a clade of GH9B endo-(1,4)-β-glucanases that appeared to have diverged from the GH9C sub-family, as well as a GH9B clade that was associated with the GH9A sub-family. This was an unexpected result so position of intron splice sites, a tool that may be used to explore the relationships between clades was investigated. It was performed not only to provide further evidence of orthology as found on the phylogenetic tree, but to determine if these relationships were seen in gene structure as well as amino acid alignment.

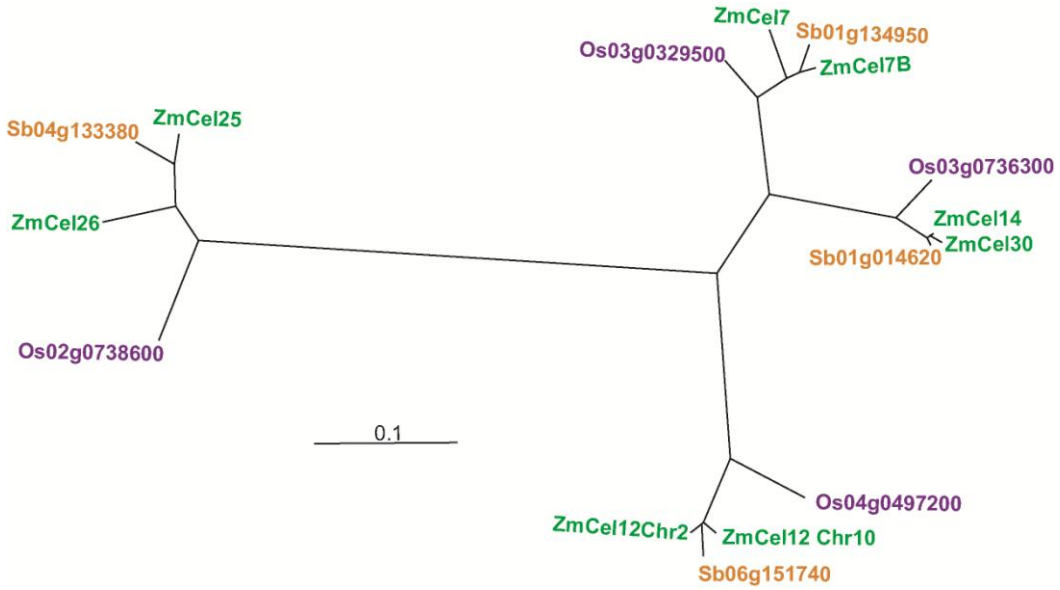


Figure 2-8: A parsimonious tree showing maize homoeologues with sorghum orthologues.

This tree includes the maize homoeologues (green), and their sorghum (orange) and rice (purple) orthologues. This tree was produced to illustrate the relative distances between the maize homoeologues and their sorghum orthologues using amino acid sequence and using rice as the outgroup. It was produced as per Figure 2-7

2.3.3. Intron and exon gene structure

Examination of intron position of endo-(1,4)-β-glucanases in cereals shows a suite of common positions. The intron/exon structure for all genes analysed can be found in Appendix D. Although there are some variations in numbers of introns, in general, there is retention of intron positions and exon sizes between those orthologues that had been determined from the phylogenetic tree. It has been suggested that where genes are not separated by large distances, splice sites and exon sizes may provide useful tools for grouping of orthologues (Irimia and Roy, 2008). Figure 2-9 shows a sequence alignment with all possible intron splice sites indicated by an arrow and a number as they relate to the amino acid sequence of representative maize endo-(1,4)-β-glucanases. Table 2-4 shows each maize gene and its respective introns and illustrates splicing patterns that are associated with the clades.

In the case of sub-family GH9A and clade B1, intron splice sites unique to these two clades provide evidence that these genes have originated from a common ancestor. This is seen by the presence of intron 1, common to all GH9A and B1 genes only, and intron 6 common to all GH9A and clade B1 genes except *ZmCEL12* and *ZmCEL6*. Figure 2-10 illustrates all of the cereal GH9A genes and the clade B1 genes with exon and intron lengths. It should be noted that exon size is not common between GH9A and clade B3. The occurrence of deletions and/or insertions since their divergence and the length of the branches on the phylogenetic tree indicate that a long period of time has elapsed since separation of these genes.

As a generalisation, while clade B1 and GH9A contain introns 1 and 6, the remainder of the endo-(1,4)-β-glucanase genes contain introns 2 and 7. A closer look at GH9C gene structure in maize indicates that intron 7 is not present, however, it is seen in three of four rice GH9C genes. This suggests loss of the intron from the GH9C genes in maize, sorghum and barley since their separation from rice. Notwithstanding the CBD, the GH9C sub-family and clade B3 all have common exon lengths, except for the first and last exon (Figure 2-11).

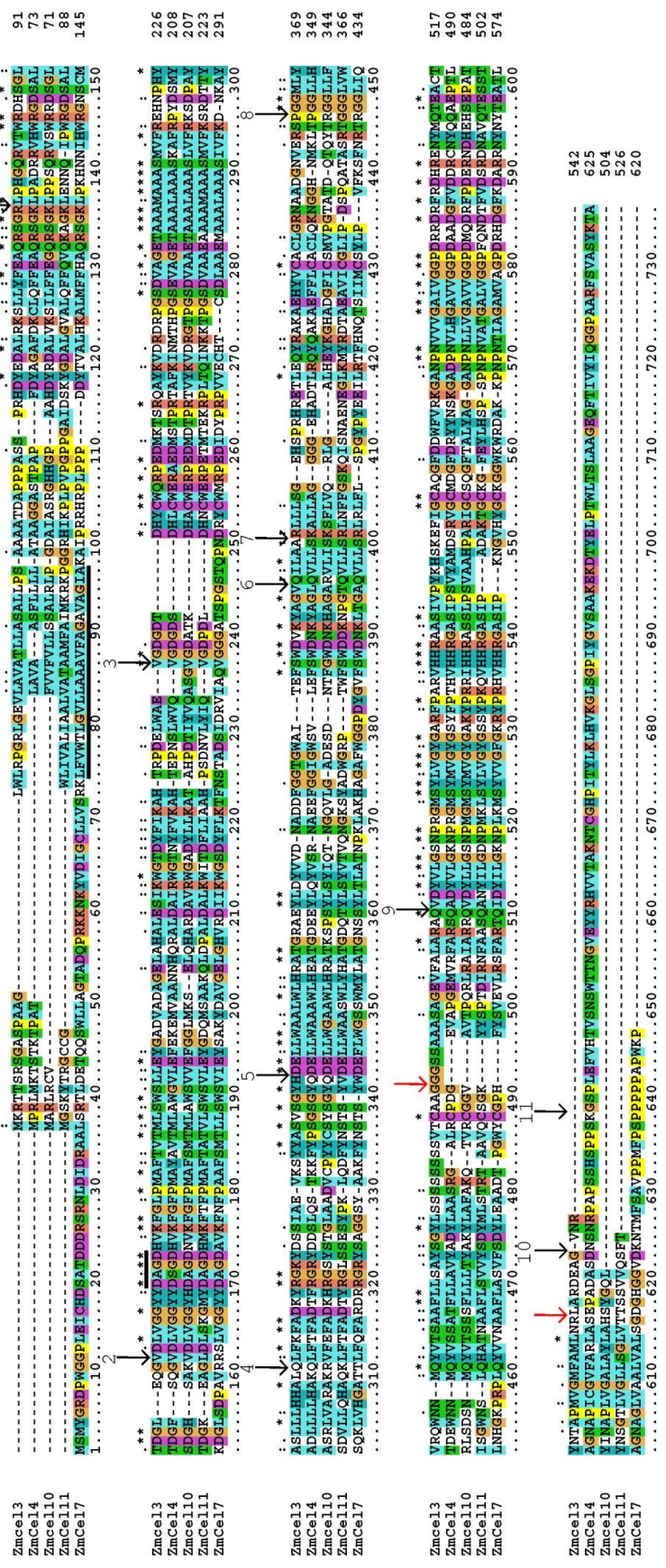

Zmcel13
ZmCel4
Zmcel10
ZmCel11
ZmCel7

Figure 2-9: The positions of the intron splice sites on an amino acid alignment of five representative maize endo-(1,4)-β-glucanases.

Black arrows with associated numbers represent the splice site positions seen in rice, maize, sorghum and barley. Red arrows are additional introns found in *Arabidopsis* (Libertini et al., 2004). The black bar under the alignment indicates the single spanning transmembrane domain (Zuo et al., 2000) and black bar over the alignment shows the catalytic site (Davies and Henrissat, 1995). *ZmCEL3*: GH9B subfamily with phylogeny to GH9C subfamily, *ZmCEL4*: GH9C subfamily, *ZmCEL10*: GH9B subfamily, *ZmCEL11*: GH9B subfamily with phylogeny to the GH9A sub-family and *ZmCEL7*: GH9A subfamily (Urbanowicz et al., 2007a).

Table 2-4: Specific introns associated with maize genes.

This table should be read in conjunction with Figure 2-9. The numbers at the top of the table are the intron numbers as they relate to the alignment in Figure 2-9. The yellow blocks indicate the introns present for the gene in that row.

Sub-family ↓	Intron number →	1	2	3	4	5	6	7	8	9	10	11
GH9A	ZmCel12											
	ZmCel6											
	ZmCel7											
	ZmCel14/30											
B1	ZmCel9											
	ZmCel11											
GH9C	ZmCel4											
	ZmCel20											
	ZmCel19/28											
B3	ZmCel24											
	ZmCel3											
	ZmCel25/26											
	ZmCel8/29											
GH9B	ZmCel15											
	ZmCel13											
	ZmCel1											
	ZmCel2											
	ZmCel16											
	ZmCel32											
	ZmCel18/34											
	ZmCel10											
	ZmCel21											
	ZmCel17											

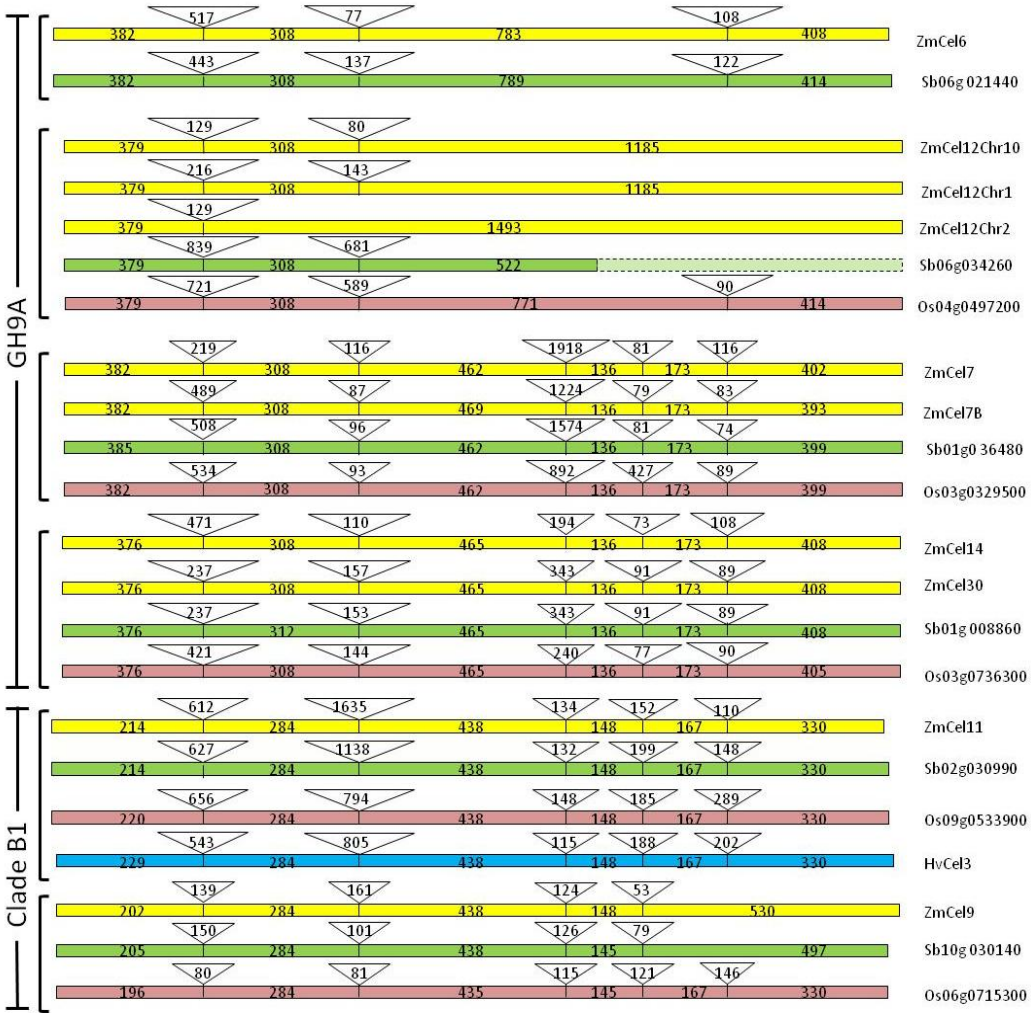


Figure 2-10: Intron/exon structure for GH9A sub-family and clade B1

Coloured bars represent genes with gene name on the right. Yellow bars are maize, green bars are sorghum, brown bars are rice and the blue bar is barley. Numbers within the bars represent exon size (nucleotides); triangles represent splice sites and intron size. Square brackets enclose separate clades of orthologous genes. Intron 1 is specific to all of these genes and intron 6 is in all but two clades, but more interestingly, from a phylogenetic perspective, is in all of the clade B1 genes. Box with dotted lines indicates that partial gene sequence is not known, eg. Sb06g151740.

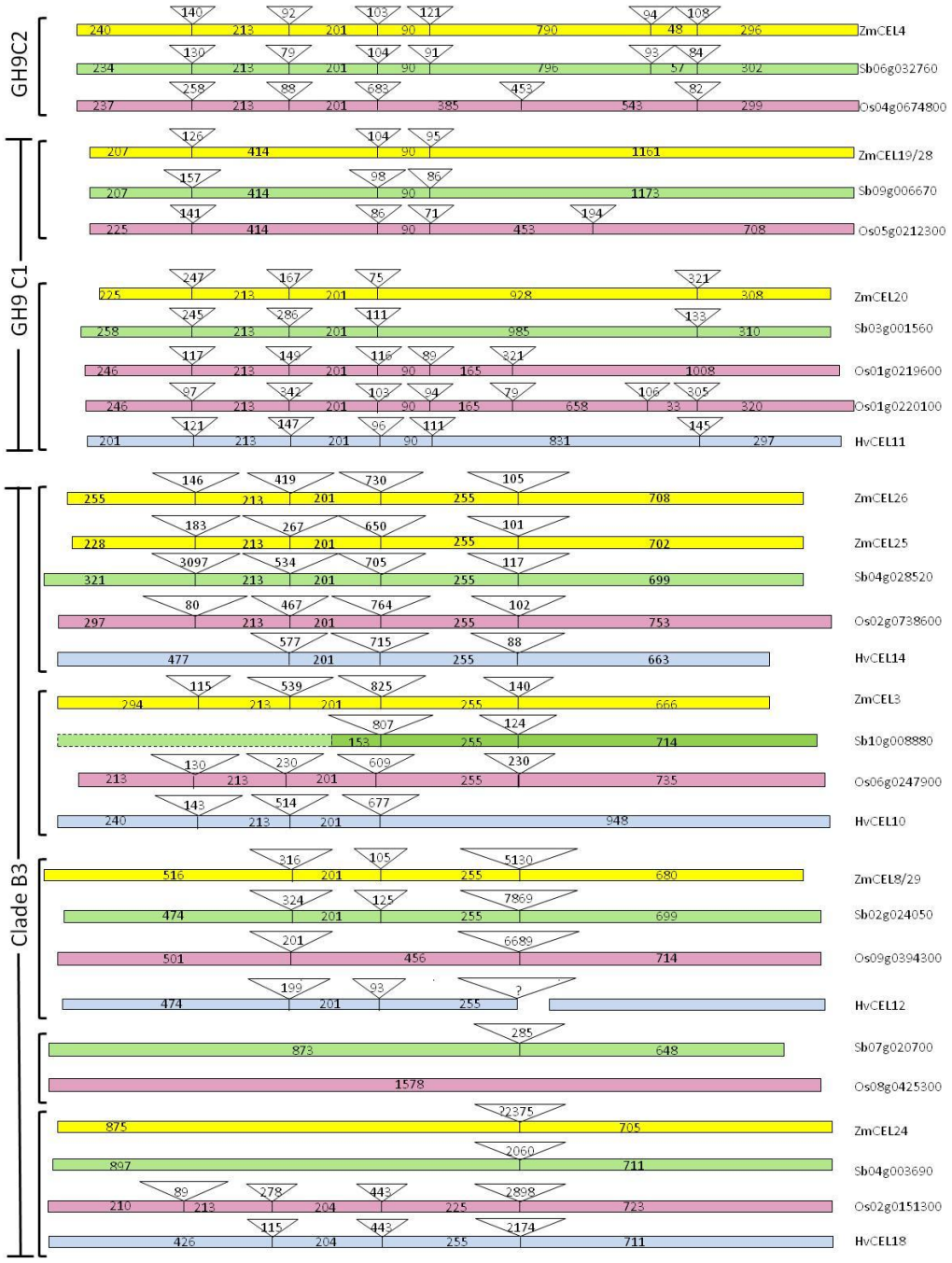


Figure 2-11: Intron/exon structure for CBD sub-family GH9C and clade B3.

As with Figure 2-10, coloured bars represent genes, with gene name to the right of the bars; yellow bars: maize, green bars: sorghum, brown bars: rice and the blue bars: barley. Numbers within the bars represent exon nucleotide size; triangles represent splice sites and intron size. Square brackets enclose separate clades of orthologous genes. Although a number of introns have been lost there is exon size commonality between and within the clades. Clades that include ZmCEL4, ZmCEL19/28 and ZmCEL20 all have a putative CBD, whereas the rest of the clades do not possess a CBD. All clades possess exons of 213, 201 and 255 (90 + 165) bps.

Where an intron has been lost, the number of bps for the resultant single exon is equal to the sum of the two exons in the orthologous or homoeologous genes. Although introns have been lost, there have been no insertions or deletions within the gene sequence. This reinforces the view that clade B3 is derived from the GH9C sub-family and time since divergence is relatively short. The clades containing orthologues of *ZmCEL15* and *ZmCEL32* also have a 213 bp second exon, which would suggest that they have originated from the same ancestor gene as the sub-family GH9C and clade B3 genes.

With regard to the *Arabidopsis* AtGH9A4 that was identified as a GH9A sub-family member by Urbanowicz et al. (2007) but clearly did not have a transmembrane domain, the number of bases in the second exon indicates that, in cereals, this clade is more closely related to the GH9C sub-family. If AtGH9A4 does indeed belong to the GH9A sub-family, gene structure suggests that its disputed N-terminal domain evolved after separation of dicots and monocots.

Intron splice sites provide a measure of gene structure and were a very useful tool in reinforcing the view that clade B1 and sub-family GH9A are relatives and also that clade B3 and GH9C are recently diverged. This analysis also showed that orthologues had almost identical gene structure, and amino acid sequence alignment provided a sound measure of orthology. The use of shared map positions on chromosomes, or gene synteny, is also a good indicator of gene orthology and is described in the next section.

2.3.4. In silico mapping of the endo-(1,4)-β-glucanase gene family in maize, sorghum and rice

Information on map positions for the rice and sorghum endo-(1,4)-β-glucanase genes was obtained by a gene name search of the *Oryza sativa* ssp indica and *Sorghum bicolour* genomes on the Gramene website. Using the B73 BAC identification codes, maize mapping data was obtained by a search of the Dupont-Pioneer and Maizesequence.org databases.

2.3.4.1. In silico mapping of endo-(1,4)-β-glucanase genes in maize

Table 2-5 shows the arbitrary gene name and maizesequence.org gene annotation, identified BACs and associated markers and approximate map position for each gene in maize. These initial positions are approximate only, as currently all of these BACs are not fully sequenced and ordered, although this changes as the website is updated with new information. There is no endo-(1,4)-β-glucanase gene evident on Chromosome 3.

A diagram of the maize chromosomes with the approximate position of each gene has been compiled (Figure 2-12). This figure shows that the position of the five homoeologous pairs is in accordance with the estimated homoeology between the two maize antecedent genomes (Wei et al., 2007). This suggests that the gene duplicates are homoeologues in all but one case, *ZmCEL12*. The *ZmCEL12* gene located on Chromosome 1 is not homoeologous with the other two *ZmCEL12* genes located on Chromosomes 2 and 10 and may be the result of a recent segmental duplication. There did not appear to be any recent tandem duplications of endo-(1,4)-β-glucanase genes in maize. Although several genes are located near one another on the map, distance between them is quite large and they are not closely related. For example, *ZmCEL1* and *ZmCEL25*, on chromosome 5 are in fact separated by 400,000 bps and are not phylogenetically related by sequence. It is, of course, possible that they are the result of the duplication event that occurred prior to divergence of the dicots and monocots.

2.3.4.2. In silico mapping of rice and sorghum endo-(1,4)-β-glucanase genes and gene synteny with maize

In rice, one recent tandem duplication has occurred since separation of rice from the other cereals, Os01g0219600 and Os01g0220100, however recent duplications were not found in sorghum and as already mentioned, only one segmental duplication of very recent origin was found in maize. Syntenous regions between rice, maize and sorghum have been more accurately identified as the result of recent availability of a near complete maize genomic sequence (Moore et al., 1995; Wei et al., 2007; Zheng et al., 2006). Regions of gene

Table 2-5: Maize endo-(1,4)-β-glucanase gene map information.

This table contains information regarding arbitrary gene name, maizesequence.org gene annotation as well as the BAC on which the gene was found, gene chromosome location, orientation, contig number and associated markers for each maize gene. For easy recognition, duplicates and homoeologues are presented in the same colour. The gene annotation on maizegenome.org is still in progress, therefore some genes have not yet been annotated and are indicated as N/A.

Gene name	BAC	Chr	EMBL BAC	Orientation	Ctg	Gene annotation	L flank marker	R Flank marker	approx. chr position (Mb)
ZmCEL12	c0140E02	1			5	GRMZM2G	umc2224	bnlg1178	12.5
ZmCEL7	c0065H02	1	209839	reverse	14	GRMZM2G 003379_T01	umc1452	bnlg2238	54.3
ZmCEL32	c0129D24	1	207872	forward	44	GRMZM2G 167669_T01	umc1848	umc1245	216.1
ZmCEL30	c0050E02	1	184871	reverse	56	GRMZM2G 147849_t01	bnlg1268	umc1885	272.7
ZmCEL12	b0144L23	2	199765	forward	74	N/A	umc1555	bnlg1064	26.3
ZmCEL8/29	c0418A05	2	207353	forward	96	GRMZM2G 703265_T01	umc1004	umc1749	175.8
ZmCEL26	c0426J19	4	185637		182	N/A	gln5	bnlg2291	166.5
ZmCEL14	b0185P18	5	195684	reverse	207	GRMZM2G 110735_T01	umc90	bnlg565	8.22
ZmCEL9	c0446C08	5	213844	reverse	220	GRMZM2G 153987_T01	umc1274	cat1	60.4
ZmCEL17	c0457K01	5	215929	forward	223	GRMZM2G 141911_T01	gtc102	umc2120	71.1
ZmCEL24	c0422H10	5	200296	reverse	225	GRMZM2G 482256_T01	bnlg1902	csu36b(rpL19)	80.1
ZmCEL1	c0495D12	5	213848	forward	251	GRMZM2G 76049_T01	umc108	umc2201	203.9
ZmCEL25	c0478O08	5	195458	forward	251	GRMZM2G 151257_01	gln4	umc2198	204.5
ZmCEL13	c0455L23	5	199482	reverse	253	GRMZM2G 393000_01	bnlg118	csu834(mss)	211.12
ZmCEL21	c0390A03	6	186810	forward	257	GRMZM2G 009025_T01	umc1143	umc1883	4.021
ZmCEL15	c0047L14	6	203408	forward	262	203408.4 FGT003	mmp163	umc2311	13.76
ZmCEL2	c0491C08	6	204519	forward	281	GRMZM2G 455642_T01	umc2317	umc1614	118.2
ZmCEL8/29	b0530D08	7	230018	reverse	309	N/A	lim333	umc1036	101.2
ZmCEL18	c0115L23	7	193413	forward	320	N/A	umc1660	uaz221(his2a)	137.2
ZmCEL11	c0206P11	7	204772	reverse	320	GRMZM2G 147422_T01	umc1660	umc1837	137.8
ZmCEL20	b0499H17	8	186574	reverse	329	GRMZM2G 453565_T01	mmp158b	rps28-R flank	26.1
ZmCEL10	c0367D10	8	203762	forward	331	GRMZM2G 331566_T01	umc1236-Rflank	umc1868-Lflank	35
ZmCEL19/28	c0003J13	8	186611	forward	349	GRMZM2G 143747_T01	umc1858	rip1	111.4
ZmCEL3	c0129H17	9	203011	reverse	368	GRMZM2G 154678_T01	-	umc1279	1.3
ZmCEL7	b0226H19	9	210056	forward	386	GRMZM2G 099101_T01	umc1654	umc1657	127.5
ZmCEL16	b0358A19	10	195855	reverse	412	GRMZM2G 125436_T01	umc1911	umc2350	118.73
ZmCEL6	c0417D13	10	205481	forward	413	GRMZM2G 066162_T01	umc2003	umc1280	127.5
ZmCEL4	c0304K02	10	200742	reverse	419	GRMZM2G 343144_T01	umc1645	bnlg1518	147.6
ZmCEL12	b0310F06	10	195329	forward	420	GRMZM2G 178025_T01	umc1556	-	149.6

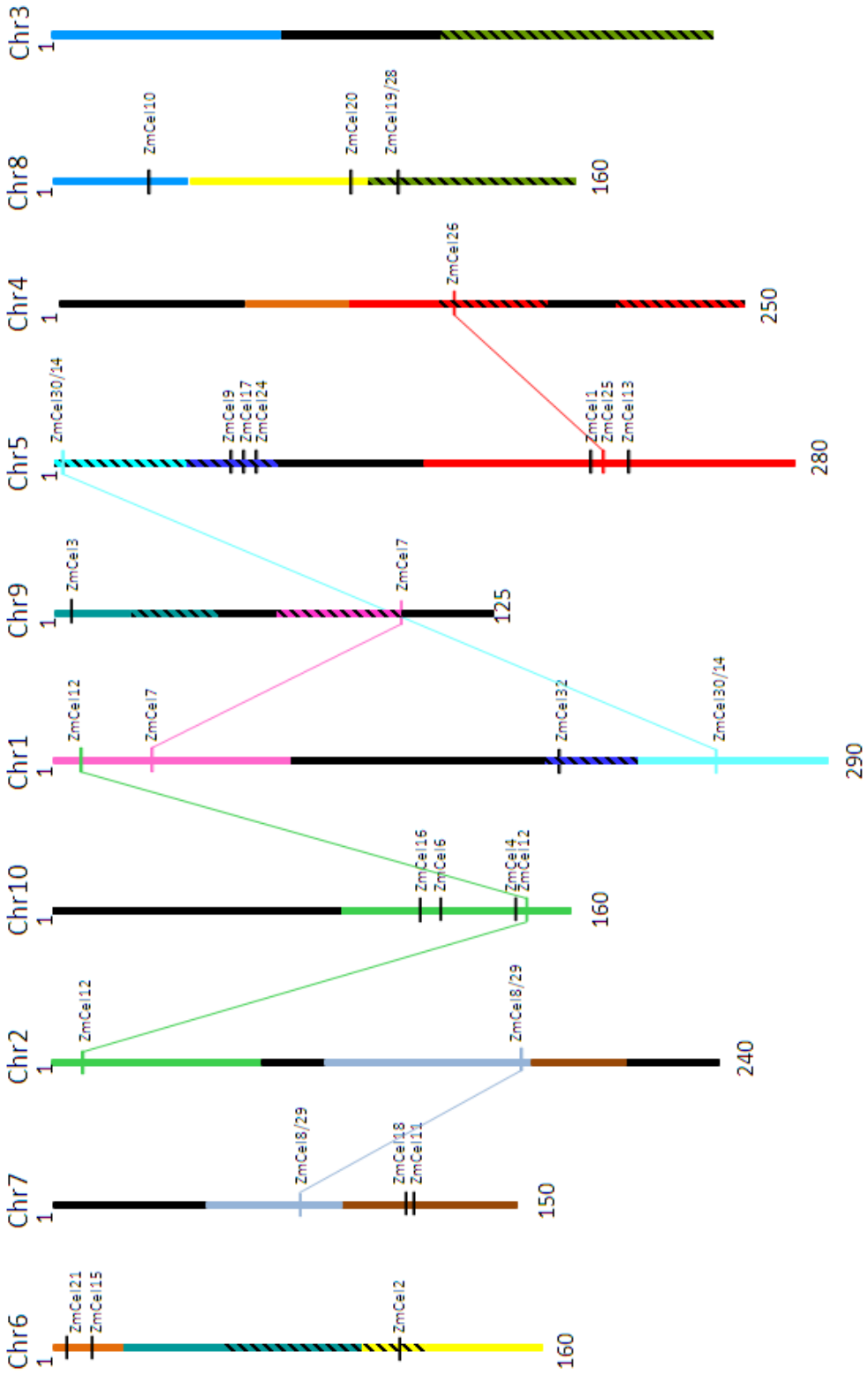

Chr3
1
Chr8
1
ZmCel10
ZmCel20
ZmCel19/28
160
Chr4
1
ZmCel26
250
Chr5
1
ZmCel30/14
ZmCel9
ZmCel17
ZmCel24
ZmCel1
ZmCel25
ZmCel13
280
Chr9
1
ZmCel3
ZmCel7
125
Chr1
1
ZmCel12
ZmCel7
ZmCel32
ZmCel30/14
290
Chr10
1
ZmCel15
ZmCel6
ZmCel4
ZmCel12
160
Chr2
1
ZmCel12
ZmCel8/29
240
Chr7
1
ZmCel8/29
ZmCel18
ZmCel11
150
Chr6
1
ZmCel21
ZmCel15
ZmCel2
160

Figure 2-12: The approximate chromosome regions of the endo-(1,4)-β-glucanase genes in maize.

Similarly coloured regions indicate areas of homoeology, as determined by linkage mapping, between the two ancient genomes for areas of the maize genome that contain endo-(1,4)-β-glucanase genes (Wei et al., 2007). Sections of the chromosomes without endo-(1,4)-β-glucanase genes are coloured black. Hatching indicates areas of possible genome reversal (Zheng et al., 2006). Coloured lines link homoeologues or duplicate endo-(1,4)-β-glucanase genes.

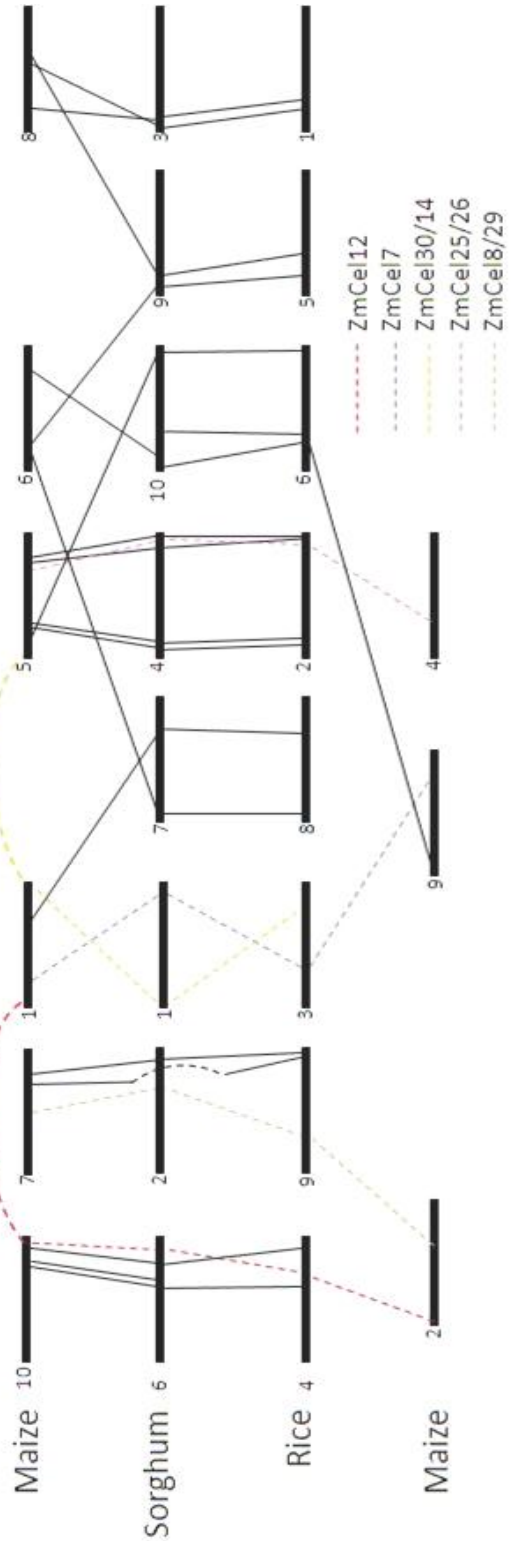

ZmCel12
ZmCel7
ZmCel30/14
ZmCel25/26
ZmCel8/29
Maize
Sorghum
Rice
Maize

Figure 2-13: Map positions of maize, rice and sorghum for the endo-(1,4)-β-glucanase orthologues.

Lines link orthologues and map positions of the three species and indicate synteny between rice and sorghum. The legend shows coloured lines linking maize homoeologues or duplicates with their rice and sorghum orthologues.

Table 2-6: Maize endo-(1,4)-β-glucanase genes with the map position of their respective sorghum and rice orthologues.

The information for this table has been obtained from DuPont-Pioneer proprietary databases, maizesequence.org and gramene.org. Approximate map positions of maize were syntenic with map positions of endo-(1,4)-β-glucanase genes in rice and sorghum (DuPont-Pioneer).

Maize	Sorghum				Rice			
			Approx. Map position				Approx. Map position	
Gene	Orthologue	Chr	From	To	Orthologue	Chr	From	To
zmcel12	Sb06g034260	6	62,203,346	62,203,693	Os04g0497200	4	24,645,224	24,648,207
ZmCel7	Sb01g036480	1	60,075,738	60,079,882	Os03g0329500	3	12,083,213	12,087,050
zmcel32	Sb07g019470	7	50,223,535	50,226,324	Os08g0387400	8	18,172,754	18,176,350
zmcel30	Sb01g008860	1	7,628,982	7,631,818	Os03g0736300	3	30,133,336	30,135,424
zmcel12	Sb06g034260	6	62,203,346	62,203,693	Os04g0497200	4	24,645,224	24,648,207
zmcel8/29	Sb02g024050	2	58,015,042	58,024,985	Os09g0394300	9	13,631,022	13,639,579
zmcel26	Sb04g028520	4	58,583,956	58,587,307	Os02g0738600	2	30,833,601	30,836,519
zmcel14	Sb01g008860	1	7,628,982	7,631,818	Os03g0736300	3	30,133,336	30,135,424
zmcel9	Sb10g030140	10	59,861,788	59,863,719	Os06g0715300	6	30,363,328	30,365,358
zmcel17	Sb04g001960	4	1,682,066	1,683,482	Os02g0123700	2	1,237,215	1,240,044
zmcel24	Sb04g003690	4	3,511,299	3,514,867	Os02g0151300	2	2,824,588	2,828,457
zmcel1	Sb04g028790	4	58,890,786	58,893,705	Os02g0733300	2	30,567,900	30,571,063
zmcel25	Sb04g028520	4	58,583,956	58,587,307	Os02g0738600	2	30,833,601	30,836,519
zmcel13	Sb04g034920	4	64,755,651	64,757,729	Os02g0778600	2	32,956,178	32,957,953
zmcel21	Sb07g001480	7	1,592,364	1,593,863	Os08g0114200	8	756,998	758,594
zmcel15	Sb09g002490	9	2,699,327	2,701,088	Os05g0129200	5	1,685,202	1,687,100
zmcel2	Sb10g009270	10	10,075,856	10,079,312	Os06g0256900	6	8,177,992	8,183,625
zmcel8/29	Sb02g024050	2	58,015,042	58,024,985	Os09g0394300	9	13,631,022	13,639,579
zmcel18					Os09g0530200	9	20,763,892	20,767,408
zmcel11	Sb02g030990	2	65,993,043	65,996,795	Os09g0533900	9	20,976,416	20,979,046
zmcel20	Sb03g001560	3	1,421,808	1,424,505	Os01g0219600	1	6,538,776	6,543,149
zmcel10	Sb03g012840	3	15,486,902	15,488,691	Os01g0312800	1	11,739,859	11,741,724
zmcel19/28	Sb09g006670	9	10,779,589	10,781,813	Os05g0212300	5	6,921,443	6,923,821
zmcel3	Sb10g008880	10	9,326,067	9,328,116	Os06g0247900	6	7,670,985	7,672,456
zmcel7	Sb01g036480	1	60,075,738	60,079,882	Os03g0329500	3	12,083,213	12,087,050
zmcel16	Sb06g017600	6	47,001,813	47,003,536	Os04g0443300	4	21,894,527	21,896,805
zmcel6	Sb06g021440	6	50,717,335	50,719,872	Os04g0497200	4	24,645,224	24,648,207
zmcel4	Sb06g032760	6	60,858,880	60,860,593	Os04g0674800	4	34,231,500	34,235,279
zmcel12	Sb06g069910	6	62,203,346	62,203,693	Os04g0497200	4	24,645,224	24,648,207
	Sb07g020700		53,568,977	53,570,779	Os08g0425300	8	20,296,621	20,298,105
No sorghum or maize orthologue					Os01g0220100	1	6,563,046	6,564,879
					Os12g0428200	12	13,658,966	13,660,907

synteny between the rice and maize genomes were determined using DuPont-Pioneer proprietary data. Table 2-6 was compiled to show map positions for the orthologues of maize in sorghum and rice. This information was used to produce a diagram showing map positions between the three species (Figure 2-13) and indicates that, superficially, the rice and sorghum endo-(1,4)-β-glucanase genes are similarly located between the two genomes. Although there have been many reports of local gene rearrangements when analysing small segments of the genome, at the macro level of investigation performed here, map positions of the genes are in keeping with the gene orthology described earlier. Due to the complexity of the maize genome, where significant gene loss and genome rearrangement has occurred as seen in Figure 2-1, the synteny is not obvious (Messing et al., 2004; Wei et al., 2007).

The mapping information, in general, agrees with the phylogenetic tree created from the amino acid sequences of the endo-(1,4)-β-glucanase genes. There is however an interesting situation that has come to light. On Chromosome 10, the GH9A genes, *ZmCEL6* and *ZmCEL12* (Figure 2-10), have an orthologue each with sorghum. There is, however, only one rice orthologue, the amino acid sequence of which aligns more closely with *ZmCEL12* than *ZmCEL6*, although intron analysis suggests that *ZmCEL6* is more closely related to rice. These two genes are separated by 22 Mb suggestive of an earlier duplication and local genome rearrangement. It is not possible to determine if the duplication and rearrangement occurred before or since separation of the rice and maize/sorghum antecedents. If *ZmCEL12* is the rice orthologue, the rearrangement occurred after separation. If *ZmCEL6* is the rice orthologue, the genome rearrangement may have occurred before separation.

Considering that maize was once an allotetraploid it is not surprising that five maize endo-(1,4)-β-glucanase genes have duplications or homoeologues. When amino acid sequences of three of these pairs were aligned with sorghum, the suggestion that a sorghum antecedent was responsible for the homoeologues was borne out. In all but one homoeologous pair, one of the maize genes was closer

to the sorghum gene than its duplicate/homoeologue. However, *ZmCEL14* and *ZmCEL30* were more closely aligned to each other than to the sorghum orthologue. Performing an evolutionary analysis of the endo-(1,4)-β-glucanase genes, was one way of more accurately defining whether and when genes duplicated, or whether they are homoeologues and also more clearly define the role of the sorghum antecedent.

2.3.5. Synonymous substitution and evolutionary distance

To explore the evolutionary distances and rates of evolutionary change within sub-families and between orthologues, codon based pairwise synonymous and non-synonymous substitution rates were calculated using two different models, as described in 2.2.12. See Appendix E for results of all analyses which included synonymous substitution rates (d_S), non-synonymous substitution rates (d_N) and the ratio of d_N to d_S, as a measure of relative evolutionary pressure being exerted on a gene pair. The number of synonymous changes per synonymous site per year, or K_S, that have occurred since separation of maize and sorghum from rice, and rice and barley from maize was able to be estimated. This provided a means of estimating the number of years since separation of maize from sorghum, and barley from rice.

2.3.5.1. Synonymous substitution rates for maize:sorghum orthologues

The rate of synonymous substitution is a measure of the relative time since the genes have separated. Results of the d_S comparison between the Nei and Gojobori and PAML codeml models are presented graphically in Figure 2-14. Results indicate that for some genes and gene groups, the two different analyses are similar. The *ZmCEL7* and *ZmCEL14/30* gene set and *ZmCEL11* gene pair show very little variation between the two models whereas all other genes had much greater synonymous substitution rates with the PAML analysis. In cases where there is a difference, the PAML codeml estimations are up to five times larger than for Nei-Gojobori. Variations between gene pairs for both models are also present. *ZmCEL21* had a synonymous substitution rate up to 5 times higher than that observed for other gene pairs.

Estimates of synonymous substitutions per synonymous site per year (K_S) were calculated using the synonymous substitutions per synonymous site estimations and assuming that rice and barley separated from maize and sorghum approximately 50 mya (Swigoňová et al., 2004). This allowed the synonymous substitution data to be "normalised". The results of this analysis are shown graphically in Figure 2-14 B and indicate that K_S estimates are not constant for all genes, that is, different genes have different rates of synonymous substitution. The "normalisation" process shows three approximate levels of substitution rates. Using the PAML codeml model, an annual substitution rate of 5-7.7 x 10^{-9} synonymous substitutions per synonymous site per year was observed with eight gene pairs, a second level containing 18 gene pairs was estimated at between 11.2 and 17.4 x 10^{-9} and a third level of substitution above 20 x 10^{-9}.

With knowledge of the K_S estimate for each gene, it was possible to next estimate the time of separation of the maize homoeologues and the orthologues of maize and sorghum. This is presented graphically in Figure 2-14 C. Because the Nei and Gojobori model involves a larger number of assumptions than the PAML codeml model the remaining data is presented for the latter model only. The homoeologous pair *ZmCEL12CHR10* and *ZmCEL12CHR2* separated most recently, at around 5 million years ago. The majority of orthologous genes have an estimated separation time of 10 to 20 mya. The gene pair *ZmCEL14* and *ZmCEL30* has an estimated separation time of 25 mya. This is unexpected because the two genes show very similar sequence, but reinforces the result of the phylogenetic tree of sorghum and the maize orthologues (Figure 2-8) that indicated *ZmCEL14* and *ZmCEL30* to be duplicates rather than homoeologues. *ZmCEL21* and its orthologous sorghum gene appear to have diverged perhaps 30 to 35 mya.

2.3.5.2. Synonymous substitution rates between barley and rice orthologues

As with the maize:sorghum analyses, rice and barley were analysed using the two

Figure 2-14: Codon based evolutionary distances for maize and sorghum genes.

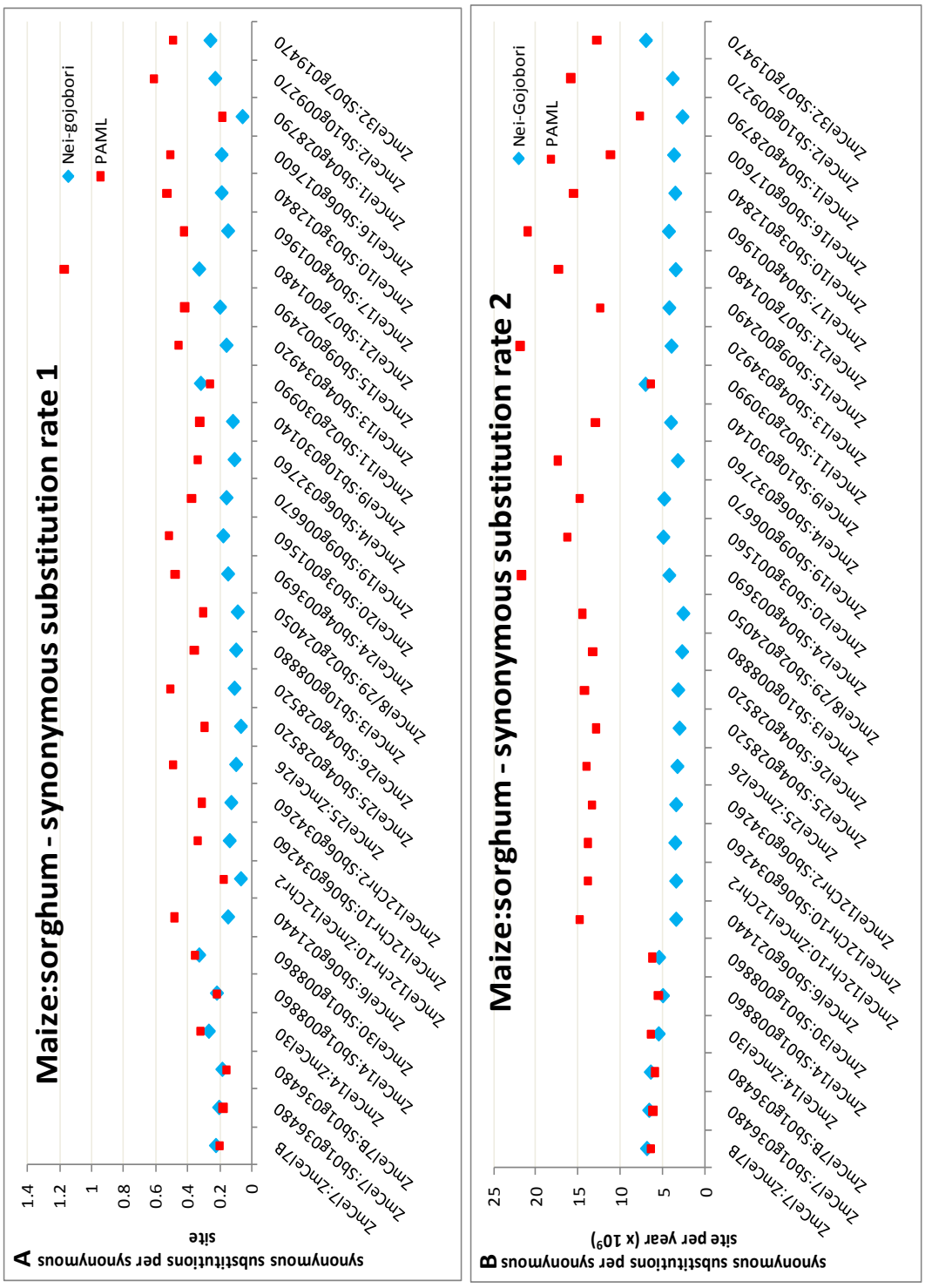
(Following pages) Estimates of synonymous substitutions per synonymous site per year (K_S) were calculated using the synonymous substitutions per synonymous site estimations and assuming that rice and barley separated from maize and sorghum approximately 50 mya (Swigoňová et al., 2004). This allowed the synonymous substitution data to be "normalised". Maize:sorghum synonymous substitution rates were estimated using the Nei and Gojobori (1986) model (blue) and PAML codeml (red). Graphs A and B illustrate the differences resulting from using the two models.

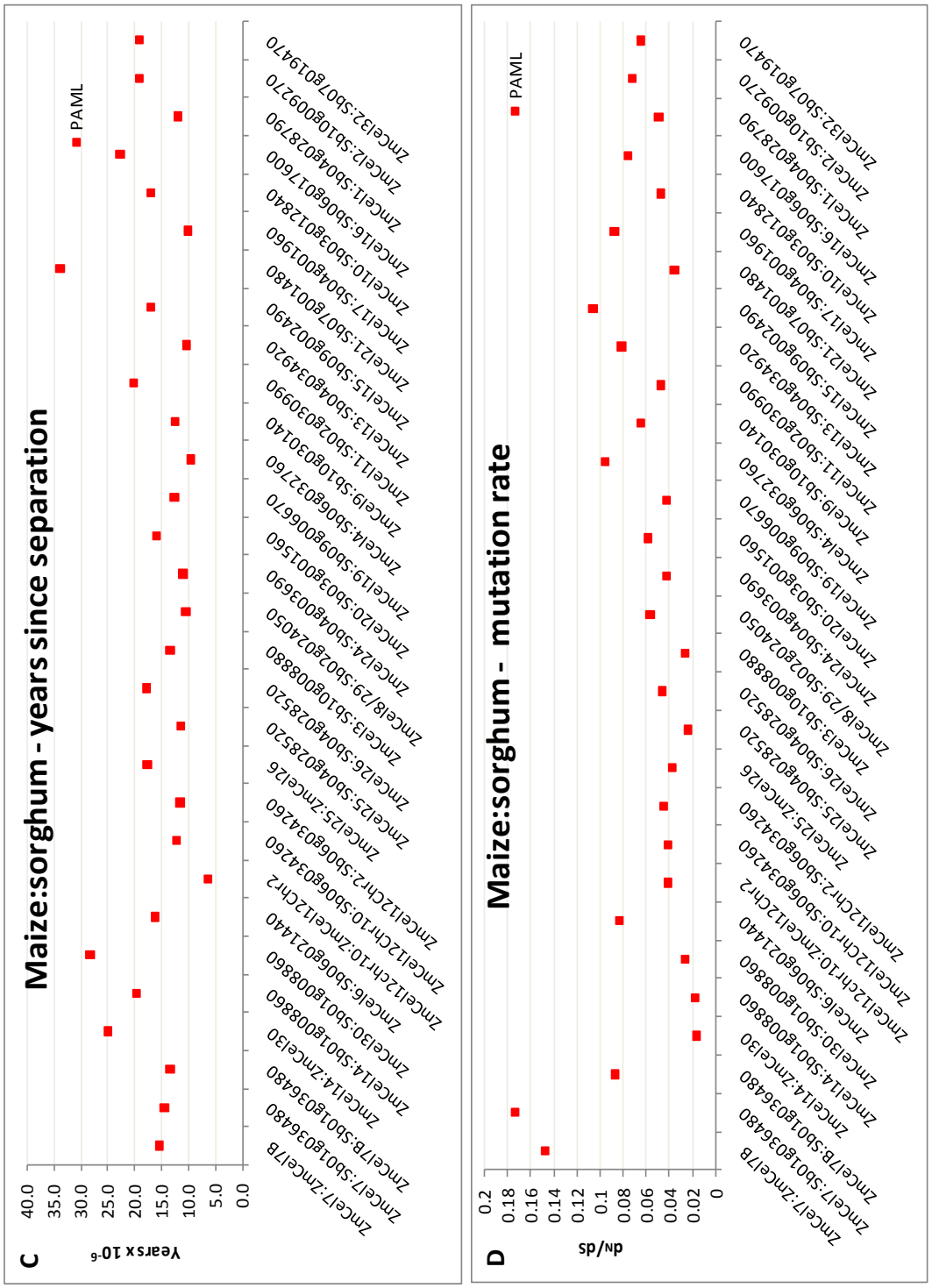
A: synonymous substitutions per synonymous site between maize and sorghum orthologues and maize homoeologues.

B:, synonymous substitution rate per synonymous site per year, using an estimated separation from rice of 50 mya.

C: estimations of the years since separation of the maize homoeologues and maize:sorghum orthologues.

D: estimations of the relative mutation rates between the maize homoeologues and the maize:sorghum orthologues.


A
Maize:sorghum - synonymous substitution rate 1
Nei-gojobori
PAML
synonymous substitutions per synonymous site
1.4
1.2
1
0.8
0.6
0.4
0.2
0
B
Maize:sorghum - synonymous substitution rate 2
Nei-Gojobori
PAML
synonymous substitutions per synonymous site per year (x 10^9)
25
20
15
10
5
0
GRMZM2G178478_P01
GRMZM2G178478_P02
GRMZM2G178478_P03
GRMZM2G178478_P04
GRMZM2G178478_P05
GRMZM2G178478_P06
GRMZM2G178478_P07
GRMZM2G178478_P08
GRMZM2G178478_P09
GRMZM2G178478_P10
GRMZM2G178478_P11
GRMZM2G178478_P12
GRMZM2G178478_P13
GRMZM2G178478_P14
GRMZM2G178478_P15
GRMZM2G178478_P16
GRMZM2G178478_P17
GRMZM2G178478_P18
GRMZM2G178478_P19
GRMZM2G178478_P20
GRMZM2G178478_P21
GRMZM2G178478_P22
GRMZM2G178478_P23
GRMZM2G178478_P24
GRMZM2G178478_P25
GRMZM2G178478_P26
GRMZM2G178478_P27
GRMZM2G178478_P28
GRMZM2G178478_P29
GRMZM2G178478_P30
GRMZM2G178478_P31
GRMZM2G178478_P32


C
Maize:sorghum - years since separation
Years x 10-6
PAML
40.0
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
D
Maize:sorghum - mutation rate
dN/dS
PAML
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
GRMZM2G017564_T01:Zmce17:Sb01g036480
GRMZM2G017564_T01:Zmce17B:Sb01g036480
GRMZM2G014124_T01:Zmce14:Sb01g008860
GRMZM2G014124_T01:Zmce14:Zmce130
GRMZM2G014124_T01:Zmce16:Sb01g008860
GRMZM2G014124_T01:Zmce130:Sb01g008860
GRMZM2G014124_T01:Zmce12Chr10:Sb06g021440
GRMZM2G014124_T01:Zmce12Chr10:Zmce12Chr2
GRMZM2G014124_T01:Zmce12Chr2:Sb06g034260
GRMZM2G014124_T01:Zmce125:Sb04g028520
GRMZM2G014124_T01:Zmce125:Zmce126
GRMZM2G014124_T01:Zmce126:Sb04g028520
GRMZM2G014124_T01:Zmce13:Sb10g008880
GRMZM2G014124_T01:Zmce18/29:Sb02g024050
GRMZM2G014124_T01:Zmce20:Sb04g003690
GRMZM2G014124_T01:Zmce19:Sb03g001560
GRMZM2G014124_T01:Zmce14:Sb09g006670
GRMZM2G014124_T01:Zmce19:Sb10g032760
GRMZM2G014124_T01:Zmce11:Sb02g030140
GRMZM2G014124_T01:Zmce13:Sb04g030990
GRMZM2G014124_T01:Zmce15:Sb04g034920
GRMZM2G014124_T01:Zmce21:Sb09g002490
GRMZM2G014124_T01:Zmce17:Sb07g001480
GRMZM2G014124_T01:Zmce10:Sb04g001960
GRMZM2G014124_T01:Zmce16:Sb03g012840
GRMZM2G014124_T01:Zmce11:Sb06g017600
GRMZM2G014124_T01:Zmce12:Sb04g028790
GRMZM2G014124_T01:Zmce132:Sb10g009270
GRMZM2G014124_T01:Zmce132:Sb07g019470

different models, Nei-Gojobori and PAML codeml. (Nei and Gojobori, 1986; Yang and Nielsen, 2000). Because the comparative results were similar to those of the maize:sorghum estimates, only the results for the PAML codeml estimations have been presented for the d_S and K_S estimates (Figure 2-15). As with the maize:sorghum results, the Nei and Gojobori model estimated lower d_S and K_S for most genes but were similar for three gene pairs, *HvCEL1*, *HvCEL6* and *HvCEL3*. "Normalisation" of the data was performed using maize as the outgroup and an estimated separation time of 50 mya. *HvCEL10* and *HvCEL12* with their rice orthologues had very high synonymous substitutions rates. One problem associated with such high values is that a $d_S > 1$ becomes quite problematic because the model is attempting to assess multiple substitutions at one site so increasing the number of assumptions being made. These two gene pairs are included in all but Figure 2-15 A, however, these estimates should be considered as inaccurate, at best. Salse et al. (2008) estimates divergence times of the Pooideae and Oryzoideae at around 46 mya. Not including *HvCEL10* and *HvCEL12*, estimates here averaged 41 mya with a standard deviation of 10 mya.

2.3.5.3. Comparison of K_S of barley and rice with maize and sorghum orthologues

When combining the results of the maize:sorghum and barley:rice orthologues as presented in Figure 2-16, two different levels of K_S are seen, with a couple of outliers. Furthermore, where known, the K_S estimates are similar between the orthologues of all four species. This pattern was further explored by plotting K_S and the proportion of synonymous sites in the sequences (Figure 2-17). The proportion of synonymous sites in each sequence also shows two distinct levels in accordance with the variation in K_S estimates for the different genes, which indicates a link between K_S values and number of synonymous sites in the sequence.

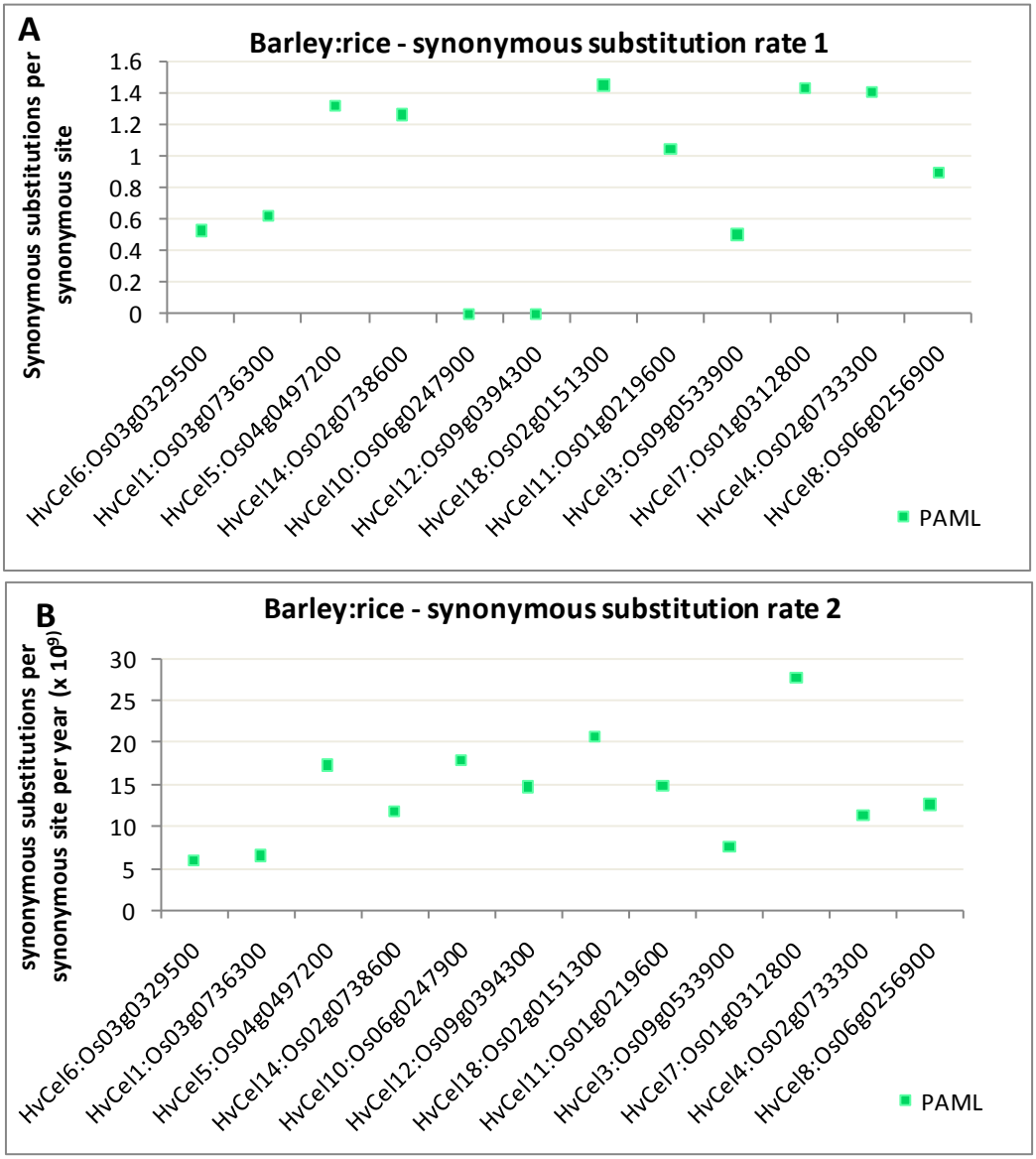
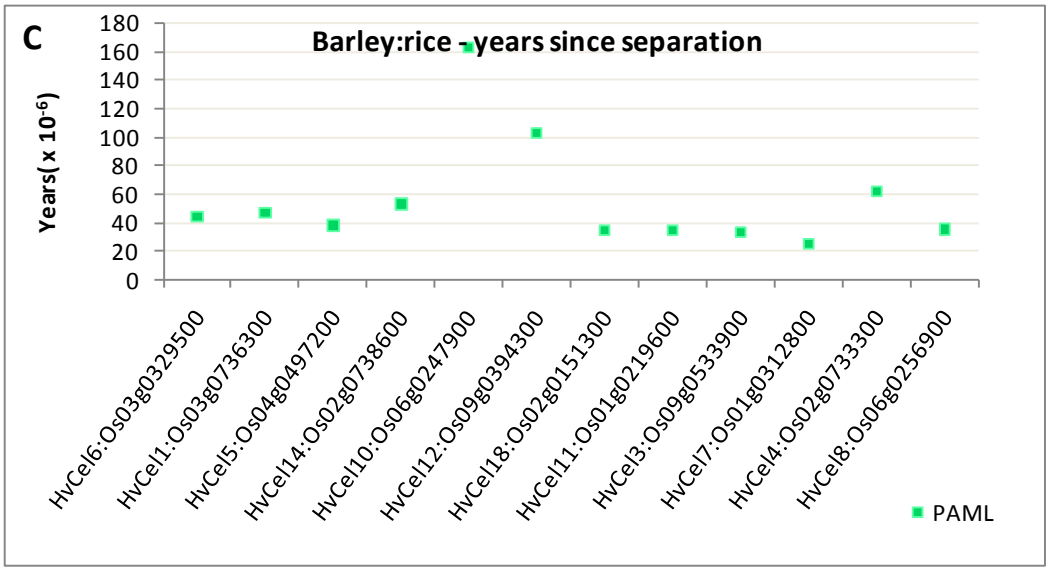
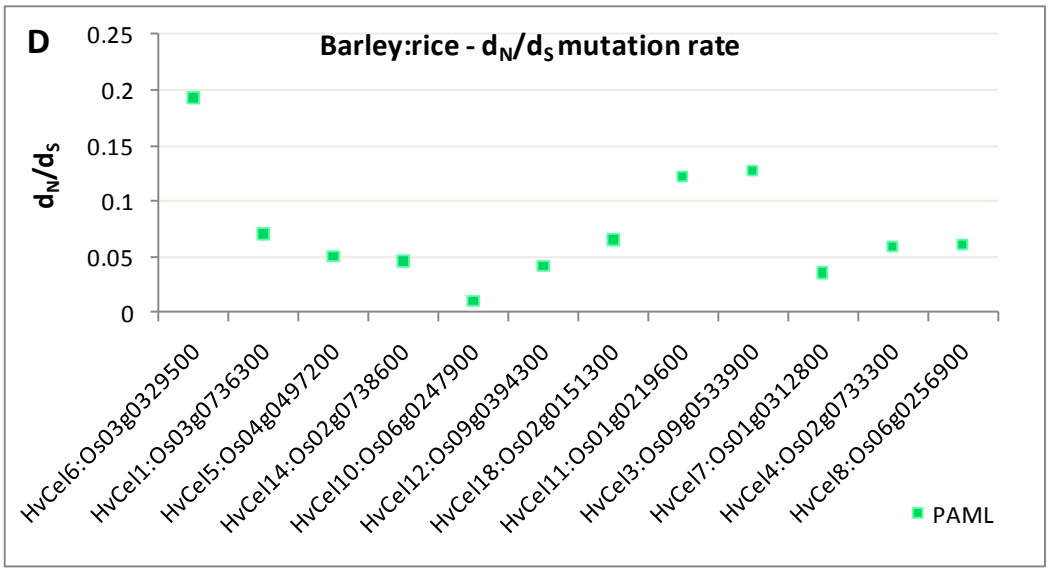


Figure 2-15: Codon based evolutionary distances for barley and rice genes.

Barley:rice distance analyses estimated using the PAML codeml model. The following graphs illustrate the differences for several parameters resulting from these estimations. A: synonymous substitutions per synonymous site. Estimates for *HvCEL10* and *HvCEL12* were not included as substitution rates were 5.83 and 3.02, which meant that the other gene pairs were not interpretable on the graph. B: "normalised" synonymous substitutions per site per year with *HvCEL10* and *HvCEL12* included. C: Years since separation are consistent with the synonymous substitution rates, with *HvCEL10* and *12* as outliers. D: Mutation rates for the barley:rice gene orthologues.


C
Barley:rice - years since separation
Years(x 10-6)
180
160
140
120
100
80
60
40
20
0
HvCel6:Os03g0329500
HvCel1:Os03g0736300
HvCel5:Os04g0497200
HvCel14:Os02g0738600
HvCel10:Os06g0247900
HvCel12:Os09g0394300
HvCel18:Os02g0151300
HvCel11:Os01g0219600
HvCel3:Os09g0533900
HvCel7:Os01g0312800
HvCel4:Os02g0733300
HvCel8:Os06g0256900
PAML


D
Barley:rice - dN/dS mutation rate
dN/dS
0.25
0.2
0.15
0.1
0.05
0
HvCel6:Os03g0329500
HvCel1:Os03g0736300
HvCel5:Os04g0497200
HvCel14:Os02g0738600
HvCel10:Os06g0247900
HvCel12:Os09g0394300
HvCel18:Os02g0151300
HvCel11:Os01g0219600
HvCel3:Os09g0533900
HvCel7:Os01g0312800
HvCel4:Os02g0733300
HvCel8:Os06g0256900
PAML

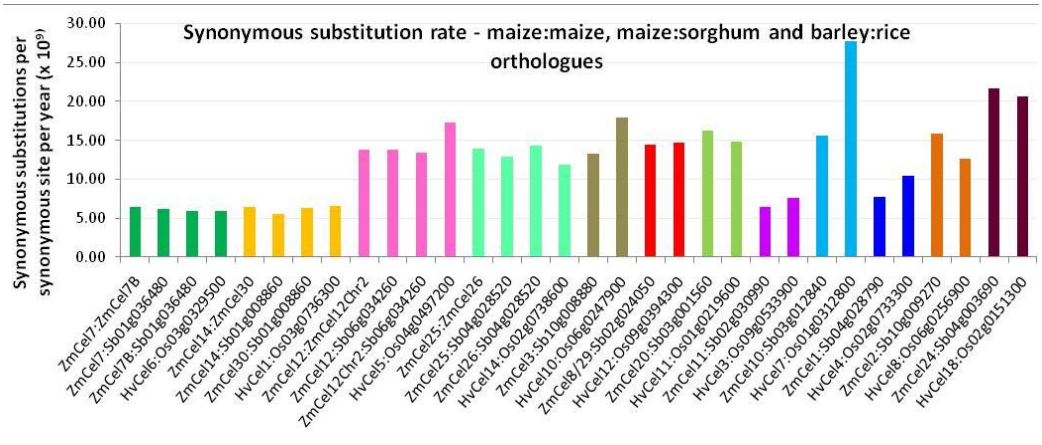


Figure 2-16: Orthologous gene pairs for maize:maize, maize:sorghum and barley:rice.

Comparison of orthologues for all species analysed shows two distinct levels of synonymous substitution. Orthologous gene sets are presented in the same colour and the graph shows that orthologues across all four species have very similar rates of synonymous substitutions per synonymous site per year.

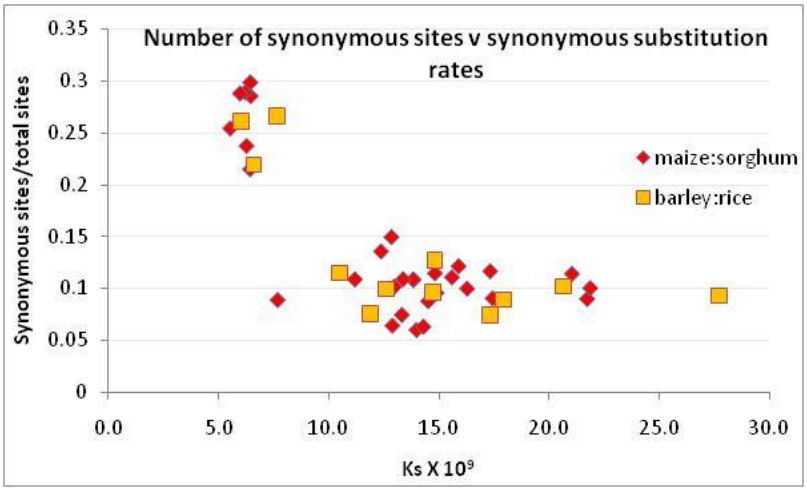


Figure 2-17: The ratio of number of synonymous sites to total sites plotted against K_S

This graph shows the separation of orthologous genes into two groups according to synonymous substitution rates and their respective relative level of synonymous sites.

2.3.5.4. *Synonymous substitution rates between some maize genes*

From the phylogenetic tree (Figure 2-7) it can be surmised that *ZmCEL12* and *ZmCEL6* are the result of an earlier duplication event. Again using rice as the outgroup, time since separation between *ZmCEL12* and *ZmCEL6* is estimated as 39.4 mya. Although somewhat speculative, considering the degree of error that may be present, this result suggests that *ZmCEL12* and *ZmCEL6* are the result of a duplication event and genome rearrangement in the maize/sorghum antecedent since separation from rice, rather than the loss of a gene in rice.

Clade B3 (Figure 2-7) shows a group of orthologues in the cereals that contain four or five genes from maize, sorghum and rice and only one *Arabidopsis* gene. Further distance analysis was done on these genes to try and identify their time of separation. For *ZmCEL3* and *25* a separation time of 56.8 mya was estimated and for *ZmCEL3* and *ZmCEL24* a separation time of 51 mya was estimated. These times coincide with the time of separation of rice and barley from maize and sorghum antecedents, and reinforces that duplication most likely occurred just prior to their separation.

2.3.5.5. *Mutation rates since separation*

Mutation rates are presented for maize and barley in Figure 2-14 D and Figure 2-15 D, respectively. As the PAML codeml estimates for d_S were, in most cases, higher than the Nei-Gojobori analysis, the mutation rates were lower. In fact, estimated mutation rates of < 0.2 for both the maize and barley data indicate a strong tendancy towards gene purification (Kimura and Ohta, 1974).

2.4. DISCUSSION

Phylogeny is defined as "....an historical pattern of relationships among organisms which has resulted from the actions of many different evolutionary processes...cladograms show relative affinities of groups of organisms or taxa..." (www.answers.com/topic/Phylogeny). Using physical features alone for classification of organisms has over the years provided some difficulties, and grasses, which includes cereal species, are no exception (Kellogg, 1998). A sequence based phylogenetic analysis, on the other hand, provides an evolutionary direction, thus allowing greater understanding of the relationships of species to each other (Kellogg, 1998). The study here relies on amino acid and codon based nucleotide ORF sequence, intron and exon structure including splice sites and chromosome map information as the basis for the phylogenetic analysis.

Combining these four analytical methods provides a chronological phylogenetic trace. The intron/exon structure is the most ancient level of gene information available (Irimia and Roy, 2008). More recently, genes with their introns and exons were assembled into plant genomes and retention of that gene order along a chromosome has, in general, been retained between species after separation from their common ancestor. Although local rearrangement of gene position occurs from time to time with chromosome breakage and recombination and gene duplication, in general, linkage analysis has shown that cereals (Poaceae) are derived from a common ancestor with a common gene order (Moore et al., 1995; Gale and Devos, 1998; Wei et al., 2007). If retention of gene order along a chromosome (synteny) exists, it provides evidence of orthology. Gene phylogeny is also determined by codon-based substitutions. Due to their stochastic nature, the synonymous codon-based substitutions that have occurred since divergence from a common ancestor are a measure of the relative distance separating two genes and therefore two species or duplicated sequences (Kimura and Ohta, 1974). In contrast, the non-synonymous codon-based changes alter protein sequence. It is somewhat simplistic to assume that all changes to the protein sequence will affect protein activity, folding, substrate specificity, etc.

Nonetheless, the number of alterations in amino acid sequence, or non-synonymous substitutions, between two orthologous genes is used as a measure of evolutionary pressure to diverge, or not.

Apart from illuminating the differences between the cereals and *Arabidopsis*, as the dicotyledon representative, the phylogenetic trees produced some unexpected findings. The issue of whether clade B1 is an antecedent of the GH9A sub-family and the genes of clade B3 are really related to the GH9C sub-family were best answered using intron/exon structure and evolutionary distances as a guide.

Evolutionary distance

The number of synonymous changes per synonymous site per year for the *Adh1* and *2* genes has been estimated at 6.5×10^{-9} (Gaut et al., 1996) and this estimate has been used in the calculation of divergence times for other gene pairs (Gaut and Doebley, 1997; Lynch and Conery, 2000; Paterson et al., 2004; Salse et al., 2008). As seen with the endo-(1,4)-β-glucanase genes this is not a problem for genes that have an estimated Ks of 6.5×10^{-9}. However, as determined by this analysis and others (Gaut et al., 1996; Swigoňová et al., 2004), the rate of nucleotide substitutions varies between genes and in the case of the endo-(1,4)-β-glucanases three different rates within the family were estimated. It is hypothesised that differences in generation time and hence rate of meiosis is the main cause of synonymous substitution variation (Gaut et al., 1992). However, within species, the endo-(1,4)-β-glucanase genes have all been subjected to similar generation times. Therefore generation time, in this case, is probably not the reason for the variations. As synonymous substitution rates are relatively stable within orthologous groups, but vary between orthologous groups across the four species examined, it is more likely that internal gene structure or sequence composition would account for such differences. Variations in synonymous substitutions per synonymous site per year, or $K_{S,}$ appeared to be related to the number of synonymous sites as a proportion of total number of nucleotides in the sequence. This seems logical. However, it was unexpected

that there were two distinct levels for the estimated proportion of synonymous sites within the gene family (Figure 2-17).

In this study, distance estimates of the time since separation of orthologues and homoeologues of maize and sorghum, falls between 10 and 20 mya. These divergence times span both those obtained by Gaut and Doebley (1997) of 10 mya for the tetraploid event in maize and 20 mya for the separation of maize and sorghum from their common ancestor, and those estimated by Swigoňová et al. (2004), who estimated divergence times of between 5.8 and 15.9, with an average of 11.4 mya for separation of sorghum and maize. Their estimation for the genomic duplication event in maize is 4.8 mya (Swigoňová et al., 2004). These contrasting results place uncertainty on the accuracy of distance analyses in trying to estimate times of separation. Even after normalisation, a large amount of discrepancy exists between calculated separation times. For example, the average time of separation of rice and barley after removal of the extreme outliers is 40 mya with a standard deviation of 10 million years! Therefore although it is assumed that the synonymous substitution rate is a stochastic process, this, in fact is not the case. For example, codon usage varies between species and availability of transfer RNAs (tRNAs) that have evolved with codon usage will affect substitution bias (Christianson, 2005). Perhaps it may be possible to regard relative rates of synonymous substitution between orthologous genes as reliable information.

Although this study showed that definite patterns exist, for example, rates of synonymous substitution per year are quite similar within orthologues but differ between sets of orthologues, some unexpected results were reported with regard to the total number of synonymous changes. The estimated divergence time for *ZmCEL21* and *Sb07g001480* is 33.4 mya and suggests that these genes are not orthologues. Synonymous substitution rates as measured against rice are within the expected range, mapping data indicates these genes are syntenous and splicing indicates orthology. One possible explanation is that a tandem

duplication event may have occurred, some 33 mya, and subsequent divergence of sorghum and maize involved loss of one of the duplicates in each species.

In the rice:barley estimates, *HvCEL10* and *12* and their rice orthologues have a high rate of synonymous substitutions per site, with resultant large separation times. Intron/exon data indicate orthology with their respective rice genes. No map data is available for barley, and if there is retention of map synteny between the two genes then orthology would not be in doubt. Mutation rates at 0.01 and 0.04, respectively, add some weight to the suggestion that these genes are orthologous, which leaves the possibility of a duplication event prior to separation, as deduced in the previous case of *ZmCEL21*.

An important purpose for codon-based synonymous and nonsynonymous substitution rate analysis is in the estimation of evolutionary direction or mutation rates. Rates of mutation, which were all below 0.2, indicate that all of the endo-(1,4)-β-glucanases are resisting pressure to evolve including the duplicates and homoeologues. This trend towards gene purification would suggest that the role of this group of proteins has remained unchanged. It should however, be iterated that several genes showed amino acid variations at the catalytic active site. Though mutation rates are comparatively small, do these variations herald a progression away from purifying evolution for these genes with an alternative role, for example, inhibitory binding to substrates?

Gene intron and exon structure

A contentious issue within the literature is whether introns evolved during prokaryote development (introns-early) and have been lost, or were inserted late in evolutionary history (introns-late). Recent evidence is suggesting that both theories have credence (Hartung Frank et al., 2002; Krauss et al., 2005). However, in recent evolutionary times, intron loss has outnumbered intron gain (Irimia and Roy, 2008). Importantly, intron position may be a useful marker for identifying orthologous and paralogous genes and from an evolutionary perspective, provides a tool for defining early gene phylogeny (Irimia and Roy,

2008). Intron position, as a resource in phylogenetic analysis, should be restricted to use where there is strong sequence conservation (Krauss et al., 2005). Fortunately this is the case with the endo-(1,4)-β-glucanase genes. In the context of the endo-(1,4)-β-glucanase gene families of maize, barley, sorghum and rice, as reported in this study, and also *Arabidopsis* (Libertini et al., 2004), there is a clear picture of common intron positions across the family that are a useful measure in determination of phylogeny and orthology. Splice sites in the endo-(1,4)-β-glucanases are common across monocots and dicots, however, the genes in the dicot representative, *Arabidopsis*, contain two additional introns not found in the cereals studied (Libertini et al., 2004). On the issue of whether introns have been gained or lost in the endo-(1,4)-β-glucanase genes, an analysis that includes a wide range of species including fungi and insects would need to be performed to draw any conclusions.

Intron/exon structure provides good support for the notion that the endo-(1,4)-β-glucanase gene family was derived from a single progenitor gene with subsequent gene duplications. Insertion or deletion of the transmembrane domain may have resulted in loss and/or gain of intron 2 and/or intron 1. Since clade B1 and intron 1 are present in *Arabidopsis* and the Poaceae it is fair to assume that these two clades originated from the same ancestor prior to separation of the monocots and dicots (Libertini et al., 2004).

CBDs are a recognised part of the cellulase machinery in microbes (Gilkes et al., 1991; Juy et al., 1992; Levy et al., 2002; Gilad et al., 2003), and it would not be unreasonable to suggest that early in plant evolution, a CBD was lost during duplication or genomic rearrangement events. Separation of GH9C into two distinct clades, with an *Arabidopsis* orthologue in each clade, allows for the possibility that the sub-family separated into three genes before the division of monocots and dicots. Except for splice sites located in the CBD, splice sites are variable but common to all or some of the GH9B and C sub-family genes. Exon size however, indicates that all GH9C and clade B3 genes are, in evolutionary terms, recently separated. The presence of one *Arabidopsis* gene and at least four

genes in the cereals suggests that these genes have evolved since their separation from the dicots. This statement is reinforced by the evolutionary analysis that indicates that clade B3 divergence may have occurred around 51 to 57 mya. Intron/exon structure, together with the distance analyses indicate that separation of clade B3 genes occurred prior to separation of maize/sorghum and rice/barley from their common ancestor and suggests a cereal specific role for these genes.

Phylogenetic analysis of the maize duplicates

Gene duplication has the benefit of protecting organisms from the deleterious effects of mutations. Redundancy, whereby both genes are transcribed equally but both at lower levels, is also an important fate for duplicated genes, as this option also preserves stoichiometry (Force et al., 1999; Lynch and Conery, 2000; Lynch and Force, 2000; Lawton-Rauh, 2003). Subfunctionalisation occurs when one of the gene duplicates is transcribed at a much lower rate. This may be a first step to either silencing or neofunctionalisation. The least likely outcome is that of neofunctionalisation, or evolution of a new function for a gene, as seen with the Darwinian model (Kimura and Ohta, 1974). Kimura et al. (1974) makes the point that for new gene functions to evolve, genes must first duplicate. This then allows one gene to maintain the function of its encoded protein, while the second gene is free to retain accumulated mutations, under the umbrella of Darwinian positive selection, to produce a new gene, or, hence to oblivion, without disadvantage to the plant. The rate of gene loss after duplication can be quite variable and in *Arabidopsis*, there is a relationship between gene function and loss (Blanc and Wolfe, 2004). Genes involved in regulation, such as transcription factors or genes in metabolic pathways are more likely to retain duplicates, allowing for increased flexibility in development. Following the allotetraploidisation of maize, the genome has experienced gene losses with a homoeologue retention rate of approximately 25%, of which 30% have orthologues with rice and/or sorghum. The majority of the retained genes are transcription factors (Schnable et al., 2009). This is consistent with earlier studies that found the genome has experienced gene losses of greater than 50% of

one or other of the homoeologous genes (Salse et al., 2008). Duplication retention is certainly not evenly spread across the genome; the *CESA* gene family, for example has a 100% retention rate (Penning et al., 2009).

Duplicate retention for the 24 maize endo-(1,4)-β-glucanase gene homoeologues studied here is 22%. This rate of loss is not unexpected as the endo-(1,4)-β-glucanase gene family is thought to be neither regulatory nor involved in metabolism. The main retention rate is seen in the GH9A sub-family, where three of the four genes have a homoeologue, as well as a duplication that occurred in the maize/sorghum ancestor. Combined with the 100% *CESA* gene retention rate, this adds fuel to the speculation that the KORRIGAN GH9A subfamily might be involved in cellulose production or at least, that these genes are working together. It may indicate that this group of genes has a role in regulation of cellulose production. Duplicated genes that are involved in metabolic pathways are often retained so that stoichiometry is preserved (Blanc and Wolfe, 2004) and this may imply that the GH9A genes are part of a cellulose production complex. Another interesting observation is that tandem repeats are not a feature of the endo-(1,4)-β-glucanase family in cereals. This is in contrast to a study in which 35% of gene families contained at least one and up to 15 tandem repeats (Messing et al., 2004).

Maize duplications and sorghum

Despite the near complete sequencing of the maize and sorghum genomes, little has been recently reported on comparison of the two genomes. Here, integration of the various analytical methods was used in an attempt to draw some conclusions regarding the homoeologous endo-(1,4)-β-glucanase gene pairs in maize and their relationship with sorghum. If maize and sorghum separated at time X and were reunited in maize at time Y, then maize homoeologues should have one duplicate diverging from sorghum at time X and the other at time Y. The divergence time of the two homoeologues should the similar to the initial separation time of the maize and sorghum antecedents. As mentioned earlier, it may be that the time since divergence and the putative re-convergence of the

maize and sorghum genomes is too short to be accurately measured using synonymous substitution rates, but it can still be attempted.

The homoeologous pair, *ZmCEL25* and *ZmCEL26*, provided convincing data in favour of the hypothesis that sorghum provided the genome to produce the allotetraploid maize. Divergence times for *ZmCEL25* and the sorghum orthologue, *Sb04g028520*, are around 11 mya with a mutation rate of 0.023 whereas *ZmCEL25* and *ZmCEL26* diverged around 17 mya as did *ZmCEL26* and *Sb04g028520* with a mutation rate of 0.038 and 0.046, respectively.

Distance analysis of the homoeologous pair of *ZmCEL7* and *ZmCEL7B* against their sorghum orthologue, *Sb1g036480*, indicates that although *ZmCEL7B* may be the closer relative to sorghum all three genes are remarkably similar when looking at synonymous substitutions per synonymous site and years since separation. Estimations of separation time are 13.4 mya for sorghum/*ZmCEL7B*, 14.5 mya for sorghum/*ZmCEL7* and 15.5 mya for *ZmCEL7*/*ZmCEL7B*. On the other hand, the mutation rate between *ZmCEL7B* and *Sb1g036480* is approximately half that for *ZmCEL7* and *Sb1g036480* and this is reflected in the phylogenetic tree of the amino acid alignment (Figure 2-8).

In the case of *ZmCEL14* and *30*, the parsimonious phylogenetic tree (Figure 2-8) indicates that *ZmCEL14* and *30* are more closely related to each other than to their sorghum orthologue. Synonymous substitutions per synonymous site favour *ZmCEL14* as being more closely related to *Sb01g008860* than *ZmCEL30*. Both synonymous and non-synonymous substitutions for *ZmCEL30*, as measured against rice are greater than *ZmCEL14* and *Sb01g008860*, which suggests that there is less purifying evolutionary pressure on *ZmCEL30* compared to the other two genes. Separation time of these genes for maize and sorghum is much greater than expected, 19.7 mya for *ZmCEL14*/*Sb01g008860*, 25 mya for *ZmCEL30*/*ZmCEL14* and 28.3 mya for *ZmCEL30* and *Sb01g008860*. These estimates suggest duplication of the antecedent gene before divergence of the common ancestor with loss of *ZmCEL30* from sorghum at or after the maize/sorghum separation. In contrast to the synonymous estimates, non-

synonymous substitutions per non-synonymous site were the lowest recorded for the endo-(1,4)-β-glucanase gene family, which means that mutation rates are significantly lower than most of the other gene pairs. Mutation rates for *ZmCEL14*/*Sb01g008860* are about half those for *ZmCEL30* and *Sb01g008860* suggesting that *ZmCEL14* and *Sb01g008860* are more likely to be derived from the same source gene.

The *ZmCEL12* group of homoeologues/duplicates is orthologous with *Sb06g034260*. There are three *ZmCEL12* genes located on the maize genome, on Chromosomes 1, 2 and 10. *ZmCEL12Chr1* is positioned on Chromosome 1 and is not in a homoeologous position with either of the other two *ZmCEL12* genes or syntenic with sorghum. The ORF also retains 100% identity with *ZmCEL12Chr10*. These results suggest *ZmCEL12Chr1* represents a very recent segmental duplication event. In this case, synonymous substitutions per synonymous site and mutation rates are similar for the orthologues, however, *ZmCEL12Chr2* and *ZmCEL12Chr10* are more closely related to each other than they are to sorghum. Except for *ZmCEL12Chr1*, map position in maize is homoeologous, as well as being syntenic with the sorghum orthologue. The *ZmCEL12* group of genes are also interesting as separation time for *ZmCEL12Chr2* and *Chr10* is only 6.3 mya which coincides with the estimated maize allotetraploidy as estimated by Swigoňová et al. (2004). Separation time for both *ZmCEL12* homoeologues with sorghum is approximately 12 mya. These separation times suggest a duplication event.

ZmCEL12 also features in a possible duplication event that involves *ZmCEL6* that occurred after divergence of the common ancestor of rice/barley and maize/sorghum. There is no rice orthologue for *ZmCEL6*. Distance analysis places the occurrence of this duplication at around 40 mya. Map synteny evidence suggests either, a breakage and rearrangement of the chromosome, producing the duplication, or a segmental duplication event that inserted the duplicate further along the chromosome from the original gene. Further comparative fine mapping of the region may solve this dilemma.

The maize allotetraploid event has been estimated to have occurred between 4.8 and 10 mya (Gaut and Doebley, 1997; Paterson et al., 2004; Swigonova et al., 2004; Salse et al., 2008). Wei et al. (2007) have not indicated from their linkage analysis that sorghum is involved, but instead hypothesise that an unnamed maize divergent is responsible. Some of the data presented here is indicative that a sorghum antecedent was involved in the allotetraploid event in maize as mooted by Gaut and Doebley (1997) and Patterson et al. (2004). It is not possible to determine if this is a direct antecedent of sorghum, or one that has branched from the sorghum line prior to maize allotetraploidisation. Because the time since maize genome duplication is, in an evolutionary sense quite short, these analysis estimates are not accurate enough to be more certain.

2.5. SUMMARY

Similar numbers of endo-(1,4)-β-glucanase genes are found in maize (29) rice (24), sorghum (23) and *Arabidopsis* (24). The phylogenetic tree shows that evolutionary distance has resulted in differential gene loss and gain between the cereals and *Arabidopsis*. A group of genes without a CBD, but related to the GH9C sub-family as deduced from the phylogenetic tree and intron/exon structure, has expanded in all of the cereals analysed since their separation from the dicots, producing a group of genes that are cereal specific. *Arabidopsis*, on the other hand, appears to have retained or gained duplicates in the GH9B sub-family that are not present in the cereals. The 29 genes in the maize genome include five homoeologues/duplicates, and codon-based analysis indicates that one of *ZmCEL7/7B* and *ZmCEL25/26* may have been gained from the recent allotetraploid event, *ZmCEL14/30* and *ZmCEL6/12* may be the result of an earlier tandem duplication event prior to separation from the sorghum ancestor, while *ZmCEL12Chr1* is a segmental duplication that has occurred since allotetraploidy. The number of synonymous substitutions per year vary between the different genes of maize, barley, rice and sorghum in the endo-(1,4)-β-glucanase gene family, but are remarkably similar within orthologous groups. Times of separation as calculated using synonymous substitutions per synonymous site per

year should be regarded with caution. Using the estimate of a general divergence of cereals of around 50 mya, this analysis estimates that maize and sorghum separated at between 10 and 20 mya, and that rice and barley separated at around 41 mya. Evolutionary analysis of the maize duplicates/homoeologues is suggestive that a sorghum antecedent provided the DNA for the allotetraploid event. As found in other studies using linkage analyses, there has been retention of gene synteny of the orthologues. It can also be said that evolutionary pressure on all genes studied is of a purifying trend.

The phylogenetic study described in this Chapter has provided some insights into the structure of the endo-(1,4)-β-glucanase gene family, with particular emphasis on the Poaceae. As a means of providing a clearer picture of the roles of some these genes, a transcriptional analysis was performed and is presented in the next Chapter.

Chapter 3

A Transcriptional Analysis of Endo-(1,4)-β-glucanase Genes in Barley and Maize

3.1. INTRODUCTION

Following the description of the maize and barley endo-(1,4)-β-glucanase gene families presented in Chapter 2, the transcription patterns of many of these genes were investigated in a range of tissues and organs, and the results are presented in this Chapter.

As a means of determining gene activity in a tissue, transcriptional analysis can provide a quantitative comparative measure. Transcript levels of the endo-(1,4)-β-glucanase genes were examined using microarray, massively parallel signature sequencing (MPSS) and quantitative real time PCR (QPCR) methods. A general profile of gene transcripts across a number of tissues was performed, but with stem lodging as the main focus of the project, a more specific analysis of developing stem tissues was carried out. This required an understanding of how the stem develops during growth in both barley and maize.

3.1.1. Development of the cereal stem

Growth of the stem

A grass stem is divided into several nodes and internodes (Figure 3-1). A node is defined as the point at which the leaf emerges from the stem and is apparent as a localised thickening, whilst the internode is the length of tissue between each node. The stem is derived from meristematic tissue situated just below the shoot apex and growth of the stem commences after germination (Esau, 1977; Dickison, 2000).

Monocotyledonous plants retain meristematic tissue immediately above the node and growth of the internode commences by division and differentiation of these meristematic cells. The cells grow upwards and away from this "basal meristem". The internode length, and hence the stem length, is determined by the number of cells produced in the meristem as well as the degree of elongation of those cells that have been pushed upwards from the meristem (Morrison et al., 1994). The pattern of internode elongation in maize as an illustration of monocotyledons generally, is well illustrated by Figure 3-2. Initially, the

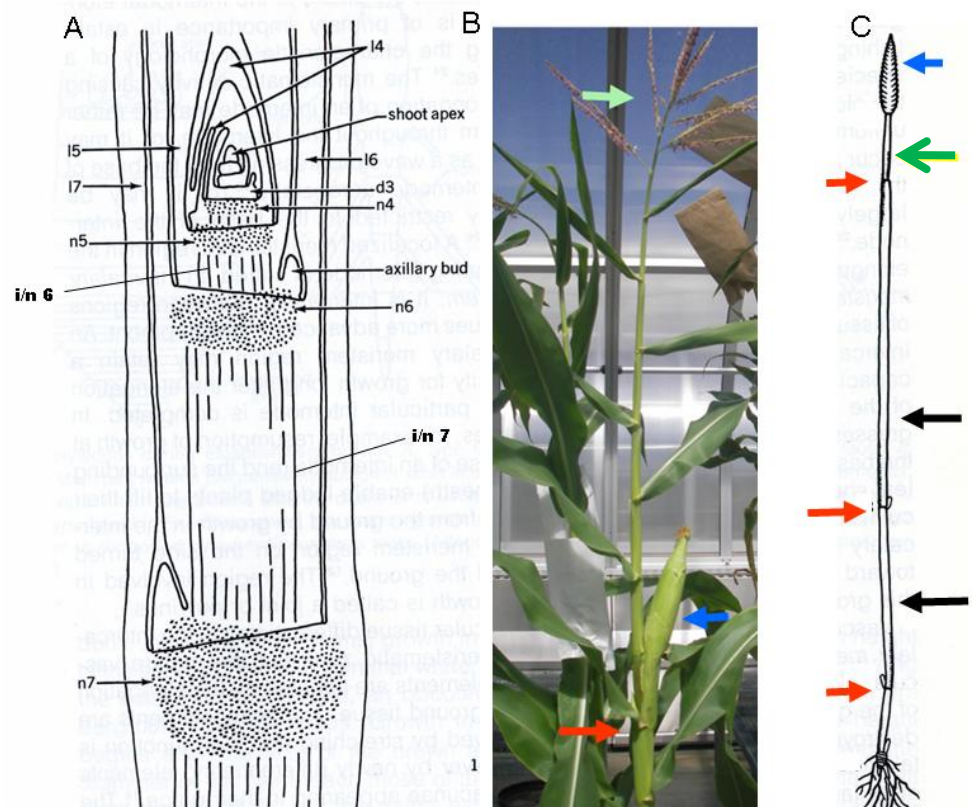


Figure 3-1: Anatomy and development of a monocotyledon plant stem.

A. Nodes (n) of an *Agropyron repens* grass stem are numbered from youngest to oldest, ie. n7 is more mature than n6. i/n refers to internode associated with the node. Leaves (l) ensheath the internode and growing stem superior to the node from which it is derived. Discs of insertion of the leaf primordia (d) are indicated by d3. Cells from the meristem of n7 and n6 nodes are dividing and pushing cells above them into the internode where they elongate to produce i/n 7 and 6 and push the meristematic region upwards as they do so. The shoot apex contains the head, or ear with all reproductive structures of the plant. (Taken from Dickison, 2000). Compare with maize, **B**. in which the male reproductive structures, or tassels (green arrow) are situated at the apex of the plant and female structures, blue arrow, (ears and silks) are located at the node. Red arrow shows position of the node from which the ear develops. **C**. a cereal stalk, such as wheat or barley, with leaves removed showing internodes (black arrows), nodes (red arrows), peduncle supporting the head (green arrow) and head with developing grain (blue arrow) (Taken from Esau, 1977).

NOTE:
This figure/table/image has been removed to comply with copyright regulations.
It is included in the print copy of the thesis held by the University of Adelaide Library.

Figure 3-2: Maize internode elongation over a 10 day period.

This photo is adapted from a Figure in Morrison et al. (1994) and shows the pattern associated with maize stem internode elongation. Paint dots that were spaced at 1mm intervals on a maize stem internode at day 1 have, for clarity, been overlayed with lines by the author. As the internode elongates, from Day 1 through until Day 10, an increase in the spacing between paint dots at the point of elongation is seen. Initially, the whole internode elongates at a similar rate but by day 4 elongation has perceptibly reduced in the top three dots and as elongation further progresses, the area of maximal elongation progresses towards the base of the internode. After day 8 elongation only occurs in the lower half of the internode, between the three lowest dots (Adapted from Morrison et al., 1994).

internode consists only of elongating cells but as these cells complete their elongation cycle, cells at the distal end of the internode begin to produce secondary cell walls and elongation for those cells ceases. Over time the elongating internode comprises a lower section of elongating cells, an upper one third of mature cells and a middle region where both elongating and maturing cells are found. These are called the elongation zone, the maturation zone and the transition zone, respectively. Eventually, all cells of the internode stop elongating and only secondary cell walls are deposited.

The cell wall during stem growth

Meristematic cells undergo a series of divisions at the nodal meristem. The new cells differentiate and begin to elongate. It is this cell division and growth extension that is responsible for the elongation of the internode. During internode elongation, the previously deposited primary wall must allow for an increase in length, while still maintaining structural support and resistance to turgor pressure. Cell elongation is thought to be initiated by expansin proteins that are activated by an auxin induced lowering of the pH within the cell wall (Cosgrove, 2005). Disruption of H-bonds between the glucuronoarabinoxylans (GAX) and cellulose that normally stabilise the cellulose microfibrils occurs. This is part of a mechanism that is believed to allow for cell wall "creep", which allows cellulose microfibrils to slide across one another so the wall is able to elongate without breaking. Cell wall polysaccharides continue to be synthesied and deposited in the wall during this phase of primary cell wall synthesis.

Another, yet to be fully defined, component of cell elongation is the removal or rearrangement of the GAX and XG within the cell wall matrix. This is thought to be achieved in part by the xyloglucan endotransglycosylase/hydrolase enzymes (XETs). *In vitro* activity of purified XETs has indicated that these enzymes might catalyse the covalent linking of XGs to (1,3;1,4)-β-glucans and XGs to cellulose (Hrmova et al., 2007). The evidence strongly suggests that they might also restructure the walls or may have a role in loosening walls during elongation (Nishitani and Tominaga, 1992; Thompson and Fry, 2001; Hrmova et al., 2009). Although the endo-(1,4)-β-glucanases are generally assumed to hydrolyse

cellulose, it is further speculated that the endo-(1,4)-β-glucanases are part of the elongation process because they are known to hydrolyse (1,3;1,4)-β-glucans and XGs *in vitro* (Hayashi et al., 2005; Yoshida and Komae, 2006; Urbanowicz et al., 2007b). Once fully elongated, the cell wall undergoes maturation with the deposition of the secondary cell wall. Indeed, wall material deposited after elongation is defined as the secondary cell wall, although changes such as lignin deposition can occur simultaneously in the primary wall.

During the processes of elongation and maturation, there are changes in relative concentrations of cell wall constituents in the stem and other tissues. Such changes are well illustrated in elongating barley coleoptiles (Gibeaut et al., 2005). Over an eight day period, the cells of the coleoptile elongate and mature with minimal cell division occurring. During this time cellulose is the most abundant cell wall constituent at 35-40 mol% and it remains unchanged throughout that period. Levels of XG also remain constant throughout at between 8 and 12 mol%, however, the degree of substitution is reduced over the eight days. During elongation, (1,3;1,4)-β-glucan increases from 1 mol% to 10 mol% but drops back to 1-2 mol% when elongation ceases (Gibeaut, 2005).

Analysis of rice stem internodes shows that in the elongating section, cellulose and non-cellulosic polysaccharide levels are similar at approximately 50% (Azuma et al., 1996). However, the maturation zone of the same internode shows a shift with cellulose levels increasing to 60% and non-cellulosic polysaccharides decreasing to 40% of the cell wall. Maize internode cellulose levels show similar trends, depending on the stage of development of the internode (Appenzeller et al., 2004). In this study the maize elongation zone cell wall contained 25% cellulose, the transition zone contained 30% and mature vascular bundles comprised 50% cellulose.

After flowering and as the grain ripens changes in the cell wall are still occurring. In rice stems a decrease in the amount of (1,3;1,4)-β-glucans that coincides with increased lodging, reduced stem strength and an increase in (1,3;1,4)-β-glucanase transcripts has been observed (Baba et al., 2001). Another study investigating energy deficient maize during grain filling shows that a breakdown in stem

parenchyma occurs at this time and causes an increase in stalk lodging (Liebhardt et al., 1968). However, relative amounts of crystalline cellulose also increase, suggesting that while part of the cell wall is being removed, cellulose microfibrils are not degraded.

As outlined above and in Chapter 1, the endo-(1,4)-β-glucanase family is implicated in secondary cell wall synthesis and modification of wall structure during elongation. Gene transcription often provides an indication of downstream biochemistry and is more easily performed than protein analysis. Therefore, as a means of providing further evidence to support these speculations, a transcriptional analysis was performed.

3.1.2. Transcriptional analysis

Several methods for transcript analysis are available, including Northern hybridisation, microarray, MPSS and QPCR. The advantage of Northern hybrisation analysis is that it can be carried out with a minimum of costly equipment. However, disadvantages include the possibility of degradation of the RNA, the method is quite time consuming and it requires the use of toxic chemicals. Further, transcription of only one or a few genes can be analysed per experiment. Microarray techniques allow for screening of large numbers of genes at one time, but they are expensive, they may, in some cases, lack sensitivity (Taniguchi et al., 2001) and they are limited by their requirement for gene sequence data.

MPSS is also a very costly procedure, but, it has the advantage that it not only provides data for an entire mRNA population in a tissue but can also be performed without any prior sequence data knowledge. Thus, it provides a very useful tool for species that have not been fully sequenced (Brenner, 2000; Reinartz, 2002; Meyers et al., 2004; Stolovitzky et al., 2005; Nakano et al., 2006; Shah et al., 2006). A brief summary of the process involved in production of MPSS data is presented as a flow chart in Figure 3-3. This technique is also purported to detect transcript in tissues at levels as low as 5 transcripts per million (Brandenberger et al., 2004). The first four bases for all signature tags is GATC, which represents the binding site for the *Dpn*II restriction enzyme. Tag

NOTE:
This figure/table/image has been removed
to comply with copyright regulations.
It is included in the print copy of the thesis
held by the University of Adelaide Library.

Figure 3-3: Flow chart to describe MPSS library production.

A brief description of the steps involved in producing an MPSS library is presented in this flow chart (Taken from Meyers et al. 2004). The *Dpn*II restriction endonuclease cut in step 2 produces the GATC motif present at the start of all signature tags.

position, however, may affect results, as tags located at the 3' end of the sequence have notably higher levels of occurrence than those in the middle of the gene, and tags at the 5' end show lower abundance. (Chen and Rattray, 2006). Chen and Rattray (2006) suggest that genes whose MPSS tags are positioned at the 5' end of their mRNA transcript may not produce detectable transcript and may account for some of those genes not being represented in MPSS databases . Apparent levels of transcript can also be reduced by problems associated with sequencing of some specific four-base "words" (Meyers et al., 2004). A total of 20 four-base "words" that cause low measured transcript have been identified, and are represented in an estimated 7.7% of potential signatures. Of these 20 four-base "words", 16 are pallindomes and it is thought that perhaps some intermolecular interaction of these molecules on the surface of the microbead impairs the sequencing process of these "words". Improperly sequenced signatures are subsequently removed during the quality control process (Meyers et al., 2004).

MPSS databases are produced with sequenced tags of 17 or 20 nucleotide bases, also referred to as 17mer or 20mer signature tags, respectively. Due to the nature of the sequencing procedure a 17mer tag theoretically increases the chance of a gene being represented, as fewer cycles of sequencing are undertaken, which provides less opportunity for errors. However, as the sequence is shorter, there is a theoretical reduction in specificity compared to the 20mer sequence tag. In reality however, the 17mer and 20mer databases have been found to produce similar results (Meyers et al., 2004).

Quantitative real time PCR (QPCR) is a very sensitive and accurate procedure for transcript profiling, but does require known sequence for part of the gene in question. Its usefulness is in quantitatively comparing gene transcription across several tissues or at different developmental stages. For this it relies on the transcript levels of multiple control genes, which are used to calculate a normalisation factor for each tissue (Vandesompele et al., 2002; Burton et al., 2004). When performing QPCR, several assumptions are made. Within tissue samples it is assumed that the cDNA pool is an accurate reflection of the amount of starting mRNA transcript and that the primer pairs are all amplifying at similar

rates (Livak and Schmittgen, 2001). For comparisons between tissues, the assumption is that the rate of transcript production for each of the control genes is constant across the different tissues (Vandesompele et al., 2002).

Considering the breadth of *in vitro* endo-(1,4)-β-glucanase enzyme substrate activity, it is probable that this gene family is transcribed across a wide range of tissues both spatially and temporally, where cell walls are being produced, strengthened and/or degraded. Some general transcriptional activity data is available for cereals. These include a public MPSS database for rice (http://mpss.udel.edu/in9311/) and the PLEXdb microarray database (http://www.plexdb.org/) (PLEXdb, 2001-2006) for barley (Druka et al., 2006). During this project, the Dupont-Pioneer proprietary maize MPSS library was made available. A more focussed analysis using QPCR on a barley developmental tissue series and a barley and maize stem developmental series was carried out. Thus, the aims of experiments described in this Chapter were firstly, to gain an understanding of transcriptional activity of the endo-(1,4)-β-glucanase family in maize and barley across a broad range of tissues and, secondly, to ascertain which genes were being transcribed during elongation and maturation of the stem internode in maize and barley.

3.2. METHODS

3.2.1. Barley Base AffyChip data search

Barley transcript data was acquired from the PLEXdb database Affymetrix Chip experiment BB3 entitled "Transcription Patterns during Barley development". (http://www.plexdb.org/modules/tools/plexdb_blast.php) (Druka et al., 2006; Wise et al., 2007). The database was searched using available cDNAs from barley endo-(1,4)-β-glucanase ESTs and contigs. Tissues in the database were from Morex and Golden Promise barley varieties and included developing grain, floral tissues and vegetative tissues from germinating embryo and seedling. These include germinating seed (coleoptile, radicle and embryo), seedling (root, crown and leaf), immature inflorescence, floral bracts (before anthesis), pistil (before anthesis), anthers (before anthesis), caryopsis at 5 days after pollination (DAP), 10 DAP and 16 DAP, embryo at 22 DAP and endosperm at 22 DAP.

3.2.2. Maize MPSS database search

The Dupont-Pioneer proprietary MPSS database comprises 17mer signature tags and contains 327 tissue libraries. Maize endo-(1,4)-β-glucanase cDNA sequences were found by searching the DuPont-Pioneer contig database, which provides Pioneer contig names for each cDNA. These contig names were used to search the DuPont-Pioneer MPSS databases for unique tags.

3.2.3. Plant tissue preparation for QPCR

3.2.3.1. Barley plant tissue series

A full description of the barley (*Hordeum vulgare* var Sloop) tissue series can be found in Burton et al. (2004) and Burton et al. (2008). Briefly, plants were grown in a greenhouse under a day/night temperature regime of 23°C/15°C or germinated either in damp vermiculite or on damp paper towels in the dark for 3 to 6 d at 20°C. Seedling leaves of about 13 cm in length were used to isolate leaf tip (the top 7 mm of the leaf) and 3 mm at the leaf base. Root tissues included root tip (1 cm, containing root cap, meristem, and elongation zone) and mature root (1-cm section about 6 cm behind the root tip, containing the differentiation and maturation zone). Floral tissues, consisting of anthers and pistil, were

collected about 2 weeks before anthesis and at anthesis. Stem tissue was taken from the upper internode, below the pre-anthesis spike (i.e. below the peduncle); cell elongation would have ceased in this segment. Extracts from coleoptiles grown in the dark at room temperature were prepared 1 to 7 days after imbibition of the grain by dissecting away the seedling leaves contained within them. Developing grain was hand pollinated and bagged, and the caryopses collected at 3-5 days after pollination (dap) and 8-10 dap. Embryos were collected 22 dap.

3.2.3.2. Barley stem series

Barley (*Hordeum vulgare* var Sloop) stems were harvested at the 3rd internode above the crown at different developmental stages, as illustrated in Figure 3-4. Because barley plants have many tillers, it is multi-stemmed, with stems at various stages of maturity. At the commencement of the experiment, stems were selected for harvest when the third internode above the crown was 1-2 cm in length. Because internode length can vary according to growth conditions at the time, to ensure consistency across the growth stages, stems were selected simultaneously and more stems than would be required were selected. Figure 3-4 indicates developmental stages selected for the barley stem series. The first harvest (B1) was performed as the entire third internode began elongating and was approximately 2-3cm in length. The second harvest (B2, B3 and B4) was performed when the internode was rapidly elongating at its proximal end (B1) but also had transition (B2) and maturation (B3) zones. The elongation zone was approximately one third of the length of the internode and was defined subjectively as being soft and flexible (personal communication: Kanwarpal Dhugga, Pioneer Hi-Bred International Inc., Johnston, Iowa, USA). The third harvest (B5 and B6) was performed when the internode had almost completed elongation, as assessed by daily measurement of internode length. At this stage the internode consisted principally of transition (B5) and maturation (B6) zones. The elongation zone was estimated to be less than 1 cm in length. The final harvest (B7) was performed 12 days after anthesis, when the secondary cell wall was advanced in maturity and grain filling was underway. In all cases, triplicate samples of 1 cm lengths of stem were harvested and all tissues were immediately placed in liquid nitrogen and stored at -80°C until required for RNA extraction.

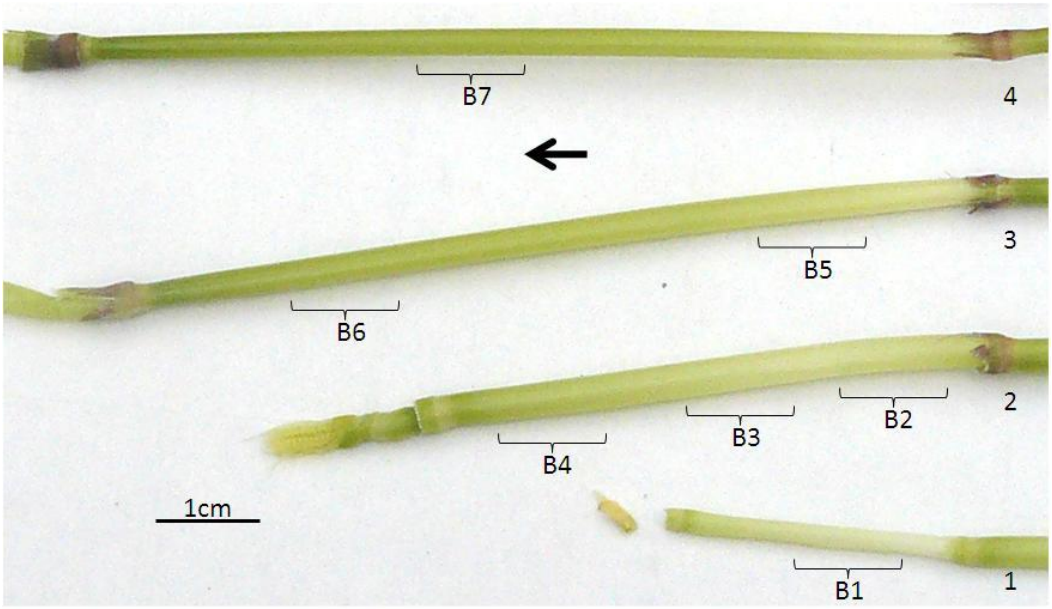


Figure 3-4: Sloop barley stem internodes with leaves removed to show the stages of development used for the barley stem series.

1. Early elongating internode, 2. Rapidly elongating internode, 3. Internode elongation almost complete, 4. Mature internode after flowering. The brackets labelled B1-B7 indicate the section of stem harvested for the particular stage of development. Sections of stem (approximately 1 cm in length) were removed for each harvest. The arrow indicates direction of growth and the top of the plant.

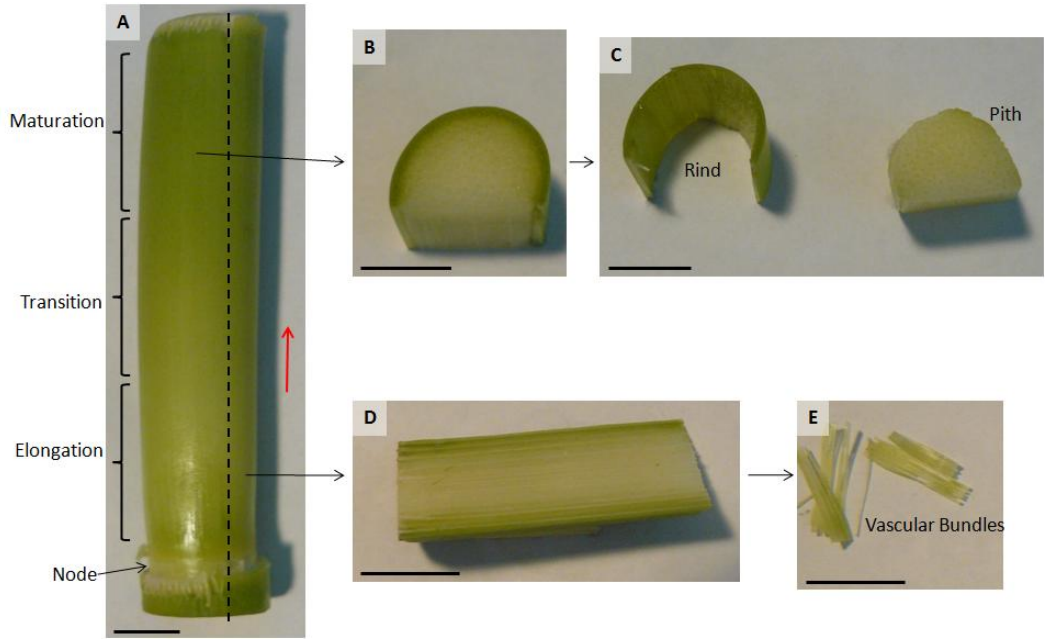


Figure 3-5: A B73 maize variety internode as prepared for the maize stem tissue series.

During preparation of the maize stem for the developmental series, the internode was dissected into the different stem tissues, including rind, pith and vascular bundles, prior to RNA extraction. Panel **A** shows an entire maize internode. The elongation, transition and maturation zones are labelled as is the node. The red arrow indicates the direction of growth and the top of the plant. The internode was split longitudinally into two sections as indicated by the dashed line. The larger section was divided into the zones according to degree of maturity and a 1 cm section cut from the centre of each zone as indicated in Panel **B**. The pith was subsequently dissected from the rind as shown in Panel **C**. Panel **D** shows the smaller section that was also divided into the zones according to the degree of maturity as discussed in section 3.2.3.3. The vascular bundles contained in the rind, and illustrated in Panel **E**, were manually stripped from the rind after breaking the section in half. The black line in the left hand corner of each panel indicates 1 cm.

3.2.3.3. Maize stem series

Maize tissues from the B73 cultivar were harvested for QPCR. As domesticated maize normally has only a single tiller, several plants were grown to enable harvesting of stems at different stages of development. Stages of maturity for the maize stem series were similar to that in barley. The following growth stages of maize were determined with reference to the Iowa State University extension website (http://www.extension.iastate.edu/hancock/info/corn.htm) (Hill, 2007). The first four internodes are contained below soil level with internode five, usually approximately one cm in length, the first visible internode above ground. Numbering the vegetative stages of development uses the fifth node as the reference point. Vegetative, or V stages of development, are reached when the collar of the leaf for the respective node appears above the sheath of the preceding leaf. R2 is reproductive stage two, or the blister stage of grain development. Internode 12, the internode found to be situated directly below the ear of B73 maize plants, was harvested at the following stages of development (Hill, 2007).

The following explanation should be read in conjunction with Figure 3-2. The first harvest was made of the early elongating internode at late V11 when the internode was at approximately day three on Figure 3-2 (Morrison et al., 1994; Hill, 2007). The second harvest occurred with the internode at approximately day seven of its development, while it was still rapidly elongating, but elongation had ceased at the distal end. This was at vegetative stage V14 and the internode contained elongation, transition and maturation zones. The next harvest was at the V16 stage of development at which point elongation had ceased and the internode was composed of transition and maturation zones only. The final harvest at R2 was of the internode when it was almost fully mature (Hill, 2007). As illustrated with barley (Figure 3-4), a one cm portion of the stem was removed from the centre of the relevant section of the internode. Internode tissues were dissected into rind, pith and vascular bundles as illustrated in Figure 3-5. For the first harvest, when the vascular bundles were too soft to remove effectively, the internode was dissected into rind and pith only. In all cases, triplicate samples of 1 cm lengths of stem were harvested and all tissues were

immediately placed in liquid nitrogen and stored at -80°C until required for RNA extraction.

3.2.4. Real-time Quantitative PCR (QPCR)

PCR reactions were performed as described in Chapter 2, section 2.2.3. Ribonucleic acid (RNA) and deoxyribonucleic acid (DNA) were size separated as described in Chapter 2, section 2.2.4. Total RNA was extracted from plant tissue using a phenol/chloroform extraction method as described in Chapter 2, section 2.2.2.1. Protein, salts, short DNA sequences, including primers and primer dimers, and other contaminants contained in the PCR reaction mixture were removed using the clean-up gel extraction column, NucleoSpin Extract II (Machery-Nagel) according to the manufacturer's instructions. This is described in Chapter 2, section 2.2.5. Sanger sequencing was performed at the Australian Genome Research Facility (AGRF) by capillary separation using ABI Prism BigDye Terminator Sequencing Reaction Kits (Applied Biosciences) on an ABI 3730*xl* sequencer as described in Chapter 2, section 2.2.6.

3.2.4.1. Production of cDNA

For barley and maize, cDNA was prepared by adding 1 to 3 µl RNA to 1 µl oligodT primer, a 15 base polyT oligonucleotide, and sterile filtered water to 12 ul, mixed and spun briefly and incubated at 70°C for two min. Tubes were immediately cooled on ice for at least two minutes and the contents briefly spun down before adding a mix of 4 µl 5X 1st strand buffer (Invitrogen), 1 µl DTT (0.1M), 1 µl 10mM dNTP mix, 0.5 µl RNAseOUT (Invitrogen), 0.25 µl Superscript III reverse transcriptase (Invitrogen) and sterile filtered water to a total volume of 8 µl. The total volume for each cDNA preparation was 20µl. The contents were mixed by gentle pipetting and incubated at 48°C for 90 min in the DNA Engine TETRAD2 Peltier Thermal Cycler before being heated to 70°C for 15 min to destroy enzyme activity. To ensure that intact cDNA had been produced it was tested by PCR using Platinum Taq polymerase (Invitrogen), amplifying product from the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) sequence in the case of barley, and for maize, amplification of elongation factor (ELF) was performed. The cDNA was stored at -20°C.

3.2.4.2. Primer design

Due to the high sequence identity of the endo-(1,4)-β-glucanases, primers for QPCR were designed from within the 3' untranslated regions of the genes. This maximised primer specificity for individual endo-(1,4)-β-glucanase mRNAs. Where possible primers were designed with a melting temperature of around 54°C, with a GC content greater than 50% and to amplify products of between 100 bp and 300 bp. Once the primers were designed, a search on the NCBI barley or maize databases was performed to ensure that there were no other genes to which these oligonucleotides may anneal.

To ensure that only one product of the correct sequence was obtained, product was amplified using the RG 6000 Rotor-Gene real-time thermal cycler (Corbett Research), and the melt curves examined for multiple products. Multiple products resulted in more than one melt curve, an uneven melt curve or a melt curve with a broad temperature range. If a single melt curve was produced, the product was purified on a high performance liquid chromatography (HPLC) column and checked for a single elution peak. The product of this peak was sequenced to ensure it was the correct product. Primers used for QPCR are given in Appendix A, Table 3, for barley and Table 4 for maize.

3.2.4.3. QPCR

All QPCR was kindly performed by Dr Neil Shirley at the Australian Centre for Plant Functional Genomics (ACPFG) according to the methods published in Burton et al. (2008). Reaction mixtures were prepared using a liquid-handling CAS-1200 robot (Corbett Robotics) and contained seven PCR standards and each of the prepared cDNAs. Each QPCR reaction mixture contained 2 µl of either the cDNA solution, the diluted standard or water with 5 µl of IQ SYBR Green PCR reagent (Bio-Rad Laboratories), 1.2 µl of each of 4 µM forward and reverse primers, 0.3µl of 103 SYBR Green in water and 0.3 µl of water for a total volume of 10 µl. For barley, normalization was carried out using primers for glyceraldehyde-3-phosphate dehydrogenase, heat shock protein 70, cyclophilin, and α-tubulin, using the geometric means of the three control genes that varied the least with respect to each other (Vandesompele et al., 2002; Burton et al.,

2004). As with barley, the maize glyceraldehyde-3-phosphate dehydrogenase and α-tubulin genes were used for normalisation, however, elongation factor and actin were used instead of cyclophilin and heat shock protein 70. The final concentrations of mRNA for the genes of interest were expressed as arbitrary units, representing the numbers of copies of mRNA per microliter of cDNA normalized against the best three of the four control genes (Burton et al., 2004).

3.2.4.4. Co-transcription analysis

A total of 122 genes known to play a role in cell wall synthesis were analysed by QPCR across the barley developmental series cDNAs. The results of relative levels of transcript were combined into a Microsoft Excel spreadsheet and, from these transcript data, a correlation coefficient matrix was produced. This enabled the determination of the co-transcriptional correlations for the endo-1,4-β-glucanase genes with each of the 122 cell wall synthesis genes.

3.3. RESULTS

3.3.1. Barley

3.3.1.1. PLEXdb Affymetrix microchip data

There are 11 endo-(1,4)-β-glucanases represented on the PLEXdb database experiment BB3: Transcription Patterns during Barley development (Druka et al., 2006). These data are presented graphically in Figure 3-6 and also in Appendix F, Table 1. There is a good deal of variation in transcript abundance between genes and tissues. For example, *HvCEL12* shows moderately high levels of transcript for all tissues, in contrast to *HvCEL11* that is highly transcribed in only pistil and caryopsis at five days after pollination. *HvCEL2* and *HvCEL7* show comparatively low levels of transcript in all tissues.

3.3.1.2. QPCR transcription across a barley developmental tissue series

A total of 12 genes were analysed for transcript levels using QPCR. The tissue series comprises a set of 16 Sloop barley tissue cDNAs that represent most parts of the plant and at various stages of development (Burton et al., 2004). The results are presented in Figure 3-7 but for a comprehensive table of results see Appendix F, Table 2. Comparative analysis of transcript patterns for the genes indicated that genes could be loosely divided into four groups. Presented in Figure 3-8A is Group 1 with *HvCEL1*, *HvCEL3* and *HvCEL14*. These genes showed the majority of transcripts in vegetative tissues, especially the leaf base and peduncle, but also, to a lesser extent, spike. Group 2 (Figure 3-8B) contained *HvCEL5* and *HvCEL10*. Here, the majority of transcripts were found in root tip, root base and leaf base but also significant activity in the anther at preanthesis. Figure 3-9A contained *HvCEL2*, *HvCEL4*, *HvCEL7* and *HvCEL8*. All of these genes had the majority of transcripts in floral tissues, but also root tip. The fourth group (Figure 3-9B), which contained *HvCEL6* and *HvCEL11* showed significant transcripts in floral tissues only. The barley tissues analysed for transcript in the Affymetrix Chip experiment BB3 entitled "Transcription Patterns during Barley Development" (Druka et al., 2006; Wise et al., 2007) are not directly comparable with the barley developmental tissue series used for QPCR. This is because the tissues used for the barley microchip data shown in

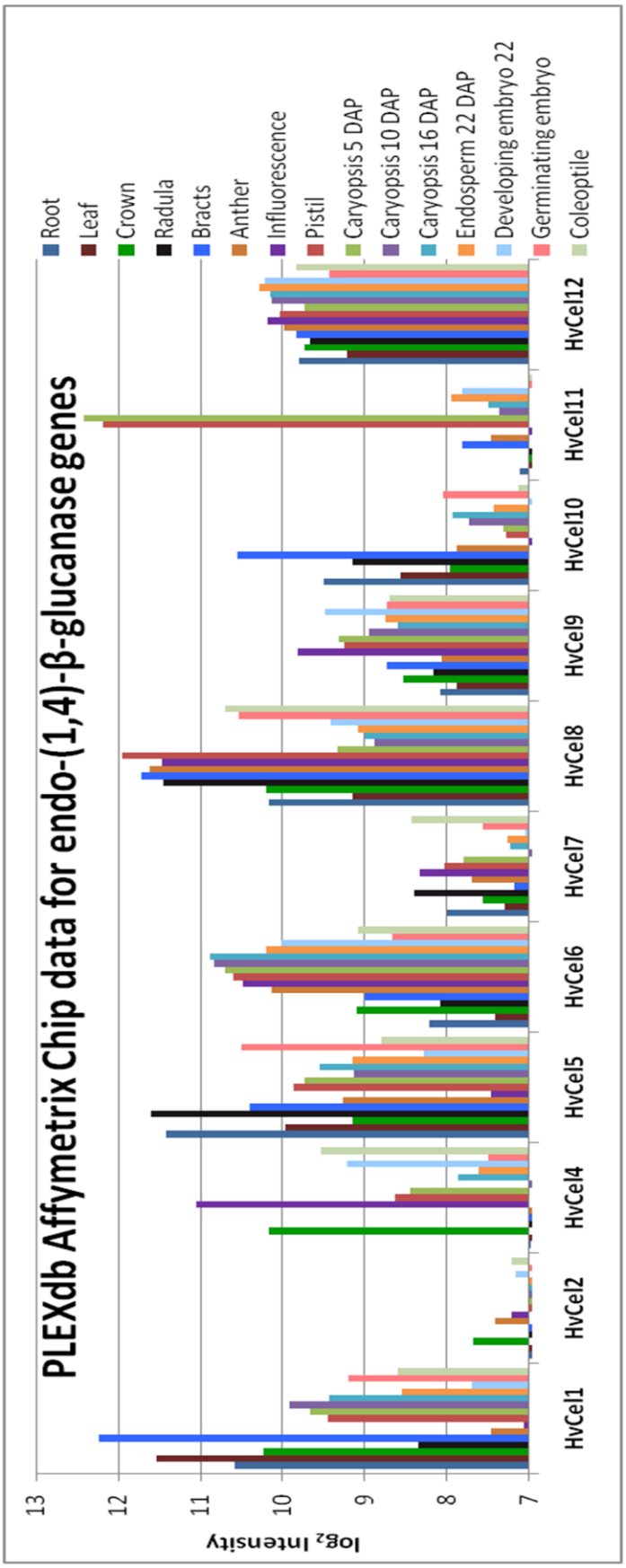

PLEXdb Affymetrix Chip data for endo-(1,4)-β-glucanase genes
Root
Leaf
Crown
Radula
Bracts
Anther
Influorescence
Pistil
Caryopsis 5 DAP
Caryopsis 10 DAP
Caryopsis 16 DAP
Endosperm 22 DAP
Developing embryo 22
Germinating embryo
Coleoptile
log2 Intensity
13
12
11
10
9
8
7
HvCel1
HvCel2
HvCel4
HvCel5
HvCel6
HvCel7
HvCel8
HvCel9
HvCel10
HvCel11
HvCel12

Figure 3-6: PLEXdb Affymetrix transcript data for endo-(1,4)-β-glucanase genes in barley variety Morex.

The data clearly shows the variety of tissues in which transcription of the endo-(1,4)-β-glucanases is occurring. It also illustrates that some genes have transcript in many tissues while others are restricted to a few tissues. *HvCEL9* relates to TC136010 a contig containing a premature stop codon. For barley, intensity levels below 7 are considered to be background fluorescence only (Druka et al., 2006). DAP = days after pollination

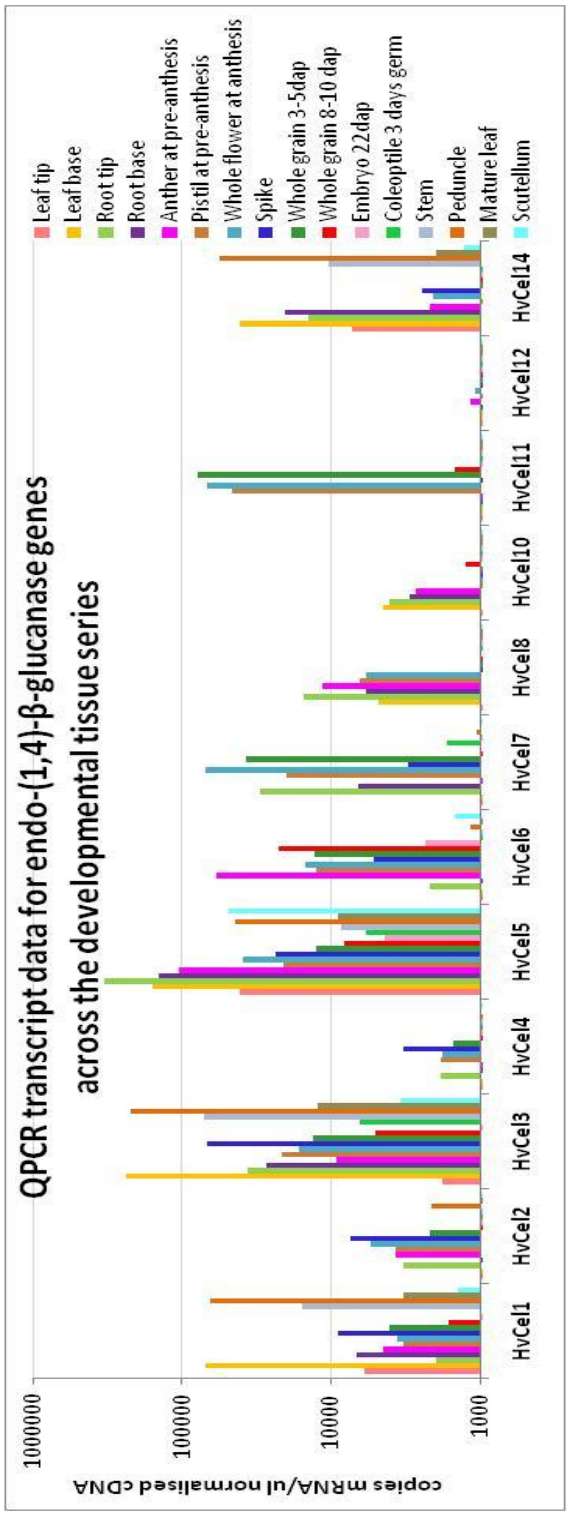

QPCR transcript data for endo-(1,4)-β-glucanase genes
across the developmental tissue series
copies mRNA/ul normalised cDNA
1000000
100000
10000
1000
HvCel1
HvCel2
HvCel3
HvCel4
HvCel5
HvCel6
HvCel7
HvCel8
HvCel10
HvCel11
HvCel12
HvCel14
Leaf tip
Leaf base
Root tip
Root base
Anther at pre-anthesis
Pistil at pre-anthesis
Whole flower at anthesis
Spike
Whole grain 3-5dap
Whole grain 8-10 dap
Embryo 22dap
Coleoptile 3 days germ
Stem
Peduncle
Mature leaf
Scutellum

Figure 3-7: QPCR levels of 12 endo-(1,4)-β-glucanase genes across the developmental tissue series.

The vertical axis is a $\log_{10}$ scale and shows relative transcript levels of the genes normalised against three control genes. The vertical axis crosses the horizontal axis at 1000 copies mRNA per µl normalised cDNA. This is an arbitrary level, below which, transcript levels are considered to be unreliable.

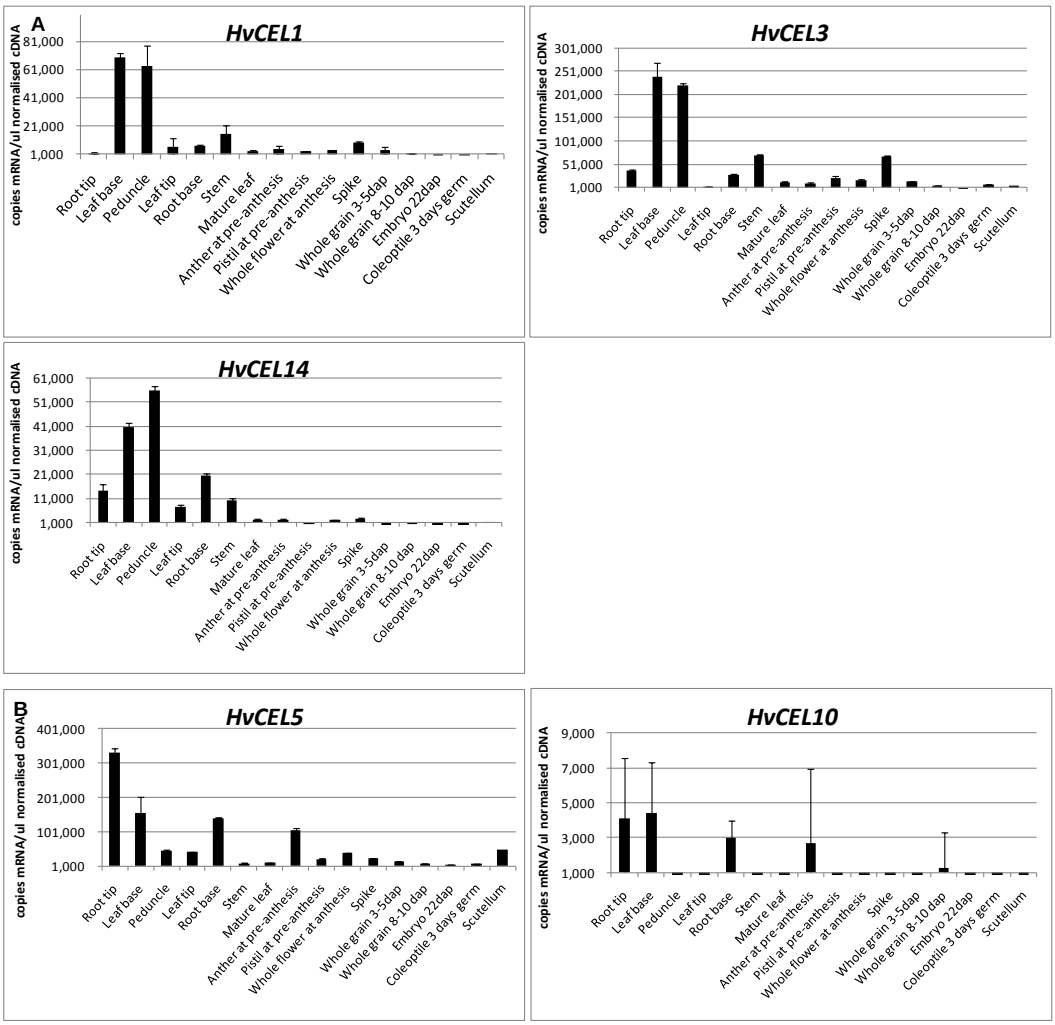


Figure 3-8: Transcript profiles for the endo-(1,4)-β-glucanase genes across the barley tissue developmental series.

Transcript below 1000 copies was arbitrarily considered to be background. **A:** QPCR data for *HvCEL1*, *3* and *14*. These genes had correlation coefficients of >0.91and showed highest levels of transcript in tissues with maturing secondary cell walls. **B:** QPCR data for *HvCEL5* and *10*. Statistical correlation of transcript between these two genes is >0.9, but standard deviations for HvCEL10 are significantly large.

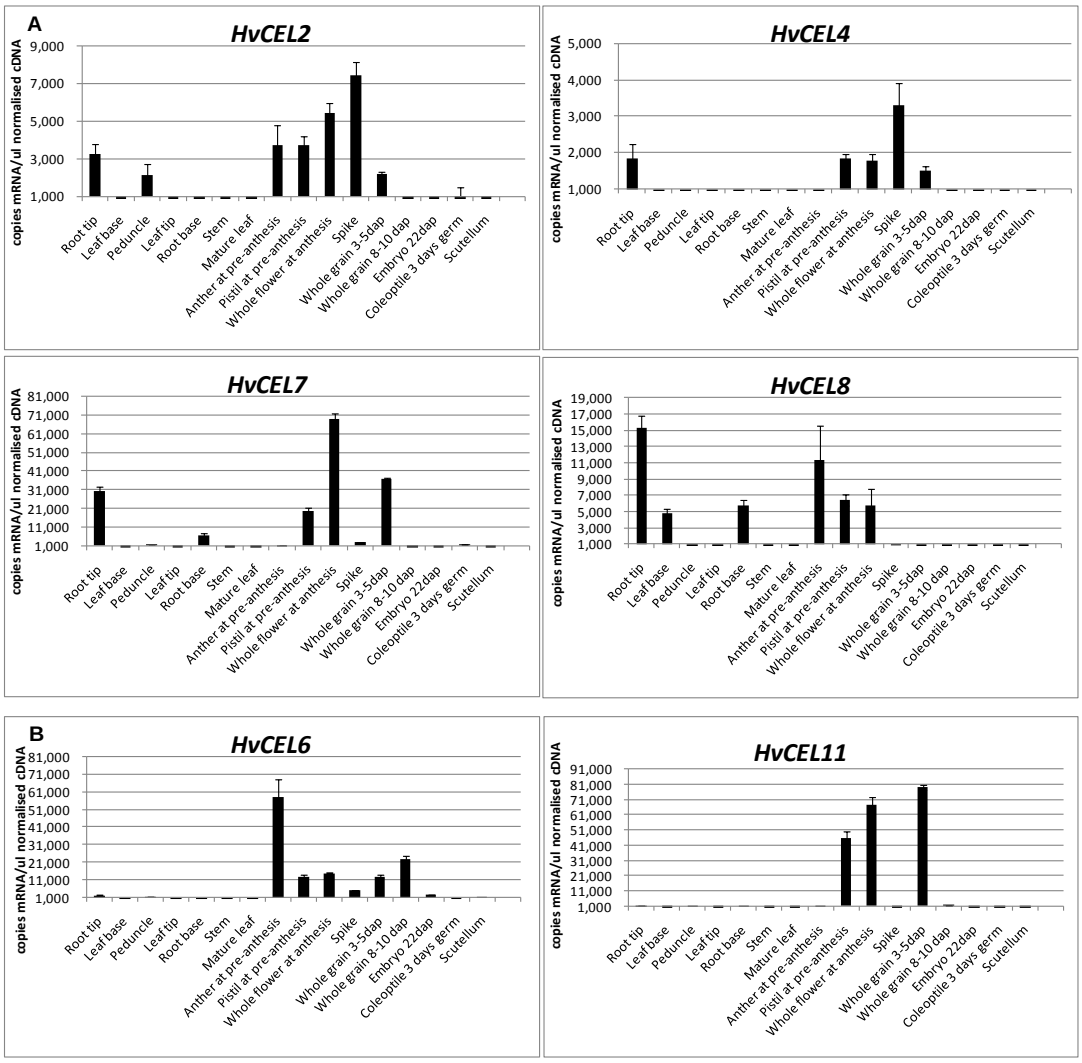


Figure 3-9: Transcript profiles for endo-(1,4)-β-glucanase genes across the barley tissue developmental series.

Transcript below 1000 copies was arbitrarily considered to be background. **A:** *HvCEL2*, *4*, *7* and *8* comprise the group of barley genes that are principally transcribed in floral tissues and early grain as well as root tip, and variously, peduncle, root base and leaf base. **B:** *HvCEL6* and *11* only showed transcript in floral tissues and early grain.

Figure 3-6 do not include any mature tissues. It is possible however, to make some comparisons, particularly with regard to floral tissues and developing grain. For example, *HvCEL11* has high levels of transcript in pistil, whole flowers and early developing endosperm only. Similarly, *HvCEL2*, *HvCEL4*, *HvCEL6*, and *HvCEL8* are transcribed in floral tissues, although at much lower levels, and are also actively transcribed in elongating tissues such as leaf base and root tip. In contrast, *HvCEL1* and *HvCEL14* show high levels of transcript in leaf base, stem and peduncle but much lower levels of transcript in floral tissues and root, and developing and germinating grain. Similarly, *HvCEL3* and *HvCEL5* show large amounts of transcript in vegetative tissues, but also have significant, although less transcript in floral tissues. Contrary to the PLEXdb data, *HvCEL10* and *HvCEL12* have very low levels of transcripts in all tissues in the barley developmental series. An explanation for this dichotomy is offered in the Discussion.

3.3.1.3. QPCR across a barley stem series

A stem-specific transcriptional analysis to investigate variation in gene activity of stem tissues as the internode elongates and matures was performed using a stem specific developmental series, as described in Figure 3-4. The barley stem series was designed to gain some insights into the endo-(1,4)-β-glucanase gene transcripts, both temporally and spatially, as the internode elongates and matures. The results are presented graphically in Figure 3-10 and Figure 3-11. A full results table is presented in Appendix F, Table 3.

The results indicate that the endo-(1,4)-β-glucanases can be loosely categorised into four groups. In group 1, *HvCEL1*, *HvCEL3* and *HvCEL14* have significant activity across all stages of development of the stem internode, but peak in the transition zone of the rapidly elongating internode (B3). Group 2, *HvCEL5* and *HvCEL10*, demonstrate a peak during the transitional phase of growth (B3), as with group 1 but with a secondary peak during early elongation (B1). Group 3, represented by *HvCEL2*, *HvCEL4*, *HvCEL6* and *HvCEL8*, have transcript levels peaking during early internode elongation (B1), and a minor peak as the grain fills (B7), although *HvCEL8* has a very large standard deviation at grain fill. The

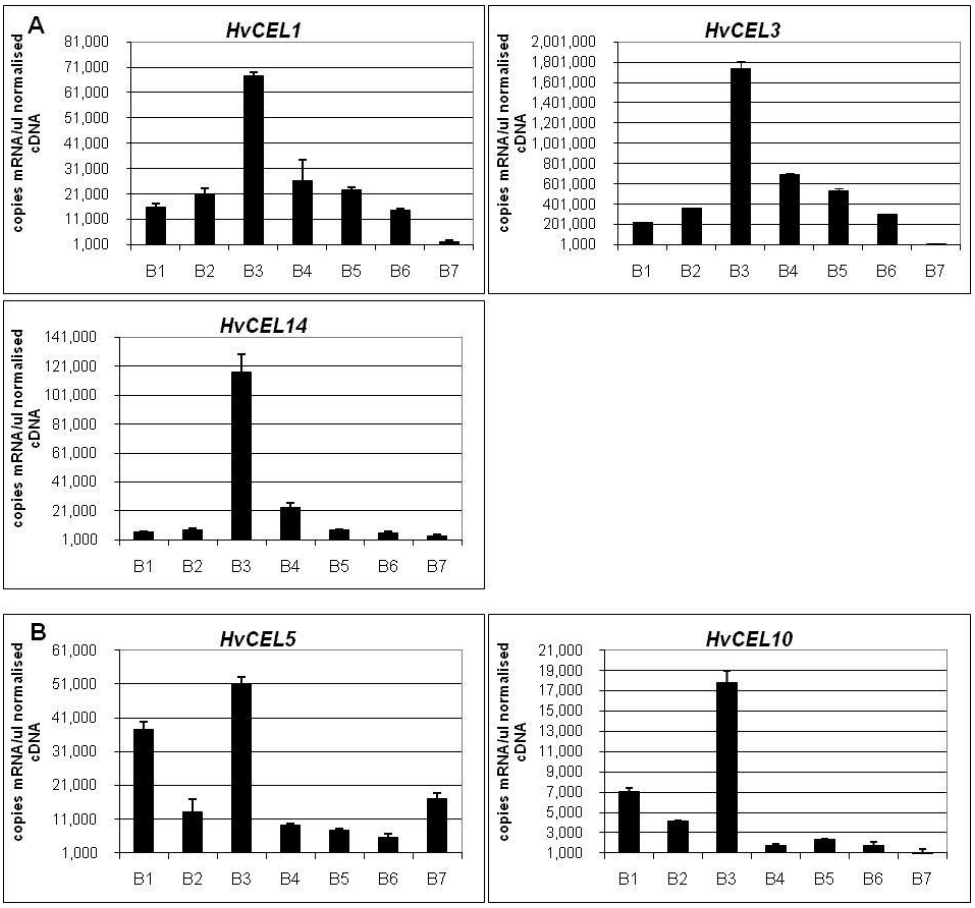


Figure 3-10: Transcript profiles of endo-(1,4)-β-glucanase genes across a barley stem tissue series**.**

Transcript below 1000 copies was arbitrarily considered to be background. **A:** all genes in this group show a marked increase in levels of transcript on stem internode tissues that are beginning to produce secondary cell walls. *HvCEL1* and *3* also show significant transcript in all other developmental stages except during grain fill. **B:** *HvCEL5* and *10* show transcript in the transitional phase of internode development as with the genes in A, but also show a secondary peak of transcript in young elongating internode. (B1-early elongation, B2-elongation zone of elongation internode, B3-transition zone of elongation internode, B4-maturation zone of elongation internode, B5-transition zone and B6-maturation zone of maturing internode, B7-internode 2 weeks post-anthesis.)

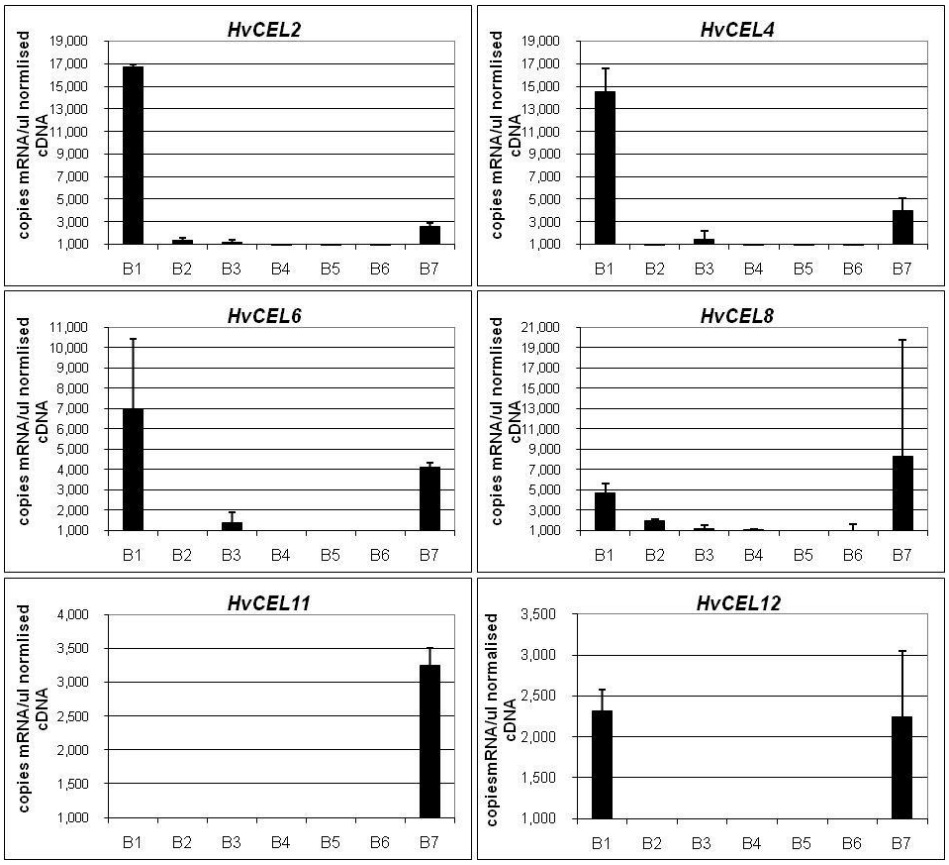


Figure 3-11: Transcript profiles of endo-(1,4)-β-glucanase genes across a barley stem tissue series.

HvCEL2 and *4* have transcript principally in young elongating internodes. Transcript below 1000 copies was arbitrarily considered to be background. *HvCEL6*, *8*, *11* and *12* have very low levels of transcript and may well not be actively involved in stem development. *HvCEL8* and *12* have high standard deviations and are therefore not statistically sound. (B1-early elongation, B2-elongation zone of elongation internode, B3-transition zone of elongation internode, B4-maturation zone of elongation internode, B5-transition zone and B6-maturation zone of maturing internode, B7-internode 2 weeks post-anthesis.)

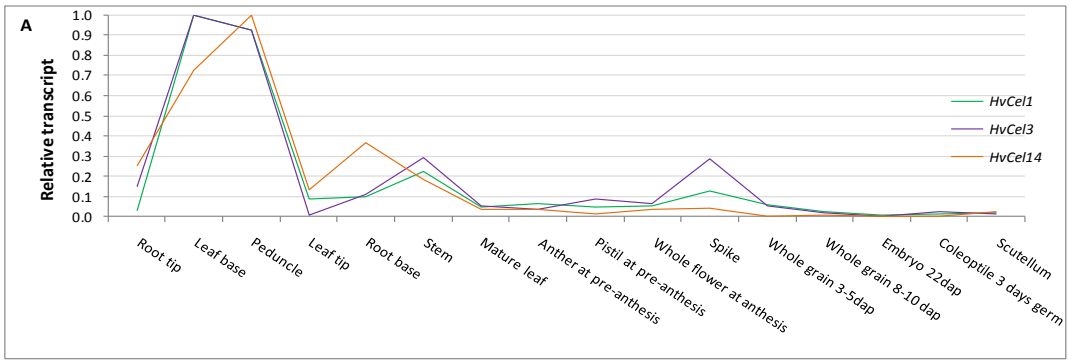
fourth group, containing *HvCEL7*, *HvCEL11* and *HvCEL12*, all have notably low levels of transcript of less than 3,500 copies per µl of normalised cDNA.

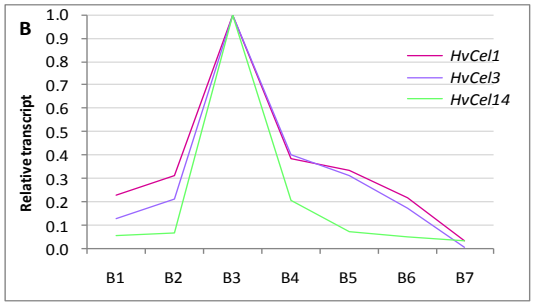
3.3.1.4. Correlation of endo-1,4-β-glucanase gene transcription in barley tissues

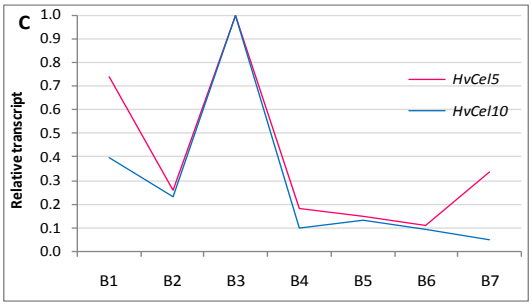
The general patterns associated with the QPCR transcript data as described in the last section were examined more closely. A correlation matrix of the levels of QPCR transcript of the different endo-1,4-β-glucanase genes in the various tissues was produced in Microsoft Excel to determine if these observed patterns were statistically significant. Graphs of relative transcript levels of genes that showed correlation coefficients (r^2) of >0.9 across the tissue and stem series are presented in Figure 3-12 and a summary of these graphs can be found in Table 3-1. Correlation coefficients for the barley tissue developmental series showed *HvCEL1*, *HvCEL3* and *HvCEL14* had an r^2 of between 0.91 and 0.98 (Figure 3-12A), and in the stem series, the r^2 value ranged between 0.96 and 0.99 (Figure 3-12B). *HvCEL5* and *HvCEL10* had an $r^2 = 0.86$ in the barley developmental tissue series, and for the stem series the r^2 value was 0.91 (Figure 3-12C). Due to the high standard deviation associated with *HvCEL10* in the barley developmental tissue series, the graph of correlations between *HvCEL5* and HvCEL10 was not included in Figure 3-12. Correlation coefficients of transcript showed that *HvCEL2*, *HvCEL4* and *HvCEL6* had an $r^2 > 0.91$ across the stem series but only *HvCEL2* and *HvCEL4* showed a significant correlation, $r^2 = 0.9$, in the tissue series. This is illustrated in Figure 3-12D and E.

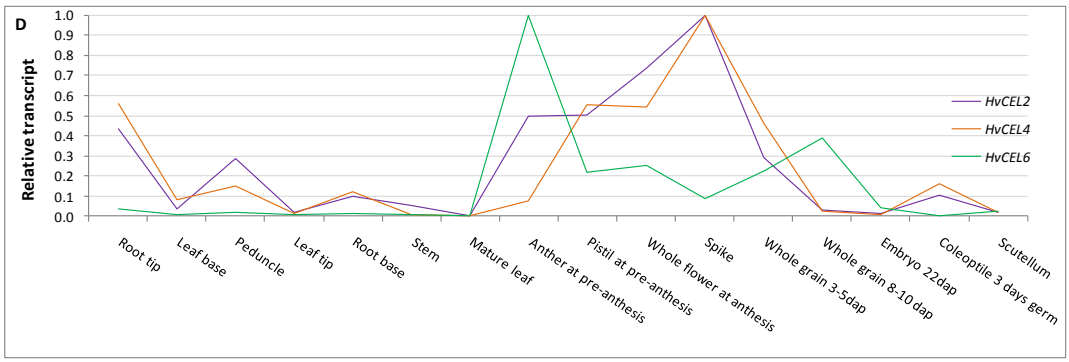
3.3.1.5. Correlation of transcript of the endo-1,4-β-glucanases with other cell wall genes

With knowledge that the endo-1,4-β-glucanase genes may play a role in cellulose production and cell elongation, transcript data from 122 genes across the barley developmental series were obtained. Correlation coefficients of co-transcription between these other genes and the endo-1,4-β-glucanases were determined. A


A
Relative transcript
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
HvCel1
HvCel3
HvCel14
Root tip
Leaf base
Peduncle
Leaf tip
Root base
Stem
Mature leaf
Anther at pre-anthesis
Pistil at pre-anthesis
Whole flower at anthesis
Spike
Whole grain 3-5dap
Whole grain 8-10 dap
Embryo 22dap
Coleoptile 3 days germ
Scutellum


B
Relative transcript
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
HvCel1
HvCel3
HvCel14
B1
B2
B3
B4
B5
B6
B7


C
Relative transcript
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
HvCel5
HvCel10
B1
B2
B3
B4
B5
B6
B7


D
Relative transcript
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
HvCEL2
HvCEL4
HvCEL6
Root tip
Leaf base
Peduncle
Leaf tip
Root base
Stem
Mature leaf
Anther at pre-anthesis
Pistil at pre-anthesis
Whole flower at anthesis
Spike
Whole grain 3-5dap
Whole grain 8-10 dap
Embryo 22dap
Coleoptile 3 days germ
Scutellum

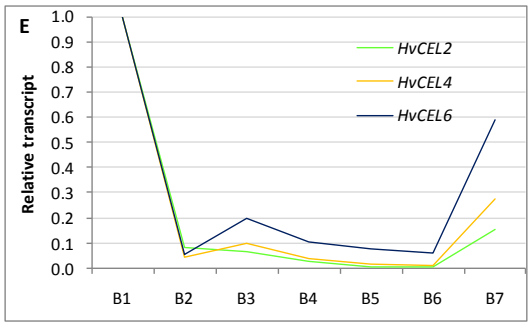

E
Relative transcript
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
HvCEL2
HvCEL4
HvCEL6
B1
B2
B3
B4
B5
B6
B7

Figure 3-12: Correlations between several genes analysed across the barley tissue and stem developmental series.

This series of graphs shows the relative transcript levels for some of these genes for each tissue with 1.0 representing the tissue in which that gene is most highly transcribed. Graph **A** shows those genes in the developmental tissue series that were highly correlated. Graphs **B** and **C** show relative levels of transcript for two sets of correlated genes across the stem tissue series. The same genes are represented in both graphs A and B. Similarly, graphs **D** and **E** represent the same genes in the tissue and stem series, however, *HvCEL6* is only weakly correlated with *HvCEL2* and *HvCEL4* in the tissue series. Actual correlations are presented in Table 3-1.

Table 3-1: Barley transcript correlations for some genes in the tissue series and stem series.

This table provides a summary of the graphs shown in Figure 3-12.

		Correlation	
Gene 1	Gene 2	tissue series	stem series
HvCEL1	HvCEL3	0.98	0.99
HvCEL1	HvCEL14	0.92	0.99
HvCEL3	HvCEL14	0.91	0.96
HvCEL5	HvCEL10	0.86	0.91
HvCEL2	HvCEL4	0.9	0.99
HvCEL2	HvCEL6	0.33	0.91
HvCEL4	HvCEL6	0.01	0.96

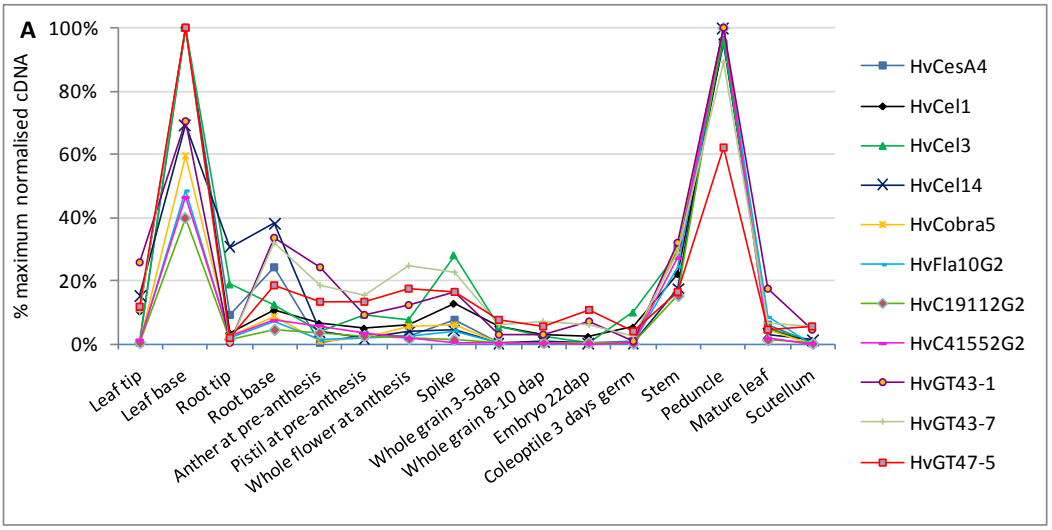
table (Table 3-2) and graphs (Figure 3-13) of all genes that showed a correlation coefficient, or r^2 >0.9 with endo-1,4-β-glucanases was detailed. *HvCEL1*, *HvCEL3* and *HvCEL14* are correlated with *HvCESA4*, known to be directly involved in cellulose synthesis in the secondary cell wall (Burton et al., 2004). Other genes showing strong co-transcription correlations included a fasciclin-like arabinogalactan protein (*HvFla10G2*), *Cobra 5*, and five transferase genes. These included β-3-glucuronyl transferase (*HvGT43-7*), β-glucurosyl transferase (*HvGt43-1*), glycosyl transferase (*HvGT47-5*), glucogenic glycosyl transferase (*HvC41552G2*) and α-galactotransferase (*HvC19112G2*). *HvCEL5* showed correlation coefficients of >0.9 with six expansin genes and xyloglucan endotransglycosylase 23 (*HvXET23*). *HvCEL8* co-transcribed with cellulose synthase-like D4 (*HvCSLD4*).

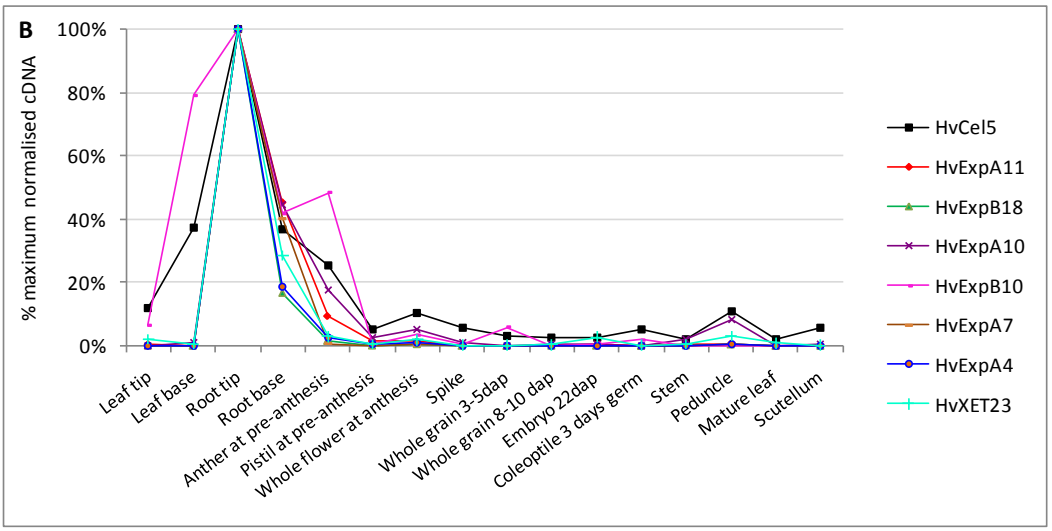
3.3.2. Maize

3.3.2.1. Maize MPSS

A search of the Dupont-Pioneer MPSS database found 24 signature tags for endo-(1,4)-β-glucanase genes and a list of genes with their signature tag sequences and positions is presented in Table 3-3. The signature tag for *ZmCEL30* has one base variation between the B73 genomic BAC sequence obtained from the maizesequence.org website and the EST sequence from the DuPont-Pioneer collection. This base change is in the GATC sequence positioned at the start of all signature tags. If this base change is a single nucleotide polymorphism, it may affect apparent transcript levels and provide some problems in interpretation of the MPSS data. Secondly, the *ZmCEL8* tag falls on an intron/exon boundary, and includes sequence from both intron and exon. This may represent a case of alternative splicing, or indicate that perhaps *ZmCEL8* is a pseudogene.

The MPSS database contains approximately 327 tissues, and data are presented as averages for the 12 core tissues, as determined by DuPont-Pioneer in Table 3-4. The core tissues comprise root, mesocotyl/coleoptiles, leaf, stalk, apical meristem, immature ear, ovary, embryo, endosperm, pericarp, silk, tassel/spikelet


A
% maximum normalised cDNA
100%
80%
60%
40%
20%
0%
Leaf tip
Leaf base
Root tip
Root base
Anther at pre-anthesis
Pistil at pre-anthesis
Whole flower at anthesis
Spike
Whole grain 3-5dap
Whole grain 8-10 dap
Embryo 22dap
Coleoptile 3 days germ
Stem
Peduncle
Mature leaf
Scutellum
HvCesA4
HvCel1
HvCel3
HvCel14
HvCobra5
HvFla10G2
HvC19112G2
HvC41552G2
HvGT43-1
HvGT43-7
HvGT47-5


B
% maximum normalised cDNA
100%
80%
60%
40%
20%
0%
Leaf tip
Leaf base
Root tip
Root base
Anther at pre-anthesis
Pistil at pre-anthesis
Whole flower at anthesis
Spike
Whole grain 3-5dap
Whole grain 8-10 dap
Embryo 22dap
Coleoptile 3 days germ
Stem
Peduncle
Mature leaf
Scutellum
HvCel5
HvExpA11
HvExpB18
HvExpA10
HvExpB10
HvExpA7
HvExpA4
HvXET23

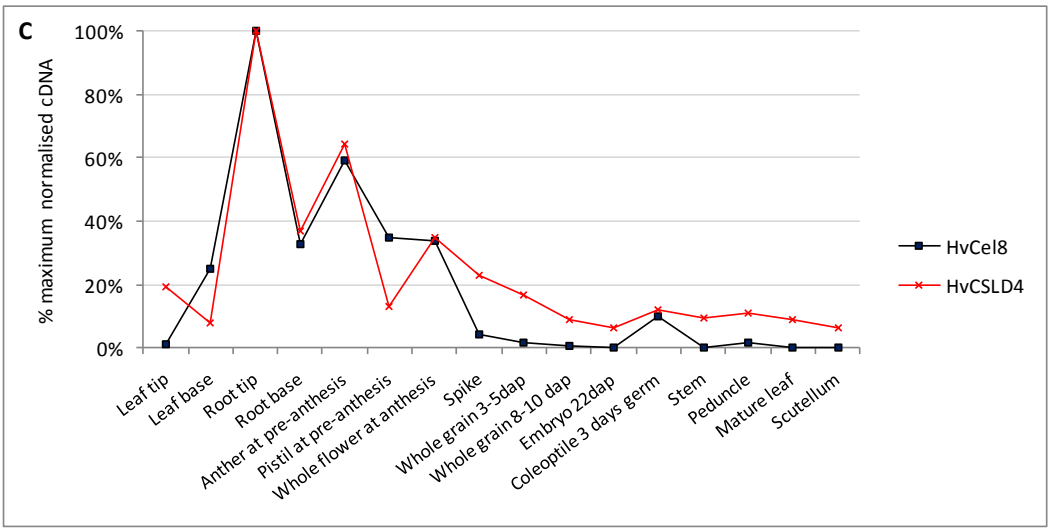

C
% maximum normalised cDNA
100%
80%
60%
40%
20%
0%
Leaf tip
Leaf base
Root tip
Root base
Anther at pre-anthesis
Pistil at pre-anthesis
Whole flower at anthesis
Spike
Whole grain 3-5dap
Whole grain 8-10 dap
Embryo 22dap
Coleoptile 3 days germ
Stem
Peduncle
Mature leaf
Scutellum
HvCel8
HvCSLD4

Figure 3-13: Graphs showing co-transcription patterns for the endo-1,4-β-glucanase gene family and other cell wall related families.

All three panels on the left illustrate transcript patterns as % maximum normalised cDNA transcript from the barley developmental series as outlined in section 3.2.5.1. **A**: *HvCEL1*, *HvCEL3* and *HvCEL14* co-transcription show major peaks of transcription in the leaf base and peduncle. Co-transcription is with 7 genes associated with secondary cell wall synthesis as described in the text. **B:** All of the genes showed maximum transcript levels in root tip. Genes co-transcribing with *HvCEL5* included 6 expansins and an XET. **C:** *HvCEL8* co-transcribed with HvCSLD4 and peak transcript occurred in the root tip.

Table 3-2: Correlation coefficients of the endo-1,4-β-glucanase family with other genes involved in synthesis of cell walls.

The table shows endo-1,4-β-glucanase genes that are co-transcribed with other genes involved in cell wall metabolism with correlation coefficients (r^2) >0.9.

HvFla10G2- fasciclin-like arabinogalactan protein
HvGT43-7- β-3-glucuronyl transferase
HvGt43-1- β-glucurosyl transferase
HvGT47-5- glycosyl transferase
HvC41552G2- glucogenic glycosyl transferase
HvC19112G2- α-galactotransferase
HvExp- expansin

	HvCel1	*HvCel5*	*HvCel8*	*HvCel3*	*HvCel10*	*HvCel14*
HvCesA4	**0.98**			**0.98**		**0.93**
HvCel1						
HvCel5						
HvCel8						
HvCel3	**0.98**					
HvCel10						
HvCel14	**0.91**			**0.90**		
HvExpA11		**0.93**				
HvExpB18		**0.92**				
HvExpA10		**0.93**				
HvExpB10		**0.92**			**0.96**	
HvExpA7		**0.92**				
HvExpA4		**0.92**				
HvCobra1					**0.92**	
HvCobra5	**0.95**			**0.94**		**0.92**
HvFla10G2	**0.92**			**0.91**		**0.92**
HvMybL				**0.93**		
HvC19112G2						**0.91**
HvC41552G2	**0.91**					**0.91**
HvXET23		**0.93**				
HvCSLD4			**0.92**			
HvGT43-1	**0.93**					**0.92**
HvGT43-7	**0.97**			**0.95**		
HvGT47-5	**0.95**			**0.93**		

and pollen. Three genes, *ZmCEL9*, *ZmCEL13* and *ZmCEL22* have transcript in pollen only, whilst a large number of genes show substantial transcript levels in meristem tissues. The data for the 10 genes with detectable transcript in maize B73 stem tissues including internode meristematic tissue, rind, vascular bundles, elongating, transition and mature zones of the internode and nodal plate have been extracted and are presented in Figure 3-13. These data show that *ZmCEL3*, *ZmCEL11*, *ZmCEL12* and *ZmCEL14* have high levels of transcript across all or most stem tissues analysed and *ZmCEL1*, *ZmCEL8*, *ZmCEL18*, *ZmCEL25*, *ZmCEL26* and *ZmCEL30* are transcribed at much lower levels and not in all stem tissues.

3.3.2.2. QPCR in a maize stem developmental tissue series

Although the structure of the maize stem is somewhat different to that of barley, an attempt was made to harvest the maize stem tissue series so that it aligned approximately with the stages of maturity of the barley stem developmental series. In general, transcript levels in maize were lower than those in barley. Agarose gels and measurement of RNA on the NanoDrop indicated that the RNA was of good quality and sufficient quantity. It is possible this may be the result of a lower proportion of mRNA in the total RNA extracted from the plant tissue, or due to higher levels of carbohydrate in the maize RNA, which might cause less efficient cDNA production during the reverse transcription of mRNA. In total, 11 genes were successfully analysed and a list of genes and their primers is presented in Table 5, Appendix A.

Figure 3-15, Figure 3-16, Figure 3-17 and Figure 3-18, should be looked at in conjunction with Table 3-5. Figure 3-15 shows that ZmCEL3, ZmCEL11 and ZmCEL14 have transcriptional activity across most or all internode tissues examined. ZmCEL12 transcript is chiefly in elongating tissues (Figure 3-16) and Figure 3-17 shows ZmCEL1 and ZmCEL18 with low levels of transcript mainly in the vascular bundles during elongation and early maturation. Figure 3-18 shows genes with low levels of transcript but with a peak in vascular bundles in the mature zone of the internode at V14.

Table 3-3: MPSS tag sequences and approximate position of tag in the ORF.

Nucleotide in *ZmCEL30*, shown in red, indicates variation between sequence of the EST on the Dupont-Pioneer database and that of the BAC database on the website maizesequence.org.

↓Gene name	Tag	Tag position	Sequence variations	Homoeologues
ZmCel1	GATCGGCTGCAAGGCCG	EXON5		
ZmCel2	GATCCATTCCTGTAAAA	3'UTR		
ZmCel3	GATCCTCGCCGCAAGGC	EXON4/EXON5		
ZmCel4	GATCTGCGGCGTCTTCG	EXON5		
ZmCel6	GATCTCGTGGGATAGAA	3'UTR		
ZmCel7Chr9	GATCTCGGGGTCAGTAA	3'UTR		7B no hits
ZmCel8	GATCAAACAGCTGCTGC	INTRON/EXON4		ZmCel29 no tag
ZmCel9	GATCCGGCCTGCTCTGC	3'UTR		
ZmCel10	GATCGGTTGCTCGCAGG	EXON3		
ZmCel11	GATCTGTGACTGTGAGA	3'UTR		
ZmCel12	GATCTGGGTGGCGAGGT	CHR10 EXON3		present in all 3
ZmCel13	GATCCATCCAAATCTCT	3'UTR		
ZmCel14	GATCCCTCCGATGTTTC	EXON6		
ZmCel15	GATCACCTGCAAGGGAG	EXON5		
ZmCel16	no tag			
ZmCel17	GATCGTCGGCGCGCTCG	MID EXON3		
ZmCel18	GATCGAGTGGATGGAGC	3'UTR		
ZmCel19	GATCCTCGCCTTCGTCA	EXON4		
ZmCel20	GATCGAGCAGAACGCGA	EXON5		
ZmCel21	GATCGGCTGCGACCAGG	EXON2		
ZmCel25	GATCCTAGCAGCAAGGT	EXON4/EXON5		ZmCel26
ZmCel26	GATCGTCGGCGGGCCGG	EXON5		ZmCel25
ZmCel29	no tag			ZmCel8
ZmCel30	GATCGTCGCTCAGGTGG	EXON2/EXON3	GATTGTCGCTCAGGTGG	ZmCel14
ZmCel32	GATCAGTTCATTTGCGA	EXON7		
ZmCel34	GATCACGACTGCTGGGA	EXON2		

Table 3-4: MPSS transcript data for maize genes in the 12 core tissues.

The table shows the average MPSS data as parts per million (ppm) in 12 core tissues for all endo-(1,4)-β-glucanase genes found on the Dupont-Pioneer MPSS database. In comparison with the barley tissue developmental series, the maize data shows similar general trends with some genes exhibiting transcript across many tissues while other genes are very tissue specific.

?Gene name	Root	Mesocotyl/Coleoptile	Leaf	Stalk	Apical Meristem	Immature Ear	Ovary	Embryo	Endosperm	Pericarp	Silk	Tassel Spikelet	Pollen
ZmCEL1	2	54	15	1	331	227	0	79	0	0	0	0	0
ZmCEL2	0.7	0	0	0.2	34	29	177	0	0	0	0	0	0
ZmCEL3	193	26	30	197.7	0.6	17.3	0	0	0	52	146	63	0
ZmCEL4	3	0	0	0	0	0	0	0	0	0	0	0	0
ZmCEL6	0.3	0	0	0	0	0	4	0	0	0	0	0	0
ZmCEL7	1	9	2	6	106	58	50	1	3	0	0	1	0
ZmCEL8	2	8	5	15	0	2	0	15	11	4	7	0	0
ZmCEL9	0	0	0	0	0	0	0	0	0	0	0	0	249
ZmCEL10	15	43	14	0.3	695	634	0	104	0	0	0	53	0
ZmCEL11	315	150	152	431	221	194	202	112	28	509	32	154	0
ZmCEL12	247	206	31	152	12	25	0	80	10	195	99	85	0
ZmCEL13	0	0	0.4	0	0	0.1	0	0	0	0	0	0	6394
ZmCEL14	38	20	26	64	3	0	30	6	95	100	55	0	0
ZmCEL15	0	0	0	0	0	0	0	0	1	0	0	0	0
ZmCEL17	0.2	0	0	0	0	0	0	0	0	0	0	0	0
ZmCEL18	4	9	6	1	10	4	0	2	0	0	0	22	0
ZmCEL19	0	0	0	0	0.2	0	2	0	0	0	0	0	0
ZmCEL20	98	7	0	0	1	5	0	49	0	6	0	0	0
ZmCEL21	34	8	1	0	119	122	0	39	1	0	0	4	0
ZmCEL25	11	0	3	6	0	0	0	0	0	0	0	0	0
ZmCEL26	2	0	3	9	0	0	0	0	0	0	0	0	0
ZmCEL30	0.3	0	0.1	2	0	0	0	0	0.2	0	0	0	0
ZmCEL32	0	0	0.1	0	0	0	0	0	0	0	0	12	0
ZmCEL34	0.1	0	0.3	0	0.5	0.2	3	0	0	0	0	0	0

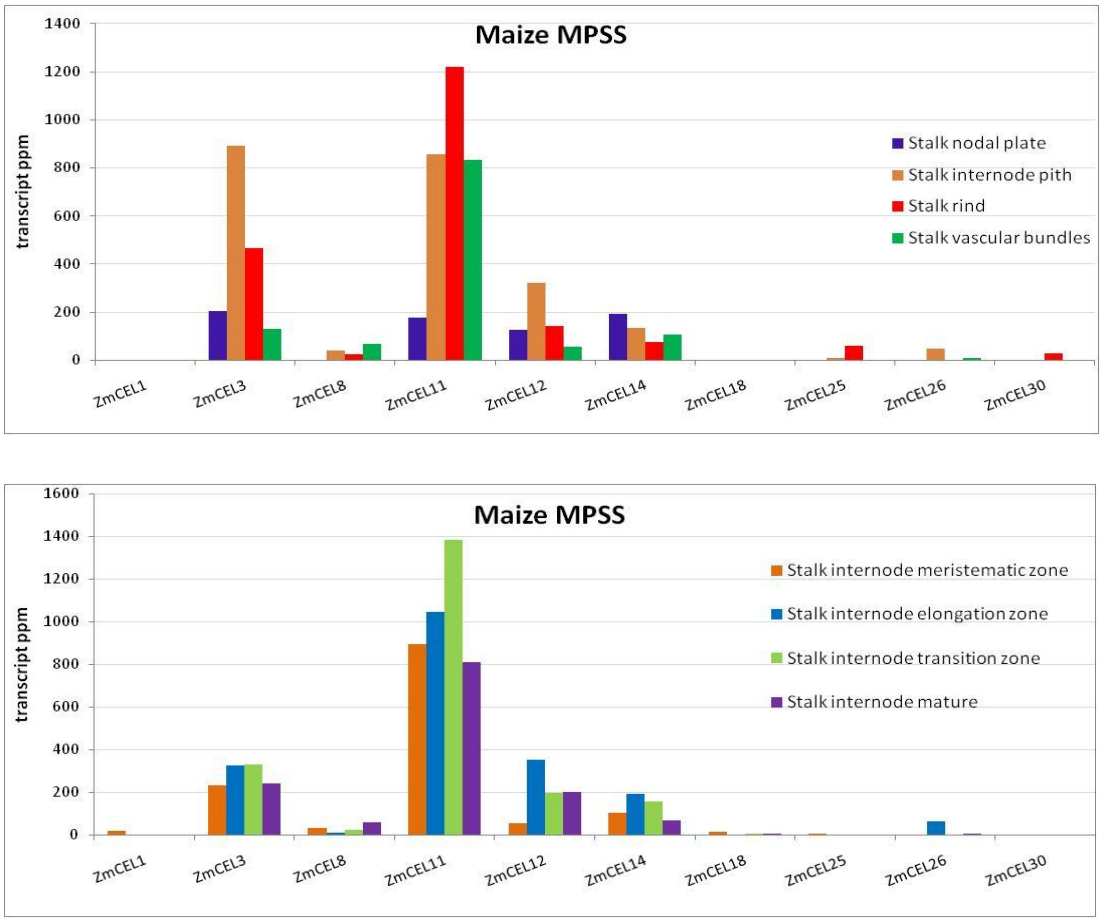


Figure 3-14: MPSS transcript data for stalk tissues in maize.

These two graphs show maize endo-(1,4)-β-glucanase genes with MPSS signature tags in stalk tissues. The top graph relates to different stalk tissues including nodal plate, and internode pith, rind and vascular bundles. The lower graph shows stalk tissues at different stages of development including internode meristematic zone, elongation zone, transition zone and mature zone. All tissues are from the V11-12 stage of development of the B73 maize variety.

Table 3-5**:** Nomenclature used for the various tissues and stages of internode development for the maize stem tissue series.

This table should be viewed in conjunction with Figures 3.15, 3.16, 3.17 and 3.18. See Section 3.2.3.3 (Morrison et al., 1994; Hill, 2007).

	Tissue	Internode 12
EARLY P	Pith	V11 early elongating internode
EARLY R	Rind	V11 early elongating internode
V14 P E	Pith	V14 elongation zone
V14 R E	Rind	V14 elongation zone
V14 VB E	Vascular bundles	V14 elongation zone
V14 P T	Pith	V14 transition zone
V14 R T	Rind	V14 transition zone
V14 VB T	Vascular bundles	V14 transition zone
V14 P M	Pith	V14 maturation zone
V14 R M	Rind	V14 maturation zone
V14 VB M	Vascular bundles	V14 maturation zone
V16 P T	Pith	V16 transition zone
V16 R T	Rind	V16 transition zone
V16 VB T	Vascular bundles	V16 transition zone
V16 PM	Pith	V16 maturation zone
V16 R M	Rind	V16 maturation zone
V16 VB M	Vascular bundles	V16 maturation zone
R2 P M	Pith	R2 maturation zone
R2 R M	Rind	R2 maturation zone
R2 VB M	Vascular bundles	R2 maturation zone

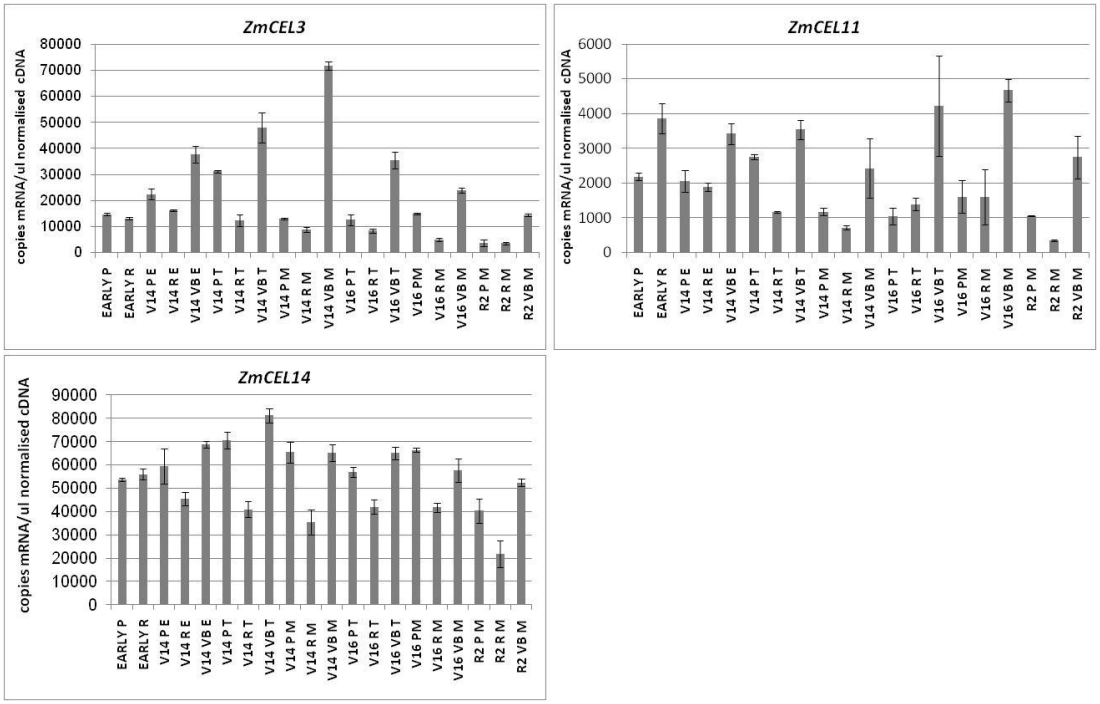


Figure 3-15: Transcript patterns for *ZmCEL3*, *ZmCEL11 ZmCEL14.*

These three genes showed transcripts for most internode tissues in the maize stem series. *ZmCEL3* and *ZmCEL11* showed marked peaks in vascular bundle tissues at all stages of development, relative to the pith and rind tissues. *ZmCEL14* has relatively high levels of transcript across all developmental stages, although rind tissues are transcribed at lower levels than the pith and vascular bundles.

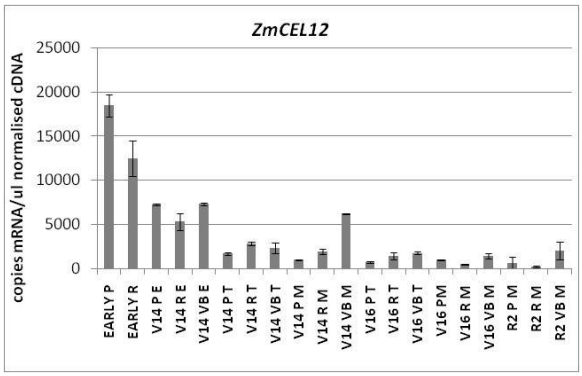


Figure 3-16: QPCR transcripts for *ZmCEL12* in the stem series.

In this graph, *ZmCEL12* shows transcript principally in young elongating tissues of the 12th internode in all young elongating tissues and in the vascular bundles during early secondary cell wall production.

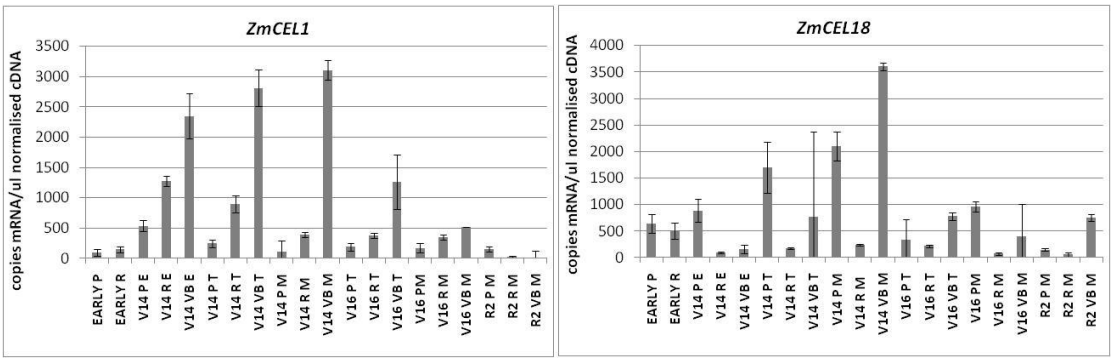


Figure 3-17: QPCR transcripts for *ZmCEL1* and *ZmCEL18.*

Analysis of *ZmCEL1* and *ZmCEL18* indicates low levels of transcript in vascular bundles of the maize stem during elongation and early secondary cell wall production in the stem tissue series.

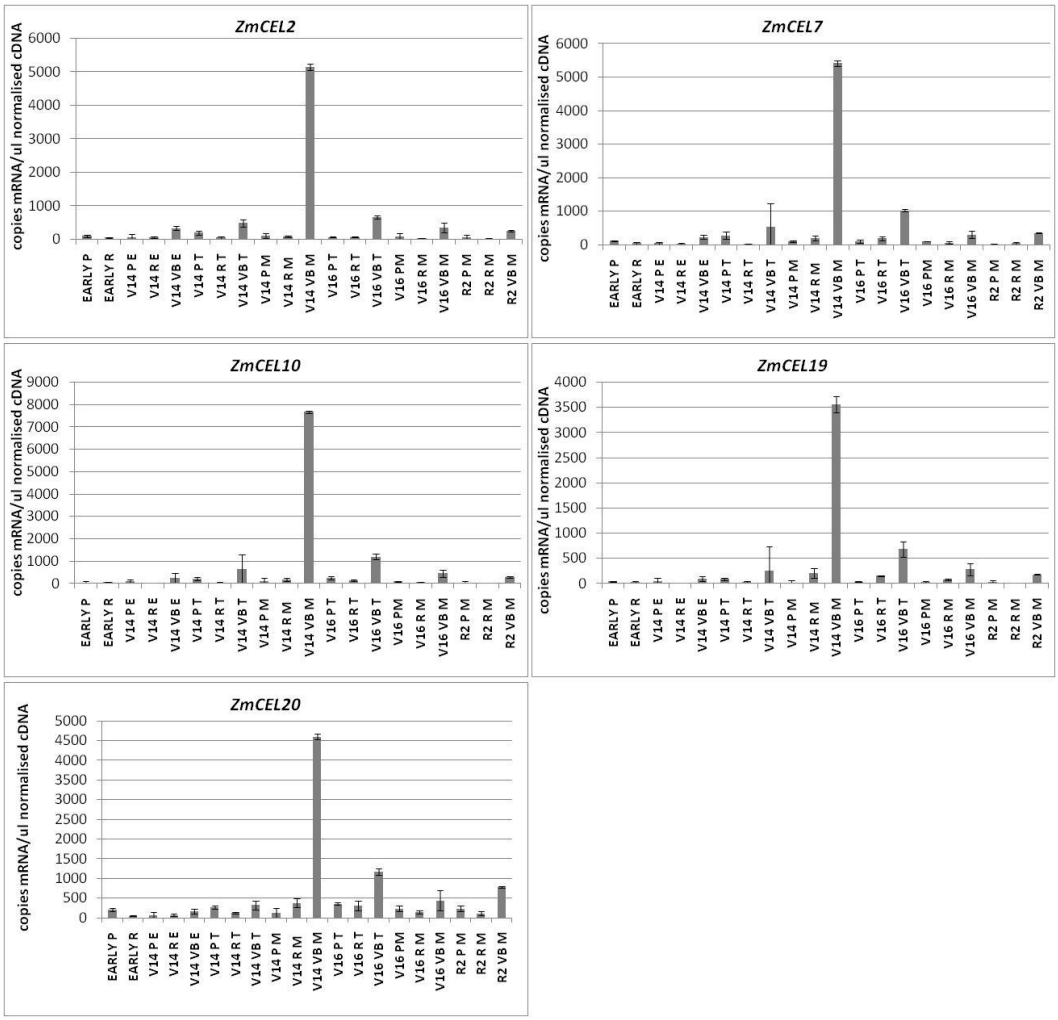


Figure 3-18: Maize stem tissue series QPCR profiles for for *ZmCEL2*, *ZmCEL7*, *ZmCEL10*, *ZmCEL19* and *ZmCEL20.*

Endo-(1,4)-β-glucanase genes transcribed in mature vascular bundles are shown here. QPCR results for the maize stem tissue series shows that transcript levels for *ZmCEL2*, *ZmCEL7*, *ZmCEL10*, *ZmCEL19* and *ZmCEL20* are low, but show a distinct peak of transcript in the mature vascular bundles of internode 12 at developmental stage V14.

3.3.2.3. Correlation of gene transcription in maize tissues

A transcriptional correlation analysis across the average of the 12 core tissues of the MPSS database showed that several sets of genes were correlated at $r^2 > 0.9$. Sets of genes with transcript correlations include *ZmCEL1*, *ZmCEL10* and *ZmCEL21*, with transcript in apical meristem, immature ear and ovary, and *ZmCEL9* and *ZmCEL13* co-transcribed in pollen. *ZmCEL2*, *ZmCEL6*, *ZmCEL19* and *ZmCEL34* are transcribed chiefly in ovary and root, however, except for *ZmCEL2*, display only background levels of signature tag abundance. When MPSS data for all stalk tissues was tested for gene correlations, only *ZmCEL25* and *ZmCEL30* showed a correlation of $r^2 > 0.9$.

A correlation matrix across the QPCR stem series transcript data showed that *ZmCEL1* and *ZmCEL3* were highly correlated at $r^2 > 0.9$, as were *ZmCEL2*, *ZmCEL7*, *ZmCEL10*, *ZmCEL19* and *ZmCEL20*. These results are presented graphically in Figure 3-19.

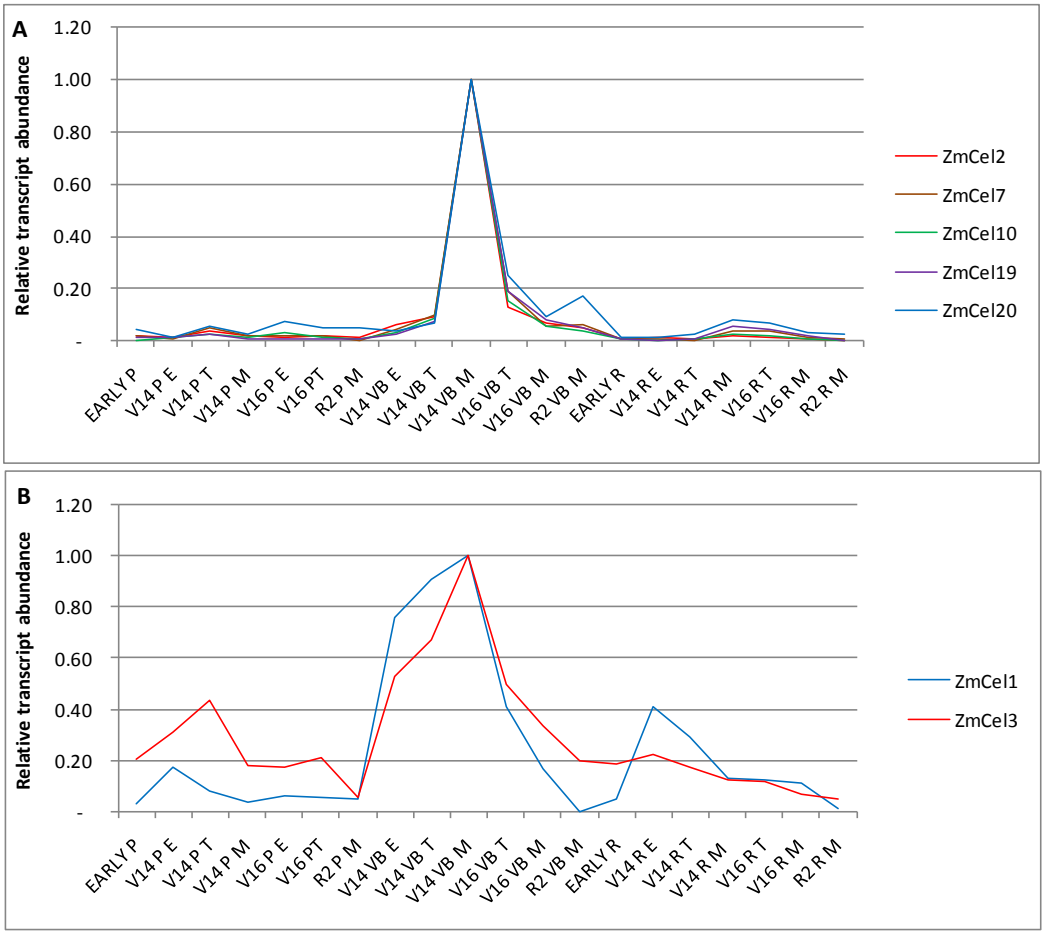


Figure 3-19: Graphs showing relative transcript levels for maize genes in the stem series.

Transcript correlations of 0.9 or greater are shown. **A**: *ZmCEL2, ZmCEL7, ZmCEL10, ZmCEL19* and *ZmCEL20* all had high transcript levels in mature zone of elongating vascular bundles, and **B**: similarly, *ZmCEL1* and *ZmCEL3* also had their highest levels of transcript in the mature zone of elongating vascular bundles, however, they also had transcript in some elongating tissues.

3.4. DISCUSSION

Several methods of transcript analysis were employed across different tissues in barley and maize. This yielded a broad range of transcript levels, and it was necessary to place some boundaries on how much transcript comprises a level above background. To this end a QPCR transcript level of <10,000 copies per µl normalised cDNA was considered low and below 1000 copies per µl cDNA was considered background noise. A moderate level of transcript was arbitrarily defined as between 10,000 and 100,000 copies per µl cDNA and above 100,000 copies was considered high. For the MPSS data, less than 5ppm was considered to be a background level of transcript, five to 50ppm was considered low (Appenzeller et al., 2004) and above 500ppm classified as high. As there were a large number of genes in the two species to be considered, each species will be dealt with separately and then the patterns of transcription between the species will be compared.

Barley transcript analysis

The barley developmental series data showed four patterns of transcription that related to both tissues and levels, which aided in grouping of the transcript data (Table 3-6). The first group includes *HvCEL1*, *HvCEL3* and *HvCEL14*, which shows variable levels of transcript across all tissues analysed. Transcript patterns of these three genes were highly correlated in the barley developmental series and also the stem series. Transcript levels were high in the vegetative tissues but were less abundant in the floral and developing grain tissues. In the stem series all three genes produced transcript for all developmental stages, and mRNA abundance peaked in the transition zone of the rapidly elongating internode.

The second pattern is exemplified by the gene pair *HvCEL5* and *HvCEL10*. In contrast to the first group, both genes show maximum levels of transcript in the transition zone of the rapidly elongating internode but also have a transcript peak in the young elongating internode. *HvCEL5* has a small peak at the grain fill stage that is not seen with *HvCEL10*. Transcript data for *HvCEL10* for the tissue

	BARLEY								MAIZE/BARLEY ORTHOLOGUES	MAIZE							
	Meristematic	Elongating	Mature	Vegetative	Floral	Developing Grain	Pollen			Meristematic	Elongation	Mature	Vegetative	Floral	Developing Grain	Pollen	
Group 1	N/A	+	+++	+++	+	-	N/A	HvCEL1	ZmCEL14	+	++	++	+++	+	++		Group 1 and Group 2
								HvCEL3	ZmCEL11	+++	+++	+++	+++	++	++		
								HvCEL14	ZmCEL25/26	+	+	+	+	-	-	-	
Group 2	N/A	+++	+++	+++	++	-	N/A	HvCEL5	ZmCEL12	+	+++	+	+++	++	+		
		+	+	+	+			HvCEL10	ZmCEL3	+++	+++	+++	+++	+	++		
Group 3	N/A	+++	+	++	+	+	N/A	HvCEL2	ZmCEL18		+			+	-	-	Group 3
								HvCEL4	ZmCEL1	+	+	+	+	++	-		
									ZmCEL2		-				+		
		+	+	+	+++	++	N/A	HvCEL6	ZmCEL7	++	-			++	+		
Group 4	N/A	+	-	+	+++	+++	N/A	HvCEL7	ZmCEL10	+++	-			+++	-		Group 4
								HvCEL8	ZmCEL1	+++	+	+	+	++	-	-	
		+	+	++	++	++			ZmCEL2	++	-				+	+	
								N/O	ZmCEL19	-	-			+	-		
		-	+	+	+++	+++		HvCEL11	ZmCEL20	+	-			++	-		
	N/A	+	+	+	-	-	N/A	HvCEL12	ZmCEL8/29	N/A	N/A	N/A	N/A	N/A	N/A	N/A	
								N/O	ZmCEL9	-	-	-	-	-	-	+	
								N/O	ZmCEL13							+++	

Table 3-6: Transcript patterns for endo-(1,4)-β-glucanase genes in barley and maize.

As described in the text, both barley and maize showed several different transcript patterns. Barley genes were loosely categorised into groups according to co-transcription patterns as indicated on the left hand side of the table and described in the text. On the right side of the table are the maize orthologues and their subjective groupings as described in the text. The maize orthologues of *HvCEL4* and *HvCEL8* could not be exactly deetermined, *ZmCEL1* or *ZmCEL2* are their closest orthologues as determined by phylogenetic analysis in Chapter 2.

+++	high relative levels of transcript
++	moderate relative levels of transcript
+	low relative transcript levels
-	no transcript detected
N/A	not analysed
N/O	no orthologue

developmental series has very large standard deviations but the correlation data indicates that these genes are co-transcribed in most tissues. *HvCEL2*, *HvCEL4* and *HvCEL6* are co-transcribed across the stem series. All of these genes have a peak of activity at the early elongating phase and a smaller peak after flowering as the grain fills. On the other hand, in the barley developmental series, *HvCEL2* and *HvCEL4* show a significant co-transcriptional correlation whereas *HvCEL6*, with very low transcript numbers in vegetative tissues and significantly higher levels in floral and developing grain produces a different pattern of transcription.

HvCEL7, *HvCEL11* and *HvCEL12* all have very low transcript abundance in the stem series but in the tissue series *HvCEL11* has high levels in pistil, flower and whole grain 3-5 days after pollination. *HvCEL7* is transcribed in root tip, pistil, whole flower and whole grain at 3-5 days post pollination. *HvCEL12* shows low levels of transcript in the tissue series as well as the stem series.

An interesting dichotomy exists for *HvCEL12* between the PLEXdb microarray data and the tissue and stem series QPCR data. The microarray data shows consistently moderate levels of fluorescence intensity for all tissues of between $\log_2 9$ and 10, whereas for both sets of QPCR data transcript levels are negligible. It is unusual for the microarray fluorescence intensity to be similar for all tissues. For all the other endo-(1,4)-β-glucanase genes relative fluorescence intensity was variable across the different tissues. Perhaps the *HvCEL12* microarray data are the result of non-specific probe binding. In other cases where tissues and their developmental stage are similar, patterns of transcription are also similar, for example *HvCEL11* and *HvCEL1*, *HvCEL3* and *HvCEL14* as well as *HvCEL5*. Although primers were carefully assessed before commencement of the analysis, this differential mRNA abundance between the microarray data and the QPCR may be the result of poor primer annealing or difficulty in amplifying the product. It has been shown that attempting to compare transcript data across different platforms, different tissues at different times of development and from different plant varieties can pose problems, however, generalised singular value decomposition (GSVD) makes it possible (Schreiber et al., 2008). Although GSVD was not performed other information was available that allowed for some speculation with regard to *HvCEL12* transcription. Interestingly, *HvCEL12* is an

orthologue of the maize homoeologues *ZmCEL8* and *ZmCEL29* (see Chapter 2). Neither of these maize genes was represented in the QPCR analysis because only short sequence and no 3' UTR was available on the EST database. Failure to detect cDNA sequence in EST databases is suggestive of low transcript abundance. Also, *ZmCEL29* was not represented in the MPSS data and *ZmCEL8* had MPSS data that showed it to have low transcript in vegetative tissues only. These findings suggest that the QPCR data for *HvCEL12* may be more likely than the higher mRNA abundance associated with the microarray data.

The role of endo-1,4-β-glucanases in cell wall synthesis was explored in the context of co-transcription with other genes known to have involvement in cell wall metabolism. A very strong relationship was seen between *HvCESA4* and *HvCEL1* and *HvCEL3*, where a co-transcription correlation coefficient, r^2, of 0.98 was determined. *HvCEL14* was also highly co-transcribed at $r^2 = 0.93$. From transcript and mutant plant analysis, *HvCESA4* is strongly believed to be part of the barley secondary cell wall cellulose synthesis machinery (Burton et al., 2004; Burton et al., 2010). Other genes with known secondary cell wall involvement that were co-transcribed with *HvCEL1*, *HvCEL3* and *HvCEL14* at $r^2 > 0.9$, included *COBRA 5* and fasciclin-like arabinogalactan protein (*HvFLA10G2*). Cobra gene mutants have been shown to dramatically reduce cell wall thickness and cellulose levels in the walls of rice and maize stems (Li et al., 2003; Ching et al., 2006). Involvement of COBRA proteins in the secondary cell wall metabolism is highly likely, however the exact nature of its involvement has not been determined. An association of FLA genes with tension wood in poplar has been determined with increased transcription and expression during tension wood production (Lafarguette et al., 2004). Tension wood is associated with high levels of cellulose.

Other genes that co-transcribed with *HvCEL1*, *HvCEL3* and *HvCEL14* included five different glycosyl transferase genes, and these gene families are believed to play a role in cell wall modification. At least 92 families of glycosyl transferases have been identified, however, most remain uncharacterised (Carbohydrate Active Enzyme database, http://www.cazy.org/; Coutinho et al., 2003; Cantarel et al., 2009). In plants, glycosyl transferases aid in the production of cell wall

carbohydrate chains by catalysing the formation of glycosyl bonds between sugar moieties and acceptor polysaccharide chains.

HvCEL5 was similarly co-transcribed at $r^2 > 0.9$ with several genes that are associated with cell elongation. This list included six expansins and a xyloglucan endotransglycosylase. Both of these gene families have been implicated in cell wall loosening to allow the cellulose crystalline microfibrils to move relative to each other and allow elongation to occur. Although the expansin family appear to possess no enzymatic capacity their role in wall loosening and cell elongation has been well established for many years (McQueen-Mason et al., 1992; Cosgrove and Durachko, 1994). Moreover, endo-(1,4)-β-glucanases were also found to enhance the loosening effect of the expansins. XETs have also been implicated in wall modification by way of hetero-transglycosylation (Fry, 1993; Thompson and Fry, 2001; Hrmova et al., 2007; Hrmova et al., 2009), co-transcription with expansins and *HvCEL5* implicate *HvXET23* with a role during cell wall loosening and cell elongation.

The majority of transcripts for *HvCEL8* were in root tips with some transcription in floral tissues, while transcripts for the stem series were confined to early stem elongation and post-flowering. *HvCEL8* co-transcribed with *HvCSLD4* at $r^2 = 0.92$. While the CSLD family of genes are not fully characterised recent literature suggests a role in early cell wall development. Rice plants with a *csld* mutant have aberrant stem and root tip cell walls (Li et al., 2009). In *Arabidopsis*, a role for *CSLD* in tip growing cells has been proposed (Bernal et al., 2008).

Maize transcript analysis

The maize genes also had varying patterns and levels of transcript abundance, as summarised in Table 3-6. Genes with consistently high transcript levels across the MPSS data and stem series were *ZmCEL3* and *ZmCEL14*. *ZmCEL11* had consistently high transcript in the MPSS database for all tissues except pollen however QPCR showed low transcript abundance generally, but with peaks in all vascular bundle tissues and early elongating rind. This may indicate that the QPCR amplification was inefficient, perhaps sequence variation reduced primer

binding efficiency or there may have been secondary structure in the DNA that slowed amplification (Livak and Schmittgen, 2001).

ZmCEL12 stem tissue transcript as measured by QPCR was restricted to elongating tissues, whereas MPSS transcript data indicated that the gene was transcribed across most tissues except pollen and tassel spikelet. *ZmCEL12* has three copies in the maize genome and all have the same MPSS signature tag. Two of these three genes have very different 3'UTRs but at the time of performing the maize QPCR, it was not known that there were three copies of this gene in the genome and it is probable that MPSS data represents total transcript for all three genes. The issue of homoeologues and duplicates is dealt with later in the Discussion.

As with the barley QPCR, some maize genes also exhibited very low levels of transcript in stem tissues. *ZmCEL2*, *ZmCEL7*, *ZmCEL10*, *ZmCEL19* and *ZmCEL20*, all had relatively low levels of transcript (between 3,500 and 7,800 copies per µl normalised cDNA) and all had elevated levels during the V14 stage of development in the vascular bundles of the stem. *ZmCEL1* and *ZmCEL18* also exhibited low levels of transcript (< 3,500 copies per µl normalised cDNA) in the stem tissue series. For all seven genes, product was successfully amplified from a cDNA cocktail of stem tissues, so some vegetative gene activity is implied. MPSS tag abundance for all of these genes showing low QPCR transcript in the stem series is quite interesting. *ZmCEL10* has very high levels of transcript in meristematic and immature ear tissues and *ZmCEL2* and *ZmCEL7* have lower levels of transcript in those same tissues. *ZmCEL20* has high levels of transcript only in root and *ZmCEL19* has very low levels only in immature ear. These results indicate that this group of genes has high tissue specificity. By contrast, *ZmCEL1* has quite high levels of transcript in a variety of tissues including root, meristematic tissues, immature ear, leaf and embryo, and similarly *ZmCEL18* has significant but lower levels of transcript in those same tissues and also in mature stalk.

ZmCEL13 was shown from MPSS data to have a very high transcript level in pollen only, and *ZmCEL9* was also exclusively transcribed in pollen but at lower

levels. As they showed no transcript in vegetative tissues, neither of these genes was analysed by QPCR across the stem tissue series. Loss of activity of these genes in the production of genetically modified field crops could prove useful for production of sterile pollen to prevent cross-pollination with nearby crops.

Comparison of transcript between maize and barley in stem tissues

It was mentioned earlier that comparison of transcript across different platforms, tissues, and developmental stages can be problematic (Schreiber et al., 2008), so comparison of different species is bound to cause even greater difficulties. Schreiber et al., (2009) compare AffyChip microarray data of wheat and barley and show that transcript for barley is close to eight times greater than that for wheat. QPCR was used to confirm that in many cases wheat had lower levels of transcript compared to barley and led to speculation that a possible shift in regulation of genes from transcription to translation had occurred. In this study, when compared to the barley QPCR data, maize transcript levels were generally much lower, and this included the control genes. Further work comparing transcript levels with that of protein levels as proposed in Schreiber et al. (2009) may help to determine whether regulation is occurring at the level of transcription or translation.

Despite these specific problems, patterns of transcriptional similarity between the two species were observed and are presented in Table 3-7. The barley *HvCEL1*, *HvCEL3*, *HvCEL5*, *HvCEL10* and *HvCEL14* genes have significant levels of transcript in the stem developmental series as well as vegetative tissues in the barley tissue series. The maize orthologues for these barley genes are *ZmCEL14*, *ZmCEL11*, *ZmCEL12*, *ZmCEL3* and the orthologous pair *ZmCEL25* and *ZmCEL26* respectively. It should be noted that the *ZmCEL25/26* genes were not analysed by QPCR as no 3' UTR sequence was available. As with the barley orthologues, all of these maize genes provide the majority of the transcript found during maize stem development.

Finally, *HvCEL6* and *ZmCEL7* are orthologous and show similar transcription patterns, with the major transcript levels during floral and early grain development. These observations provide evidence that a degree of evolutionary

Barley orthologue	Barley tissue series	Barley stem series	Sub-family	Maize MPSS	Maize stem series	Maize orthologue
HvCEL1	mostly vegetative tissues	elevated all tissues, peak in mature zone of elongating internode	GH9A	all tissues except ovary, pollen and tassel	high all stem tissues	*ZmCEL14*
HvCEL3	most tissues, not embryo	elevated all tissues, peak in mature zone of elongating internode	GH9B clade B1	high, all tissues except pollen	all vascular bundle tissues	*ZmCEL11*
HvCEL5	high transcript levels in all tissues	high in all stem tissues, peak early elongating internode and mature zone elongating internode	GH9A	high most tissues except ovary and pollen	elongating internode tissues	*ZmCEL12*
HvCEL10	vegetative tissues only	elevated early elongating internode and mature zone elongating internode	GH9B clade B3	all tissues, except ovary, embryo, endosperm, pollen	vascular bundles and all elongating tissues	*ZmCEL3*
HvCEL14	vegetative tissues only	elevated mature zone of elongating internode	GH9B clade B3	vegetative tissues only	not analysed	*ZmCEL25/26*
HvCEL6	floral tissues and developing grain, low root tip	early elongating internode and post flowering	GH9A	meristem and ovary, also low vegetative, embryo, endosperm	low mature zone elongating vascular bundles	*ZmCEL7*

Table 3-7: A comparison of transcript between barley and maize orthologues.

The blue cells on the left show a summary of transcript in the tissue developmental series and the stem series (Figure 3-4) for selected barley genes. On the left in yellow, the maize genes from the MPSS data (DuPont-Pioneer) and the maize stem series (Figure 3-5). The table shows the similarities in transcription of orthologues between the two species as determined by the phylogenetic analysis in Chapter 2.

conservation of transcriptional activity is present between barley and maize.

The six orthologues just discussed are from the GH9A, GH9B3 and GH9B1 subfamilies of genes as described in Chapter 2, section 2.3.2 and are interesting for several reasons. In the phylogenetic analysis of the endo-(1,4)-β-glucanase gene family, GH9A and GH9B1 appear to be derived from the same ancestral gene although categorised into different sub-families on the basis of protein domain structure. Furthermore, the GH9A sub-family, also referred to as KORRIGAN, is reported to be associated with cellulose production (His et al., 2001; Lane et al., 2001; Molhoj et al., 2001; Sato et al., 2001; Szyjanowicz et al., 2004; Bhandari et al., 2006). In the cereals examined here, the GH9B3 clade has expanded from one gene to four in maize, or five genes in sorghum and rice (Chapter 2, section Endo-(1,4)-β-glucanase amino acid sequence alignment). This expansion has occurred after the separation of cereals from the dicots, with only one equivalent orthologue in *Arabidopsis*. High levels of transcription of five of the six genes in barley and maize stem tissues supports speculation that these genes are involved in cellulose and/or cell wall synthesis, adding to the co-transcription evidence of genes involved in cell elongation, cell wall modification and cellulose synthesis.

Transcript analysis and homoeology in maize

The fate of duplicate genes in a genome can be several-fold. Silencing or pseudogene production is the most usual outcome (Lynch and Conery, 2000). However, less gene loss occurs with polyploidisation, as is the case in maize, because stoichiometry, or relative proportions of components in biochemical processes, is preserved (Lynch and Conery, 2000). Redundancy, whereby both genes are transcribed equally but both at lower levels, is also a possible fate for duplicated genes, as this option also preserves stoichiometry (Force et al., 1999; Lynch and Conery, 2000; Lynch and Force, 2000; Lawton-Rauh, 2003). Subfunctionalisation occurs with duplications when "members of a pair experience degenerative mutations that reduce their joint levels and patterns of activity to that of the single ancestral gene" (Lynch and Force, 2000). This may be a first step to silencing, or neofunctionalisation, that is, evolution of a new

function for one of the genes. The least likely outcome is that of neofunctionalisation, as seen with the Darwinian model (Kimura and Ohta, 1974). Transcriptional analysis may be a useful tool in monitoring the fate of gene duplicates or homoeologues. The homoeologously duplicated gene *ZmCEL12* was mentioned earlier in relation to variation in activity within the stem developmental series QPCR compared with that of the MPSS data.

The QPCR transcript for *ZmCEL12* was relatively much lower than expected because MPSS data owed high transcript levels for this gene. There are three genes present on the genome for *ZmCEL12*, and it is possible that redundancy has occurred with these genes and that QPCR is possibly only measuring transcripts from one gene. Using primers derived from the 3'UTR region may assist in differentiating transcripts for two of these three genes. The transcript sequence for *ZmCEL12Chr1* and *ZmCELChr10* is identical but *ZmCEL12Chr2* is a homoeologue and not a duplicate, and has a unique 3'UTR .

The gene pair *ZmCEL14* and *ZmCEL30* are most likely duplicates that separated prior to allotetraploidy (Chapter 2, figure 2.14C). MPSS data indicates that while *ZmCEL14* is actively transcribed at relatively high levels in a large variety of tissues, particularly vegetative tissues, *ZmCEL30* has low transcript in only two tissues, namely pericarp and rind (Table 3-4). It is possible that *ZmCEL30* has been made redundant, however, the signature tag sequence for this gene on the EST database has a single base difference to that of the genomic sequence found on the *maizesequence.org* website, perhaps the result of a single nucleotide polymorphism. This complicates interpretation of the MPSS data and very likely results in the much lower transcript for *ZmCEL30*. This dilemma could be resolved with QPCR using unique primers for each homoeologue.

MPSS data show that *ZmCEL25* and *ZmCEL26* transcripts are found exclusively in vegetative tissues (Table 3-4) and although *ZmCEL26* appears to have slightly lower levels of transcript in fewer tissues, this may simply be due to tag position (Table 3-3). No QPCR data are available, but the MPSS data suggest that there is redundancy with these genes.

ZmCEL7 is one of a homoeologous pair but MPSS data and QPCR data are only available for one of the pair. Does the absence of MPSS transcript data for the second homoeologue mean subfunctionalisation, pseudogene formation or simply that no tag was produced for the second homoeologue? Again, QPCR analysis would help in solving this problem.

The final homoeologous pair is *ZmCEL8* and *ZmCEL*29. At the time of QPCR analysis, only 60 bases of uncertain reliability of the *ZmCEL8* gene 3'UTR was available and no 3' UTR had been defined for *ZmCEL29*, so QPCR was not possible for these genes. *ZmCEL29* does not have an MPSS tag and furthermore, the MPSS tag for *ZmCEL8*, present in the cDNA contig sequence, spans an intron and exon boundary of the BAC sequence (Table 3-3). This may mean that this gene is alternatively spliced or is perhaps a pseudogene. The *ZmCEL29* sequence is partial, as the BAC on which it is positioned has a break in sequence at the 5'end of the gene. This provides the possibility that *ZmCEL29* is being silenced or has become a pseudogene. As with *ZmCEL7*, lack of an MPSS tag does not in itself mean that the gene is not actively transcribed, and again, QPCR may help resolve this problem.

Endo-(1,4)-β-glucanases and the cell wall

So little is really known of the *in planta* activity of endo-(1,4)-β-glucanases that it is difficult to do more than speculate on their actual roles in stem development. The *in vitro* hydrolytic activities of only a few of these genes in other species are known and gene transcription is not necessarily a measure of protein expression, but let us for the moment assume that it is and follow the trail of endo-(1,4)-β-glucanase genes in maize as the cells of the stem internode divide, elongate, mature and senesce (Figure 3-20). The endo-(1,4)-β-glucanase, KOR1, is located at the cell plate in Arabidopsis during cytokinesis and mutations in this gene produce cells with incomplete cell walls (Zuo et al., 2000). In the maize internode, meristematic tissues provide a source of cells undergoing cytokinesis and the orthologue for *KOR1* in maize is either *ZmCEL7* or *ZmCEL14*. The DuPont-Pioneer MPSS data in maize indicates that 14 endo-(1,4)-β-glucanases, over one half of the total number of maize endo-(1,4)-β-glucanase genes with a

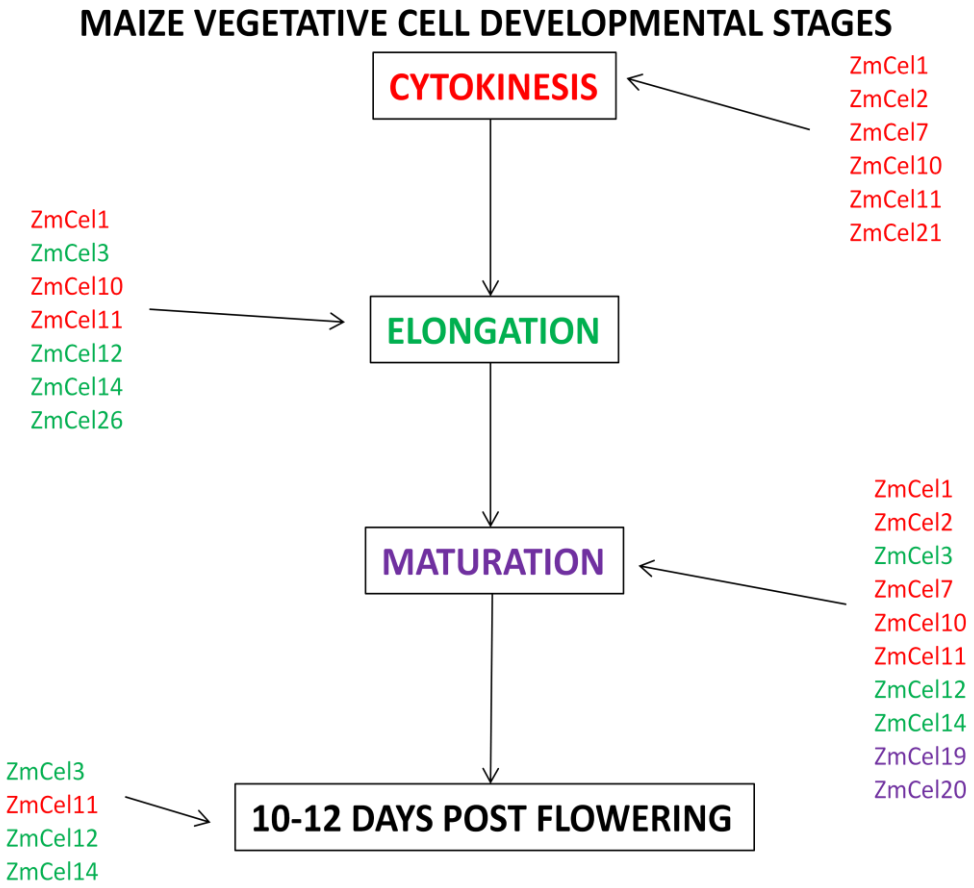


Figure 3-20: Shows a schematic representation of the life of a cell with the maize genes actively transcribed at each stage of development.

Genes transcribed at cytokinesis are in red and maintain that colour scheme for all stages of development. Green represents those genes initially found to be transcribed at elongation and purple represents those genes first transcribed in maturing cells. This schema demonstrates the possible involvement of the endo-(1,4)-β-glucanase gene family in cell wall production.

signature tag, are transcribed in the apical or internode meristem (Esau, 1977; Dickison, 2000). Six of those 14 genes have significant transcript levels.

As the cell starts to elongate it can be envisioned that further endo-(1,4)-β-glucanase activity is required to assist in cell wall creep, since loss of such activity perturbs cell elongation (Nicol et al., 1998; Sato et al., 2001). During elongation, cell wall deposition will also be occurring (Ray, 1962) and in rice, an insertional mutation of the AtKOR1 orthologue, OsGLU1, produced a dwarf plant with cells that failed to elongate fully (Zhou et al., 2006). Combined data for the maize MPSS database and stem tissue series shows 11 genes with transcripts in the elongating tissues, seven of which were at significant levels.

The end of elongation is followed by deposition of the secondary cell wall and an increase in cellulose (Azuma et al., 1996; Appenzeller et al., 2004). At this stage of development, endo-(1,4)-β-glucanases may be part of the cellulose production mechanism, but *in vitro* activity of the enzymes suggest they may also be involved in cell wall matrix modification (Molhoj et al., 2001; Hayashi et al., 2005; Yoshida and Komae, 2006; Urbanowicz et al., 2007b). In maize, 10 genes have transcript in maturing tissues. The final stage in the life cycle of the cell is that of transportation of previously stored energy in the form of carbohydrates to the developing grain. This is known to involve transport of soluble starch, but also may incorporate a component of hydrolysed cellulose or non-cellulosic polysaccharides (Liebhardt et al., 1968; Roulin et al., 2002). Approximately 10 days after flowering, as the grain develops, maize has four genes in the mature internode that are being actively transcribed.

Although the data presented above do not represent all genes in all tissues, nor do transcripts represent active proteins, they do illustrate the point that the endo-(1,4)-β-glucanase genes are active across the whole spectrum of vegetative cell development, from initial cell division to post flowering.

3.5. SUMMARY

For barley, Affymetrix Chip transcript data was obtained from the PLEXdb (Druka et al., 2006; Wise et al., 2007) database, and QPCR data was obtained from a tissue series kindly donated by Dr Rachel Burton, containing 16 different barley tissues. QPCR was also performed on a barley stem series comprising internode tissues at different stages of development. The maize transcript data was obtained from the Dupont-Pioneer proprietory MPSS database and a maize B73 stem tissue series comprising the different stages of the developing internodes. Although this analysis is by no means comprehensive, it illustrates the temporal and spatial variation in transcription of genes encoding endo-(1,4)-β-glucanases. It also signals the possibility that several of the members of endo-(1,4)-β-glucanase gene family may be working coordinately in cell wall formation.

Despite the problems associated with comparing transcripts from various sources, the data here strongly suggest that in barley the major endo-(1,4)-β-glucanases involved in stem development are *HvCEL1*, *HvCEL3*, *HvCEL5*, *HvCEL10* and *HvCEL14*. Moreover, transcript patterns for several of these genes were closely correlated. These five barley genes were orthologues of the maize genes *ZmCEL3*, *ZmCEL11*, *ZmCEL12*, *ZmCEL14* and *ZmCEL25* and *ZmCEL26*, which also provide the majority of transcript in vegetative tissues. Together with orthologues *HvCEL6* and *ZmCEL7*, whose main transcripts were found in floral and developing endosperm, these findings suggest evolutionary conservation of transcription of endo-(1,4)-β-glucanase genes between barley and maize. Transcript patterns between other orthologues of maize and barley are not necessarily similar, which may be due to dissimilarity between the tissues compared for these two species or to some drift in activity of these genes since separation of the common ancestors of the two species. Analysis of co-transcriptional patterns of the different endo-(1,4)-β-glucanase genes with other cell wall metabolism genes in barley showed that a strong association of both a temporal and spatial nature exists. This confirmed a role for several endo-(1,4)-β-glucanase genes in cell elongation and secondary cell wall metabolism.

In the next Chapter the complexity of cell wall breakdown is explored. It describes attempts at expressing endo-1,4-β-glucanase genes heterologously in microbial systems with a view to combining several coexpressing enzymes in cell wall loosening and breakdown experiments.

Chapter 4

Heterologous Expression of Endo-(1,4)-β-glucanases in Escherichia coli and Pichia pastoris

4.1. INTRODUCTION

Current published literature provides evidence that the endo-(1,4)-β-glucanases are involved in plant cell wall synthesis during all stages of development of the cell wall. This includes cell plate formation (Zuo et al., 2000), primary cell wall synthesis and cell elongation (Nicol et al., 1998; Sato et al., 2001; Zhou et al., 2006) and secondary cell wall synthesis (Joshi et al., 2004; Bhandari et al., 2006). The results of the transcript analysis presented in Chapter 3 showed that a number of the barley endo-(1,4)-β-glucanase genes are variously transcribed during several stages of stem development, including post-flowering. A correlation analysis showing high levels of co-transcription between several members of the endo-(1,4)-β-glucanase gene family is also presented in Chapter 3, Figure 3-12 and Table 3-1. Furthermore, a correlation analysis of transcripts linked several members of the endo-(1,4)-β-glucanase gene family with genes from other families involved in cell wall synthesis (Figure 3-13 and Table 3-2). These genes include HvCESA4, known to be part of the secondary cell wall cellulose synthesis machinery. Strong co-transcription was also found to exist between several endo-(1,4)-β-glucanases and an XET, and several expansins and glycosyl transferases, all of which are implicated in cell elongation.

Cell elongation necessitates loosening or breakdown of the cross-linkages that stabilise the crystalline cellulose micofibrils in the cell wall. The effects of hydrolytic enzyme activity on the cell wall during cell elongation have been investigated *in vitro* using cellulose/xyloglucan composites produced by *Acetobacter xylinus* (Chanliaud, 2004). The composites contained 78% cellulose and 22% xyloglucan. Incubation of these composites under tension with the *Trichoderma viride* endo-(1,4)-β-glucanase resulted in depolymerisation of xyloglucan and a concomitant increase in the composite's ability to withstand strain forces, that is strength was increased, but the wall became less elastic. Further evidence of the involvement of endo-(1,4)-β-glucanases in elongating

tissues has been found using plants with that have reduced endo-(1,4)-β-glucanase activity. *Arabidopsis* plants carrying either the *kor1-1* or *kor1-2* mutant, show extreme dwarfism due to failure of the cells to elongate (Nicol et al., 1998; Zuo et al., 2000). In rice, dwarf plants with cells that have failed to elongate were created by an insertion mutant in the KORRIGAN endo-(1,4)-β-glucanase orthologue, OsGLU1 (Zhou et al., 2006).

For some time, expansins have been reported to play a crucial role in cell wall loosening (Cosgrove and Durachko, 1994; McQueen-Mason and Cosgrove, 1994). Moreover, several members of the expansin family of genes have been shown to produce transcript during internode elongation in rice (Lee and Kende, 2001; Lee and Kende, 2002). Using similar *Acetobacter xylinus* cell wall composites to those described earlier, incubation under tension with expansins alone provided significant cell wall extension (Whitney et al., 2000). This work served to reinforce that of Cosgrove and Durachko (1994) in which cucumber hypocotyls were shown to elongate when incubated with expansins alone when all other cell wall enzymes had been destroyed. It was however, also found that the addition of a bacterial cellulase enhanced wall extension (Cosgrove and Durachko, 1994).

A role for endotransglycosylase/hydrolases (XET/XTH) in cell elongation and cell wall loosening has been proposed (Nishitani and Tominaga, 1992; Thompson and Fry, 2001). Using the *Acetobacter xylinus* composite described above, incubation with the tomato XET/XTH alone, while stress forces were being applied, had only minor effects on strength and elasticity but creep capacity was increased (Chanliaud, 2004). The principle activity proposed for the XET/XTHs is that of cell wall remodelling using endohydrolase and transglycosylase activity (Thompson and Fry, 2001; Hrmova et al., 2007). *In vitro* studies indicate that HvXET5 from barley catalyses covalent linkages between xyloglucans and cellulose or (1,3;1,4)-β-D-glucans (Hrmova et al., 2007). Barley cell walls contain predominantly arabinoxylans rather than

xyloglucans that link and shroud cellulose microfibrils, however, the HvXET5 enzyme was not shown to have activity using arabinoxylans as a substrate. Hrmova et al. (2007) propose that the large size of the XET/XTH family allows for the possibility that other members of the family may have alternative substrate specificities that most likely include arabinoxylans. More recently, establishment of the role of XET/XTHs has been demonstrated *in vivo* (Van Sandt et al., 2007). The SkXTH1 from *Selaginella* was used to show that the XET/XTH family provide a cell wall loosening role in onion epidermis under tension.

With the knowledge that the endo-(1,4)-β-glucanases are co-transcribed with XET/XTHs and expansins during internode elongation, it is possible that if these three proteins are incubated together in barley cell wall preparations cell wall loosening will occur. An experiment was proposed to evaluate the effect that coexpressing these proteins *in vitro* with a barley cell wall preparation would have with regard to cell wall loosening. Analysis of the cell wall preparations after incubation with the hydrolytic enzymes and expansins may throw some light on the real role for these proteins during cell wall loosening. Purified barley endo-(1,4)-β-glucanase, XET/XTH and expansin enzymes were to be obtained using heterologous expression. Several factors need to be considered when producing protein heterologously, one of which is whether to use a prokaryotic or eukaryotic system for expression.

Prokaryotic organisms do not glycosylate proteins therefore successful expression of active plant enzymes that are normally glycolsylated during post-translational modification may be problematic. The prokaryotic system has been used successfully for expression of GH9 endo-(1,4)-β-glucanases. A GH9B endo-(1,4)-β-glucanase from the termite *Nasuitermes takasagoensis* was successfully heterologously expressed in *E.coli* BL21 (DE3) cells (Khademi et al., 2002). The expressed purified protein was used to determine its crystal structure. More recently, Urbanowicz et al. (2006) produced a fusion protein

comprising the catalytic domain of *Thermobifida fusca* TfCel6A and the tomato carbohydrate binding domain SlCel9C1, which was heterologously expressed in *E. coli* BL21 (DE3) cells. The protein was enzymically active and was subsequently used to determine the activity of the protein on a variety of substrates.

Although successful heterologous expression has been performed in prokaryotic systems, evidence suggests that glycosylation of some endo-(1,4)-β-glucanase enzymes is a requirement for activity. *In planta*, the molecular mass of the tomato GH9A endo-(1,4)-β-glucanase has been calculated as either 93 or 88 kDaltons (kD) compared to the predicted mass of 68.5 kD (Brummell et al., 1997) The increased molecular weight of the protein *in planta* is the result of glycosylation. Heterologous expression of a *Brassica napus* KORRIGAN orthologue was performed using *Pichia pastoris* as the host (Molhoj et al., 2001). The resultant expressed protein was heavily glycosylated and showed an apparent molecular mass of more than twice that predicted. Deglycosylation resulted in a protein with a molecular mass close to that predicted but with a 95% reduction in relative enzyme activity. The authors concluded that glycosylation is important for enzyme function. In a similar experiment, heterologous expression in *P. pastoris* was performed using a *Populus tremula x tremuloides* PttCel9A membrane bound endo-(1,4)-β-glucanase (Master et al., 2004). Again the expressed protein was highly glycosylated with up to 90% of relative activity destroyed by deglycosylation. By way of contradiction, heterologous expression of two pepper endo-(1,4)-β-glucanases in *Pichia* did not result in measurable glycosylation, as both proteins resolved in SDS PAGE at their predicted molecular mass (Ferrarese et al., 1998). So it would appear that a eukaryotic system may sometimes be a requirement for the expression of active endo-(1,4)-β-glucanase enzymes.

Other considerations relating to heterologous expression of endo-(1,4)-β-glucanases includes the conditions required for enzyme activity. Conserved sites

shown in the alignment in Figure 4-1 include a Ca^{++} ion binding site and those involved in ligand binding as well as a salt bridge. Khademi et al. (2002) indicated that a Ca^{++} ion binding site is situated in the substrate binding cleft in association with the catalytic site (Figure 4-2). This site is conserved in the endo-(1,4)-β-glucanase family as indicated by the alignment in Figure 4-1. The requirement of Ca^{++} ions for activity of plant endo-(1,4)-β-glucanases is further supported by Master et al. (2004) who showed that enzyme activity increased from zero to 80%, relative to known standards, when $CaCl_2$ concentration was increased from zero to 30 mM.

Another factor influencing enzyme activity is pH. The effect of pH 2.5 is to deform the Ca^{++} binding site and thus reduce Ca^{++} binding, which results in reduced activity (Khademi et al., 2002). At pH 6.5, Khademi et al. (2002) propose that the final conserved glutamate residue, as marked in Figure 4-1, which is the most distant conserved ligand binding site from the start site, is deprotonated. Deprotonation produces a rotation of the residue with subsequent loss of enzyme activity. It would appear from this and other evidence that peak activity of plant endo-(1,4)-β-glucanases occurs at between pH 4.5 and 6 (Molhoj et al., 2001; Master et al., 2004; Urbanowicz et al., 2007b).

Disulphide bonds are often formed to stabilise the protein structure. The formation of disulphide bonds does not appear to be conserved across the family or between species (Master et al., 2004). Master et al. (2004) propose two non-conserved stabilising disulphide bonds. The termite enzyme has a disulphide bond, also non-conserved (Khademi et al., 2002). Thus the requirement for disulphide bonds to stabilise the protein structure is not fully determined.

The work described in this Chapter includes the use of both prokaryotic and eukaryotic protein expression systems that were set up with the aim of producing active barley endo-(1,4)-β-glucanase enzyme. Consideration was given to many of the possible permutations that may be required for a positive outcome, however, attempts to express enzymically active heterologous protein were

largely unsuccessful. The Discussion section contains strategies that may be useful for future successful production of active protein.

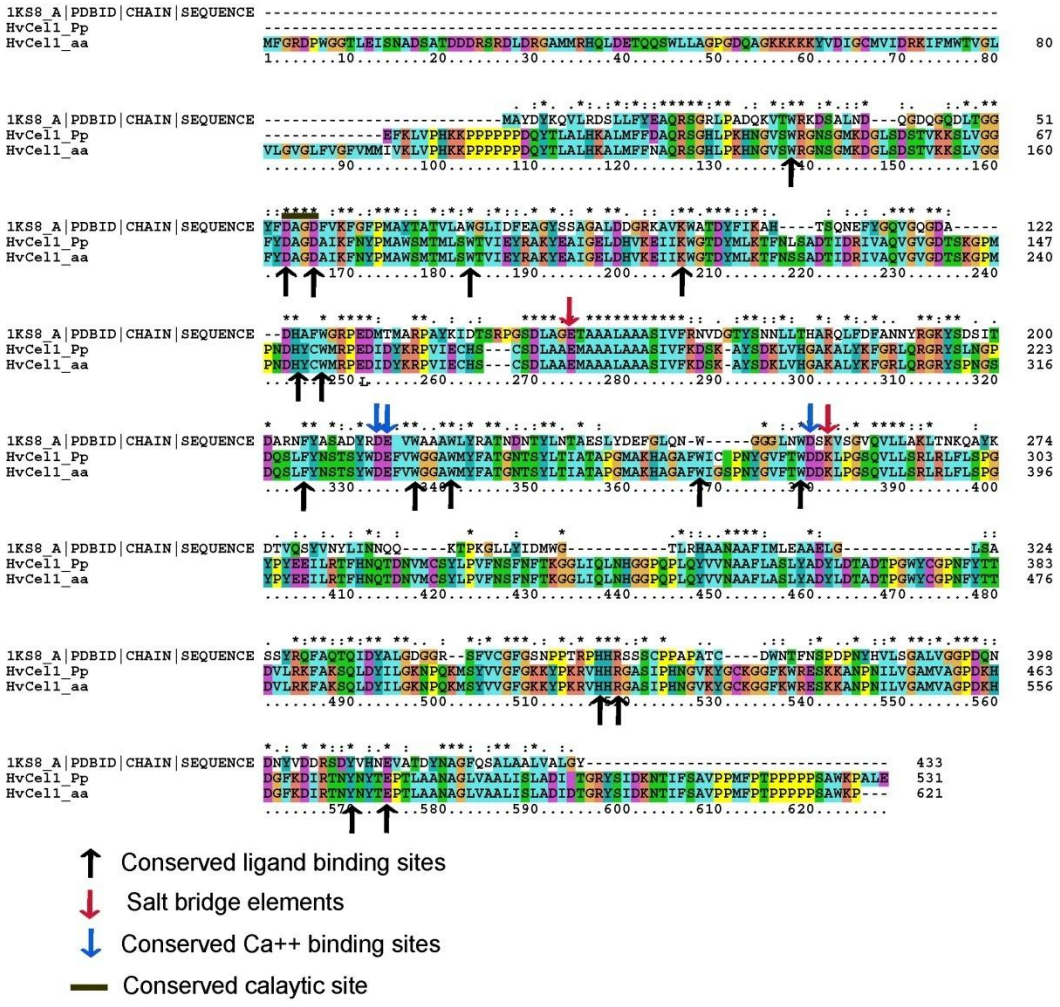


Figure 4-1: The amino acid alignment of the termite *Nasutitermes takasagoensis* and two barley HvCEL1 constructs.

The alignment shown above compares conserved sites of the termite *N. takasagoensis* (1KS8_A|PDBID|CHAIN|SEQUENCE) GH9 cellulase, the HvCEL1 protein without the transmembrane and cytosolic domains (HvCel1_Pp) and the wildtype HvCEL1 (HvCel1_aa). The conserved ligand and Ca^{++} binding residues (blue arrows), salt bridge elements (red arrows) and catalytic site (black bar) are indicated. Black arrows indicate conserved ligand binding sites. The Ca^{++} binding residues are located close to the catalytic site in the folded protein as illustrated in Figure 4-2 (Khademi et al., 2002; Master et al., 2004).

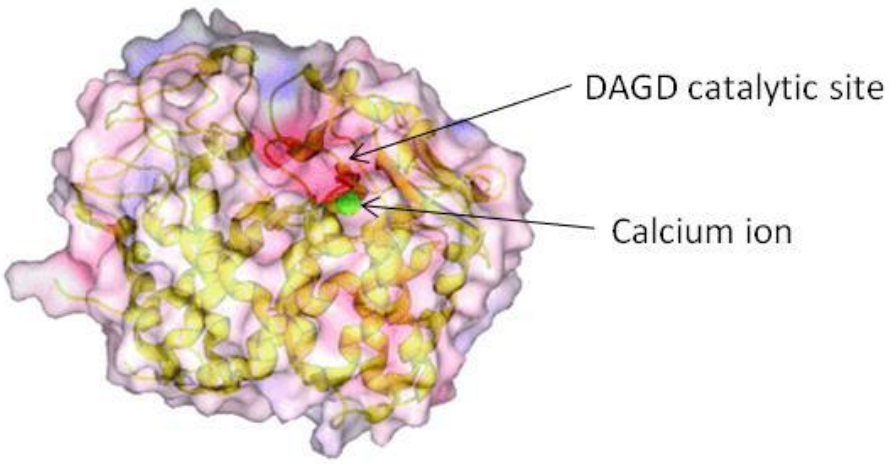


Figure 4-2: The termite *Nasutitermes takasagoensis* GH9 cellulase enzyme.

This drawing shows the location of the calcium ion (green) with respect to the catalytic site (red) as illustrated by Khademi et al. (2002).

See:
http://journals.iucr.org/

4.2. METHODS

4.2.1. Expression of endo-(1,4)-β-glucanase genes in *E. coli*

Expression of genes in *E. coli* was performed using two different vectors and expression induction methods as well as the use of several strains of *E. coli*. Expression in the bacterial vector system used here utilised the bacteriophage T7 mRNA polymerase, which produces much greater levels of transcript than the bacterial mRNA polymerase. Because the T7 mRNA polymerase is not present in prokaryotes, the host cells are modified by the addition of the gene. Induction of transcription of the T7 polymerase is facilitated by the addition of isopropyl-thio-galactopyranoside (IPTG), or lactose in the absence of glucose.

4.2.1.1. Expression vectors and E. coli strains

For expression in *E. coli*, all entry and expression vectors were "Gateway enabled". This system of cloning is briefly described in Chapter 5, sections 5.2.1.2, 5.2.1.3 and 5.2.1.4.

4.2.1.2. pET-DEST42

The bacterial expression vector pET-DEST42 (Invitrogen) is Gateway enabled and contains the lethal ccdB gene in the cloning region of the vector (Figure 4-3) (Invitrogen, 2002b). Also contained within the cloning region is a chloramphenicol resistance gene. The pET-DEST42 vector has an ampicillin resistance gene for the selection of transformed cells using 100 µg per ml ampicillin. Expression using the pET-DEST42 vector requires a ribosome binding motif AGGAGA and nine bases preceding the start codon. In these experiments the start codon was followed by six consecutive histidine codons, referred to as a 6xHis tag, before the coding region for the protein of interest. Therefore a stop codon was inserted before the 6xHis tag that is embedded at the 3' end of the coding region in the vector.

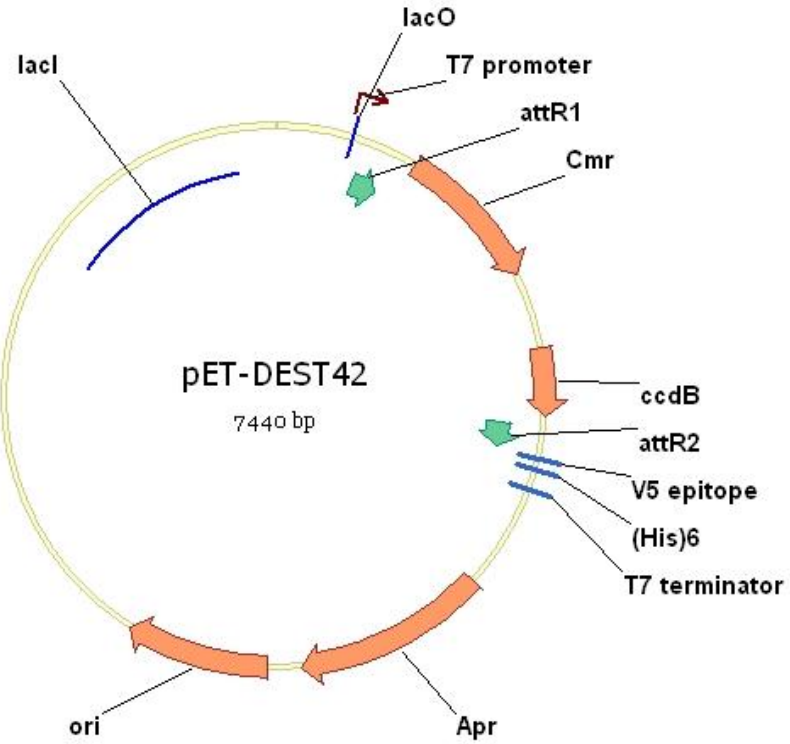


Figure 4-3: The pET-DEST42 expression vector.

The pET-DEST-42 vector contains a bacteriophage derived T7 promoter and terminator and Ampicillin (Apr) and chloramphenicol (Cmr) resistance genes. A 6xHis tag is situated at the 3' end of the ORF site before a stop codon. Note that the lethal ccdB gene is situated in the cloning site. The vector is Gateway enabled by provision of the attR1 and attR2 sites (Invitrogen, 2002a).

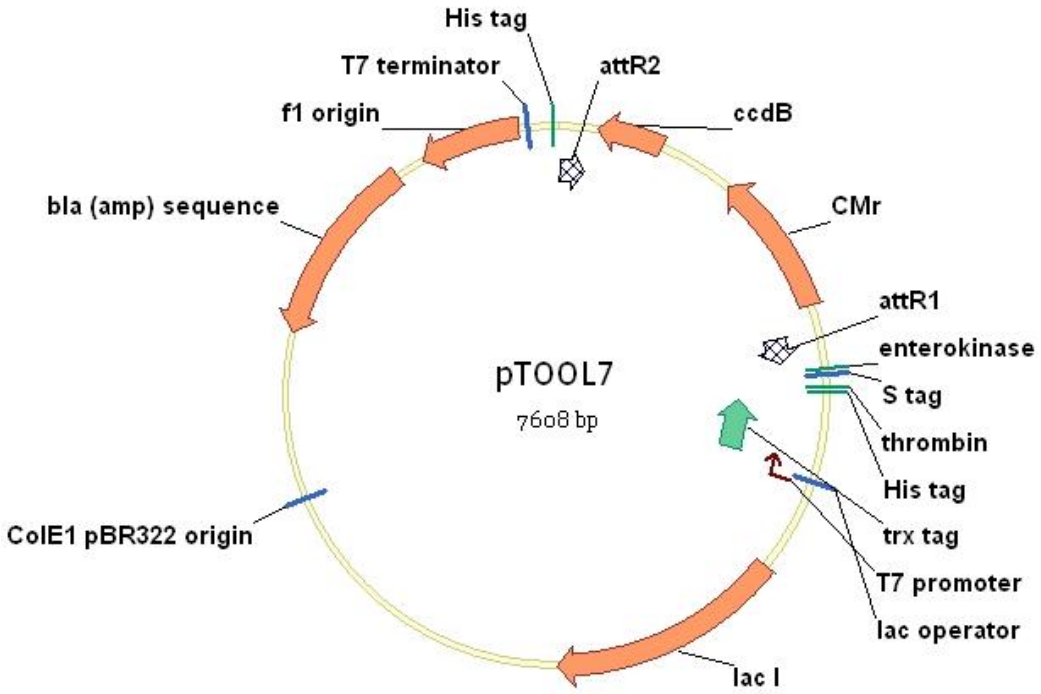


Figure 4-4: The pTOOL7 expression vector.

This vector is based on the pET32a (Novagen) vector. It has similar features to those associated with pET-DEST42 (0) but has been Gateway enabled in-house (Invitrogen). Other features include the trx tag that encodes thioredoxin as well as the protease digestion sites, thrombin and enterokinase . The target sequence is cloned between the attR1 and attR2 sites. A 6xHis tag is located at the 5' and 3' ends of the cloning site.

4.2.1.3. pTOOL7

The pTOOL7 bacterial expression vector is based on the pET32a vector (Figure 4-4) (Novagen, 2006). The pET32a vector has similar features to those associated with the pET-DEST42 vector but is not Gateway enabled. However, pTOOL7 has been modified such that it is functional within the Gateway system. Other features of pTOOL7, compared to pET-DEST42, include the thioredoxin Trx·Tag fusion tag, separate thrombin and enterokinase tags and two 6xHis tags. In theory, the Trx·Tag enables correct folding of the expressed protein and thus improved solubility in the cytoplasm. The thrombin and enterokinase tags allow for site specific protease cleavage of the target protein and the NH_2-terminal 6xHis tag from the Trx·Tag. Selection of transformed cells is enabled by an ampicillin resistance gene.

4.2.1.4. BL21 star

The BL21 star *E. coli* cell line carries the λ DE3 lysogen which contains the bacteriophage T7 RNA polymerase gene driven by the lacUV5 promoter (Invitrogen, 2002a). Transcription using T7 RNA polymerase is enabled by IPTG or lactose. A mutation at the C-terminal end of the RNase E enzyme reduces mRNA degradation without affecting rRNA and results in greater levels of expression of the target protein.

4.2.1.5. Origami (DE3)

As well as the benefits of the BL21 star cells, the Origami (DE3) strain of *E coli* cells provides increased levels of cytoplasmically-soluble protein by facilitating disulphide bond formation. Mutations in the thioreductase and glutathione reductase genes reduces misfolding of protein by allowing minimal reducing conditions in the cytoplasm and therefore increased disulphide bond formation. This plasmid is selectable using either 100 μg per ml kanamycin or 10 μg per ml tetracycline antibiotics.

4.2.1.6. Origami B (DE3)

This *E. coli* line is derived from the Origami (DE3) cells and is modified by the addition of the *lacZY* mutation that enables improved control of heterologous expression by providing sensitivity to varying concentrations of IPTG (Novagen). This is a useful addition if too much expression is preventing correct folding of the protein with consequent storage of the insoluble protein as inclusion bodies.

4.2.1.7. Amplification and cloning of gene products

Products for cloning were amplified using the Elongase polymerase (Invitrogen) as described in Chapter 5, section 5.2.1.1. The polymerase reaction mix contained 14 µl water with 2.5 µl DMSO, 2.5 µl Buffer A, 2.5 µl Buffer B, 1 µl each of 5 mM dNTPs, 4 µM forward and reverse primers, 1 µl template and 1 µl Elongase polymerase. Cycling conditions were similar to those used in Chapter 5, section 5.2.1.1. Primers used for amplification of the gene sequences are presented in Table 4-1. Amplified product was cleaned using the Macharey-Nagel Nucleospin kit as described in Chapter 2, section 2.2.5 prior to cloning into the pCR8/GW/TOPO entry vector (Invitrogen). If the product was more than 24 hours old it was A-tailed as described in Chapter 2, section 2.2.7.1. The construct was transformed into DH5α *E. coli* cells. Sequencing of the cloned insert was performed by AGRF and the insert was subsequently transformed into the appropriate expression vector via an LR reaction.

4.2.1.8. HvCEL1

The template used for amplification of HvCEL1 cDNA was the previously used pCR8/GW/TOPO-*HvCEL1* entry vector construct for transformation of barley as described in Chapter 5, sections 5.2.1.1, 5.2.1.2 and 5.2.1.3. The forward primer was designed to produce a cDNA fragment encoding a protein with a truncation of the NH_2-terminal so that the cytosolic and transmembrane domains were removed and the expressed protein contained only the catalytic domain (refer

Table 4-1: Primer sequences used for amplification of cDNAs.

The primers used to amplify product for cloning into pET-DEST42 and/or pTOOL7 for heterologous expression in *E.coli* are presented here with the length of the amplified product.

Product name	Forward Primer	Reverse primer	Product (bps)
HvCEL1	HvCel1_Ec_pCR8F-AGGAGACTATCTATCATGCACCATCACCATCACCACAAGCTCGTCCCGCACAAGAAGC	HvCel1_Ec_pCR8R-CTAAGGTTTCCAAGCTGACGGCG	1617
HvCEL3	HvCel3_dSP_F-GCGGTGGTGGCCATCATGAAGCG	HvCel3_stop_R2-CTAGGTGAATGTCTTTGCGACGGAC	1524
HvCEL11_FL	HvCEL11_dSP_F-GGCACCACCCCGAACTACCGCG	AL_TC_CK_QR-CACTTTATAGCGCAACAATACCATTC	1806
HvCEL11_CD	HvCEL11_dSP_F-AGGAGACTATCTATCATGCACCATCACCATCACCACGGGGGCGGTGGCACC	HvCEL11_Ec_CD_PCR8R-TCAGCCCATGAACGGCGCGTTGTTGTA	1443
HvCEL11_CBD	HvCEL11_Ec_CBMF-AGGAGACTATCTATCATGCACCATCACCATCACCACAGGTTTGATCGGTCAGGGCTCGAC	HvCEL11_Ec_pCR8R_TCAGACGAGCTTGTAGCCGGTGACC	408

Table 4-2: Predicted total length and weight of expressed proteins.

Products for *HvCEL1*, *HvCEL3* and *HvCEL11* were sequenced and inserted into several vectors. The table shows the vectors that the genes were cloned into, the number of amino acids in the expressed protein (aa) and the estimated weight in kDaltons calculated using the Vector NTI program (Invitrogen).

Product name	Vector	Predicted protein length (aa)	Predicted protein weight (kDaltons)
HvCEL1	pET-DEST42	534	59.5
HvCEL1	pTOOL7	614	68.5
HvCEL1	pPICαA	530	59
HvCEL3	pTOOL7	583	63.3
HvCEL3	pPICαB	537	58.3
HvCEL11_FL	pTOOL7	677	73
HvCEL11_CD	pTOOL7	556	60.5
HvCEL11_CBD	pTOOL7	211	23

to Chapter 1, Figure 1.15 for a diagram of the endo-(1,4)-β-glucanase sub-family structures). This truncation was introduced to increase the chance of producing soluble protein in the cytoplasm of the host. From the pCR8/GW/TOPO entry vector, *HvCEL1* was recombined into the expression vectors pET-DEST42 and pTOOL7 using Clonase II (Invitrogen) as described in Chapter 5, section 5.2.1.4. Predicted weight and length of the HvCEL1 expressed protein is presented in Table 4-2.

4.2.1.9. HvCEL3

The *HvCEL3* cDNA was amplified using a "cocktail" cDNA template as described in Chapter 2, section 2.3.1.2. This cDNA was derived from a mixture of vegetative tissues including caryopsis, developing embryo, stem and rachis. The translated protein of this gene has a signal peptide and catalytic domain. The forward primer was designed to truncate the first 72 bps from the wildtype ORF, thus eliminating the signal peptide to prevent excretion or sequestration of the protein (Table 4-1). After cloning into the pCR8/GW/TOPO entry vector the product was sequenced, and transferred into the pTOOL7 expression vector using Clonase II (Invitrogen) as described in Chapter 5, section 5.2.1.4.

4.2.1.10. HvCEL11

HvCEL11 comprises three separate domains; a signal peptide, a catalytic domain (CD) and a carbohydrate binding domain (CBD). Primers were designed to amplify three different sequences (Table 4-1) all of which lacked the signal peptide in the final product. The first, *HvCEL11_FL*, included the CD and CBD, the second, *HvCEL11_CD* included the CD only, and the final construct, *HvCEL11_CBD*, included only the CBD. The cDNA used for amplification of the sequences was derived from pistil. All constructs were cloned into the pCR8/GW/TOPO entry vector and sequenced before transferring into pTOOL7 using Clonase II (Invitrogen) as described in Chapter 5, section 5.2.1.4.

4.2.1.11. Expression of endo-(1,4)-β-glucanases in E. coli

Overnight cultures of 2 ml Luria Broth inoculated with cells containing the expression vector and the appropriate selective antibiotics were incubated in a 50 ml tube on a rotary shaker at 37°C. The following morning, the overnight culture was topped up with 18 ml Luria broth containing the appropriate antibiotics and placed on a rotary shaker at 37°C for one hour. IPTG was added to a concentration of between 0.05 and 1 mM to induce expression of the T7 mRNA polymerase. The culture was further incubated on a shaker at temperatures ranging from 17°C to 37°C for four to eight hours to allow sufficient quantities of protein to be produced.

Alternatively, the cultures were grown in autoinduction media. In this case, induction of expression occurred when all of the glucose, which inhibits T7 translation, had been metabolised by the bacteria. This left lactose as the main source of energy. Lactose induces translation of the T7 mRNA polymerase and hence induces translation of the target gene. Autoinduction media was sterilised by autoclaving and contained 2% 50X 5052 solution and 5% 20X NPS in Luria broth. A 500ml 50X 5052 solution was prepared with 125 gm glycerol, 12.5 gm glucose, 50 gm lactose and 365 ml sterile filtered water. The 20X NPS solution contained 66 gm NH_4SO_4, 136 gm KH_2PO_4 and 124 gm Na_2HPO_4 in 1 litre filtered sterilised water and adjusted to pH 6.75. Luria broth contained 10 gm tryptone and 5 gm yeast extract in 925 ml of filtered sterilised water and was autoclaved. Overnight cultures were grown in Luria broth with appropriate antibiotics. The culture was added to autoinduction media at the rate of 100 µl culture per ml media. This was placed on a rotary shaker at 25°C for 24 hours.

At the completion of the induction incubation, the tubes were centrifuged at 3220 x g for 10 minutes at 4°C. The supernatant was discarded and the pellet resuspended in 0.5 ml phosphate buffered saline to which the protease inhibitor phenylmethylsulphonyl fluoride (PMSF) was added at a concentration of 1 mM. The phosphate buffered saline contained 137 mM NaCl, 2.7 mM KCl, 4.3 mM

$NaHPO_4$ and 1.76 mM KH_2PO_4 and was adjusted to pH 7.4. The mixture was transferred to 1.5 ml microfuge tubes and placed on ice for 10 minutes. The cells were disrupted by sonication using the Sonifier B-12 (Branson Sonic Power Co.) at an intensity of 3 for 3 x 15 seconds. The sonicated solution was centrifuged at 13,000 x g for 10 minutes at 4°C and the supernatant placed in a second 1.5 ml microfuge tube. The supernatant was defined as the "soluble" fraction and the pellet as the "insoluble" fraction. All fractions were stored at -20°C prior to protein detection by Coomassie staining and Western blot.

4.2.2. Expression of endo-(1,4)-β-glucanases in Pichia pastoris

Heterologous protein expression was carried out using one of the pPICZα (Invitrogen) yeast expression vectors in the Protease A deficient SMD1168H strain of *Pichia pastoris* (Invitrogen).

4.2.2.1. pPICZα

The expression vector pPICZα (Figure 4-5) contains a methanol inducible promoter, alcohol oxidase 1 (AOX1), and an α-factor secretory signal, which directs the expressed target protein for secretion into the media (Invitrogen, 2008). The target sequence must be cloned in-frame with the secretory signal. The vector contains the Zeocin antibiotic resistance gene for selection in *E. coli* and *P. pastoris*. For purification or detection by western blot the vector contains a *c-myc* epitope and a 6xHis tag at the COOH-terminal end. To ensure expression of the recombinant fusion protein the cDNA must be cloned into the vector in frame with the secretory signal peptide at the NH_2-terminal and with the COOH-terminal tags. Three forms of the vector, A, B and C, are available and differ only at the multiple cloning site. These differences include frame shifts and a variation of the 5' restriction enzyme site to provide flexibility for cloning the target sequence into the vector.

Prior to preparing a sequence for cloning into the pPICZα vector, the gene of interest was searched for restriction enzyme sites similar to those in the vector multiple cloning site. This enabled selection of restriction sites that were not contained within the gene sequence. The gene sequence was also modified using specific primers. This was to ensure that the gene of interest was in frame with the α-factor secretory signal peptide at the NH_2-terminal end and the c-myc epitope and 6xHis tag at the COOH-terminal end of the sequence. Primers also were designed to incorporate the chosen restriction enzyme sites. Primers are presented in Table 4-3.

The genes were amplified using the Elongase polymerase as described in Chapter 2, section 2.2.1.3, using primers designed as described in the previous paragraph. As these cDNAs had already been cloned into pCR8/GW/TOPO entry vector, this was used as the template. Once the cDNA was amplified and the new sequence cloned into the pCR8/GW/TOPO entry vector, the construct was transformed into chemically competent DH5α *E. coli* cells by heat shock as described in Chapter 2, section 2.2.8.1. The alkaline lysis method, as described in Chapter 2, section 2.2.8.2, was performed to extract the construct from the cells for sequencing to ensure that the insert was correct and that the restriction enzyme sites were intact.

4.2.2.2. HvCEL1

HvCEL1 was cloned into the expression vector pPICZαA using the restriction enzyme *EcoR*I at the 5' end and *Xba*I at the 3' end of the insert. Once the sequence had been confirmed as correct, it was double digested from the pCR8/GW/TOPO vector using a mix including 0.5 µl each of *EcoR*I (NEB) and *Xba*I (NEB) restriction enzymes, 2.5 µl NEB *EcoR*I buffer, 1 µg total DNA and filtered sterile water to 25 µl. To prepare the expression vector for ligation of the digested *HvCEL1*, a similar digest using 1 µg pPICZαA was also incubated with the same enzymes. Both digests were gel purified on 1% agarose and the DNA

bands cut out of the gel and cleaned on Macherey-Nagel Nucleospin columns as outlined in Chapter 2, section 2.2.5 in preparation for the ligation reaction.

Ligation of the digested gene product into the pPICZαA vector was facilitated with 0.5 µl T4 ligase (Invitrogen), 2 µl 5X ligase buffer (Invitrogen) and 0.5 µl filtered sterile water. DNA for the ligation reaction of vector:insert was at an approximate ratio of 3:1, with 66 ng pPICZαA and 22.4 ng *HvCEL1* insert added to the mixture. The ligation mix was incubated overnight at 15°C and subsequently transformed into One Shot Top Ten Chemically Competent *E. coli* cells (Invitrogen) by the heat shock method as described in Chapter 2, section 2.2.8.1. The cells were plated onto low salt LB agar containing an autoclaved solution of 1% (w/v) Tryptone, 0.5% (w/v) NaCl, 0.5% (w/v) yeast extract and1.5% (w/v) agar, adjusted to pH 7.5, and Zeocin was added at a rate of 25 µg per ml for selection of the successfully transformed cells. Transformed cells were reincubated in a liquid culture of autoclaved low salt LB media containing 1% (w/v) tryptone, 0.5% (w/v) yeast extract and 0.5% (w/v) NaCl, adjusted to pH 7.5, and Zeocin was added at a concentration of 25 µg per ml. After overnight incubation at 37°C, the plasmids were extracted by alkaline lysis as described in Chapter 2, section 2.2.8.2 and the insert sequenced across the restriction enzyme ligation sites to ensure that the insert was correctly ligated.

4.2.2.3. HvCEL3

HvCEL3 was amplified and cloned similarly to *HvCEL1* for expression in *P pastoris* but with some differences. The pPICZαB vector was employed using the *Pml*I restriction enzyme site at the 5' end and *Not*I at the 3' end. This required a forward primer with the *Pml*I restriction enzyme site and a reverse primer with the *Not*I site (Table 4-3). Once the insert had been confirmed as correct it was excised from the pCR8/GW/TOPO entry vector and at the same time the pPICZαB vector was digested as per *HvCEL1*, however, with some

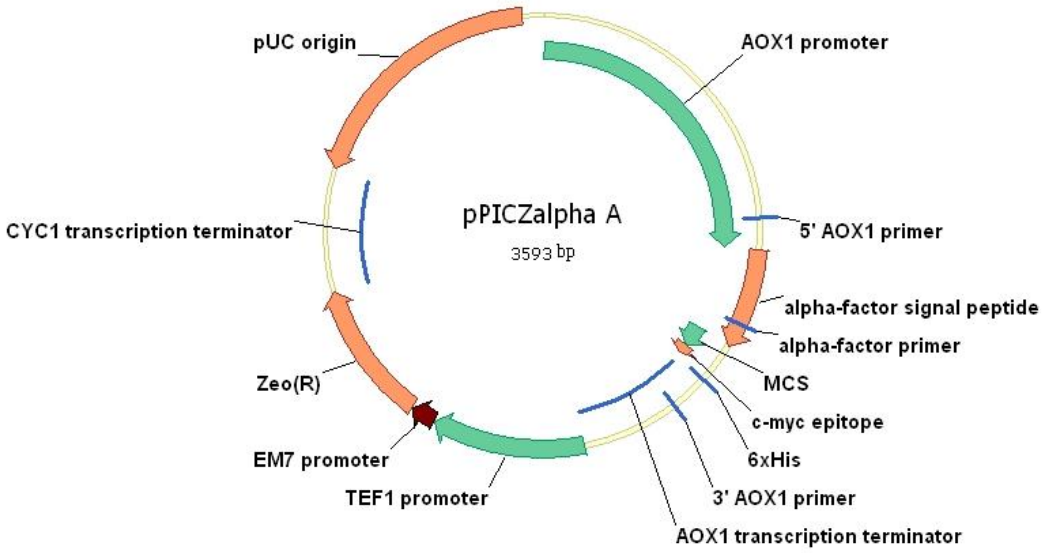


Figure 4-5: The pPICZαA eukaryote expression vector

This vector contains prokaryotic and eukaryotic promoters for Zeocin transcription (Invitrogen, 2008). The AOX1 methanol induced promoter ensures translation in the presence of methanol and the α-factor signal peptide directs the protein out of the cell. Both the 6xHis tag and c-myc epitope enable purification as well as antibody recognition for western blot analysis.

Table 4-3: Primers used to amplify sequences for cloning into pPICZα

These primers contained the restriction enzyme site required for the cloning reaction. 5' restriction enzyme sites are indicated in red and 3' restriction enzyme sites are written in blue. For HvCEL1, the *EcoR*I restriction site was used for the 5' end of the sequence. This restriction enzyme site is also contained within the pCR8/GW/TOPO cloning site so a small linker sequence, as written in green, was inserted 5' to the *EcoR*I site of the primer to facilitate cleavage.

Gene name	Forward primer	Reverse primer	Length (bps)
HvCEL1	HvCel1_Pp_pCR8F-CTATACTAGAATTCAAGCTCGTCCCGCACAAGAAG	HvCel1_Pp_pCR8R-TCTAGAGCAGGTTTCCAAGCTGACGGC	1601
HvCEL3	HvCel3_Pml1_F-CACGTGGCGGTGGTGGCCATCATGAAGCG	HvCel3_Not1_R-GCGGCCGCGGTGAATGTCTTTGCGACGGA	1535

Table 4-4: Digest rates for NEB and Super Duper buffers.

The restriction enzymes *Pml*I and *Not*I were deemed incompatible for the purpose of a double digest by NEB so three different buffers were tried in order to overcome this problem as described in detail in the text. This table indicates the digestion rates expected for the two enzymes using three different buffers (http://www.neb.com/nebecomm/DoubleDigestCalculator.asp?). No digestion rate is published for the Super Duper buffer, which is a proprietary product.

NOTE:
This figure/table/image has been removed to comply with copyright regulations.
It is included in the print copy of the thesis held by the University of Adelaide Library.

differences. Using the NEB restriction enzyme double digest calculator (http://www.neb.com/nebecomm/DoubleDigestCalculator.asp?) as a guide, it was not advised to perform a double digest between *Pml*I (NEB) and *Not*I (NEB) and the suggestion was to carry out a single digest for each restriction enzyme. The problem was one of buffer/restriction enzyme incompatibility. To increase the chance of success, three separate digests were performed in which1 µl each of *Pml*I and/or *Not*I restriction enzyme was used with 3 µl of three different buffers (Table 4-4) and 3 µl DNA made up to 30 µl with filtered sterile water. The most successful digest was that using NEB buffer 2.

The *Pml*I restriction enzyme cuts with blunt ends, which reduces the rate of successful ligation. Ligation into the pPICZαB vector was performed as per *HvCEL1* using a higher ratio of vector to insert and a longer incubation time. Here, 38 ng *HvCEL3* and 75 ng pPICZαB at an approximate 2:1 vector to insert ratio was used, and the incubation was performed for 24 hours. Following ligation, transformation into One Shot Top Ten Chemically Competent *E. coli* (Invitrogen) was performed.

4.2.2.4. Transformation of P. pastoris strain SMD1168H by electroporation

Ideally, 5 to 10 mg of pPICZα containing the insert is required for transformation of *P. pastoris*. This is acquired by incubating the previously transformed *E. coli* cells in autoclaved low salt liquid LB containing 1% (w/v) Tryptone, 0.5% (w/v) yeast extract and 0.5% (w/v) NaCl, adjusted to pH 7.5 with Zeocin added to a concentration of 25 µg per ml. Incubation was performed at 37°C with shaking and the plasmid extracted using the Qiagen miniprep kit as described in Chapter 5, section 5.2.1.4. The quantity of vector with insert was measured using the NanoDrop (NanoDrop Technologies). The insert was linearised with the *Pme*I restriction enzyme by incubating 25 µl plasmid, 0.2 µl enzyme, 3 µl NEB Buffer 4, 0.25 µl bovine serum albumin and 1.55 µl sterile filtered water at 37°C for three hours. To ensure that complete digestion of the plasmid had occurred, a 3

µl aliquot was size separated on a 1% agarose gel and visualised under UV light as described in Chapter 2, section 2.2.4.

The SMD1168H *P. pastoris* cells were prepared for electoporation by firstly incubating an aliquot of cells overnight at 30°C in 5 ml YPD media containing an autoclaved solution of 1% (w/v) yeast extract and 2% (w/v) peptone to which was added 2% (w/v) dextrose filtered using Millex-GV polyvinylidene fluoride (PVDF) 0.22 µm pore filters (Millipore). A 2 ml aliquot of this culture was added to 250 ml YPD and grown until the cells had ideally reached an optical density (OD) of 1.3 to 1.5 measured at 600nm. After 6 hours of incubation, the cells had reached an OD of only 0.8 and despite there being fewer cells than recommended they were prepared for electroporation.

The Beckman Coulter Avanti J-E centrifuge, rotor size 16.250 was cooled to 4°C, the cells were centrifuged at 5,000 revolutions per minute (rpm) for 10 minutes and the supernatant removed. The cells were carefully resuspended in 250 ml of water cooled to 4°C and placed on ice. The cells were recentrifuged and resuspended a second time in cold water. The centrifugation was repeated and the pellet resuspended in 50ml of cold 1M sorbitol and the solution recentrifuged. The cells were resuspended in approximately 1 ml sorbitol remaining in the flask after removal of the majority of the supernatant.

Prior to electroporation, all solutions and cuvettes were placed on ice for at least one hour. For electroporation, 80 µl of the *Pichia* cell suspension was added to 5 µl of linearised pPICZα containing the target sequence. This mixture was placed in the Gene Pulsar Cuvette with a 0.1 cm electrode gap (BioRad Laboratories) and checked to ensure that the suspension was at the bottom of the cuvette. The cuvette was placed in the Gene Pulsar Xcell PC module electoporator (BioRad Laboratories) and pulsed at 1,500 Volts, 400 ohms and a capacitance of 25 µFarad. Immediately following electroporation, 500 µl of cold sorbitol was added to the cuvette and the contents placed in a 50 ml tube. The cells were incubated without shaking for three hours.

Sterile YPDS agar plates had been prepared previously and contained 1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) dextrose, 1% (w/v) agar and 1M sorbitol. Zeocin for selection of transformants was added at a concentration of 100 μg per ml. The electroporated cells were added to the plates at two volumes. 50 μl was spread on one plate and 300 μl was spread on a second plate. The plates were incubated at 30°C for two to three days until colonies had grown to approximately 1 mm in diameter.

4.2.2.5. Heterologous expression in P. pastoris

Buffered glycerol complex medium (BMGY) and buffered methanol complex medium (BMMY) solutions were used for protein expression in SMD1168H *P. pastoris*. These solutions are similar except that glycerol is contained in BMGY and for BMMY, methanol replaces the glycerol. The solutions were prepared as follows. A mixture of 10 g yeast extract and 20 g peptone in 700 ml filtered water was autoclaved. Phosphate buffer containing 132 ml 1M K_2HPO_4 and 868 ml 1M KH_2PO_4 adjusted to pH 6 with KOH was autoclaved as was a 10% (v/v) solution of glycerol. A 10X Yeast Nitrogen Base with ammonium sulphate (YNB) solution was prepared with 13.4 g YNB in 100 ml filtered sterile water and was filter sterilised with Millex-GV polyvinylidene fluoride (PVDF) 0.22 μm pore filters (Millipore). A 5% (10X) solution of methanol was prepared and similarly filter sterilised. Biotin was prepared as a 500X solution containing 20 mg of biotin in 100 ml filtered sterile water and filter sterilised using a Millipore 0.22 μm PVDF filter. The final solution contained the autoclaved 700 ml solution with yeast extract and peptone to which was added 100 ml 10X phosphate buffer, 100 ml 10X YNB and 2 ml 500X Biotin. For BMGY, 100 ml of 10% (10X) glycerol was added whereas for BMMY, 100 ml 5% methanol was added. These solutions were stored at 4°C until required.

4.2.2.6. Protein expression in P. pastoris

Initially, the transformed colonies produced after Zeocin selection were screened

for those with the highest levels of protein expression. The colonies were picked from the YPDS plates, streaked onto fresh YPDS plates with 100 μg per ml Zeocin and also cultured in BMGY media at 30°C for 24 hours. The incubation was performed with 5 ml or 2.5 ml BMGY in 50 ml or 10 ml tubes, respectively. After 24 hours, the cultures were centrifuged at 2000 rpm for five minutes and the BMGY replaced with a similar volume of BMMY media. The cultures were incubated on a shaker at room temperature. Methanol, at a rate of 0.5% (v/v) was added daily to the cultures and samples were collected from the tubes over a six day period. Periodic sampling was performed to determine the point at which the greatest protein expression with least protein degradation had occurred. Collected samples were centrifuged at 4000 x g for 10 minutes. The supernatant was collected and frozen.

After the initial screening, clones with the greatest levels of protein expression were incubated in volumes of 250 ml or 500 ml BMGY in one or two litre conical flasks, respectively. These larger volume incubations were treated similarly to the screening cultures with replacement of BMGY with BMMY after 24 hours and daily additions of 0.5% (v/v) methanol. Cultures were incubated for up to 6 days at which point the tubes were centrifuged in the Beckman Coulter Avanti J-E centrifuge, rotor size 16.250, at 5000 rpm for 30 minutes at 4°C before collection of the supernatant for further analysis.

4.2.2.7. Sample analysis by dot blot

Dot blot screening was performed on all of the supernatants of the 2.5 and 5 ml incubations to determine which were producing the heterologously expressed protein. A 5 μl aliquot from each culture was applied to a Nitrobind Cast, 0.22 micron, Pure Nitrocellulose membrane (GE Water and Process Technologies) that had been soaked in Tris-buffered saline (TBS) containing 50 mM Tris-HCl and 150 mM NaCl. The dot blotted membrane was initially blocked with 1% (w/v) skim milk powder in TBS for one hour. Incubations with the 6xHis

monoclonal mouse antibody and the conjugated anti-mouse antibody and the TBS washes were performed as described in section 4.2.5.

4.2.2.8. Protein purification

Trichloracetic acid precipitation

Concentration of the protein in the supernatant was initially achieved by precipitating 300 µl supernatant with 60 µl 50% (w/v) solution of trichloracetic acid. The solution was incubated at 4°C for 15 mins on a rotor and centrifuged at 13,000 x g for 15 mins. The supernatant was discarded, the pellet washed with acetone, the sample recentrifuged and finally dried. Reconstitution of the pellet was performed using 25 µl filtered sterile water and the samples prepared for western blot analysis by the addition of 5 µl 6X sample buffer.

Ammonium sulphate precipitation

For the ammonium nitrate precipitation, the supernatant was precipitated using a 70% (w/v) saturated solution comprising 472 g of $(NH_4)_2SO_4$ per litre supernatant. The solution was incubated with a mechanical stirrer for three hours at 4°C and centrifuged for 30 mins on a Beckman Coulter Avanti J-E centrifuge, rotor size 16.250 at 10,000 x g at 4°C. The supernatant was removed and the pellet dissolved in 2.5 ml phosphate buffered saline. The phosphate buffer comprised 50 mM PO_4^{3-} and 300 mM NaCl containing 0.47 g NaH_2PO_4, 12.5 g Na_2HPO_4 and 17.5 g NaCl dissolved in a litre of filtered sterile water adjusted to pH 8.

Desalting of precipitated protein

Desalting of the ammonium precipitated protein was facilitated with a PD-10 Sephadex size separation column. The principle behind size separation is that the smaller salt particles move slowly through the matrix as they are more easily enmeshed within it. The larger particles, or proteins in this case, flow through the column more rapidly and are therefore first to elute. A 90% to 95% protein

recovery rate is expected. The column was prepared by washing with 25 ml phosphate buffer and the flow-through discarded. A 2.5 ml aliquot of the sample was applied to the column and the flow-through again discarded. The sample was eluted with 3.5 ml phosphate buffer and the flow-through containing the protein was collected. The column was regenerated with 100 ml filtered sterile water.

Ni-NTA purification of desalted protein

Nickel-nitrilotriacetic acid (Ni-NTA) agarose (Qiagen, 2008) takes advantage of its affinity for histidine to purify proteins carrying a 6xHis tag. All steps of the Ni-NTA purification were performed at 4°C. The Ni-NTA agarose was washed in phosphate buffer by the addition of 1 ml buffer to 300 µl Ni-NTA agarose. The mixture was briefly centrifuged and the supernatant discarded. The desalted protein sample was added to the resin, incubated on a rotor at 4°C for two hours and centrifuged. The supernatant was discarded and the Ni-NTA/sample washed twice in 1 ml 10 mM imidazole to remove non-specifically bound proteins. The Ni-NTA/sample was applied to a Qiagen Ni-NTA spin column (Qiagen) washed once more with 10 mM imidazole, centrifuged briefly and the flow-through discarded. Protein bound to the Ni-NTA was eluted by washing the column four times with 0.25 ml 250 mM imidazole and the flow-through captured by brief centrifugation. With each addition of 250 mM imidazole, the column was incubated for one minute. The eluate was stored at 4°C or -20°C for further analysis.

4.2.3. SDS PAGE

All of the samples subjected to sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE) were firstly denatured as follows. A 6X sample buffer solution containing 3.5 ml stacking buffer (0.5 M Tris-HCl and 0.4% (w/v) SDS pH 6.8, 1.5 ml glycerol, 0.5 gm SDS, 0.47 gm dithiothreitol (DTT) and a few grains bromphenol blue to achieve a dark blue colouration in the

solution was prepared. The soluble fraction samples were prepared by boiling for five minutes in sample buffer at a ratio of one part sample to three parts 6X sample buffer. The insoluble fraction was first resuspended in 300 µl stacking buffer to which 100 µl 6X sample buffer was added and the tubes boiled for 5 mins. Following boiling, the tubes were centrifuged for five minutes at 13,000 x g.

Poly-acrylamide denaturing gels were prepared in preformed 1.5 mm NC2015 cassettes (Invitrogen). The 10% resolving gel with a volume of 10 ml consisted of 2.5 ml 40% (w/v) acrylamide/Bis solution in a ratio of 29:1 acrylamide:bis-acrylamide (BioRad), 2.5 ml resolving buffer comprising 0.5 M Tris-HCl and 0.4% (w/v) SDS at pH 8.8 and 5 ml filtered water, with 70 µl 30% (w/v) ammonium persulphate (APS), and 10 µl N,N,N',N'-Tetramethylethylenediamine (TEMED) added last and immediately before pouring the gel. The resolving gel solution was carefully poured to a level approximately 2.5 cm below the top of the cassette and immediately covered with filtered water to prevent drying out, exclude air and facilitate setting. The stacking gel was prepared as for the resolving gel using 1 ml 40% (w/v) Acrylamide/Bis solution (BioRad Laboratories, California, USA), 2.5 ml stacking buffer, which is as described for the resolving buffer but at pH 6.8 and 6.5 ml filtered water with 70 µl 30% (w/v) APS and 10 µl TEMED added immediately before pouring. Once the resolving gel had set for approximately 30 minutes, the water was removed, the space above the gel blotted dry with Whatman filter paper and the stacking gel poured into the top of the cassette. Immediately, 1.5 mm combs with either 10 or 15 teeth were carefully placed in the stacking gel to ensure that no bubbles were trapped in the gel. The cassette and its contents were left for approximately one hour to cure.

Once the gels had set they were assembled into the XCell Surelock Electrophoresis Cell (BioRad) according to the manufacturer's instructions. Two different buffers were variously used for running the gel. These included Tris-

glycine buffer containing 25 mM Trizma base, 192 mM glycine and 0.1% (w/v) SDS and MES buffer containing 2.25 mM 2- (morpholino)ethanesulfonic acid, 2.25 mM Trizma base, 0.05% (w/v) SDS and 0.05 mM ethylenediaminetetraacetic acid (EDTA). Between 15 and 30 µl of sample was loaded into the wells of the gels. One well in each gel also contained a 5 µl aliquot of the molecular marker, SeeBlue Plus2 Prestained Standard (Invitrogen) and variously, BenchMark His-tagged Protein Standard (Invitrogen). Size separation of the proteins was performed at 120 volts and was halted when the bromphenol blue dye line was within 1 mm of the bottom of the gel. The gel was removed from the cassette and either stained with Coomassie blue or a western blot was performed.

4.2.4. Staining gels with Coomassie Blue

The Coomassie stain binds to histidine, arginine and the aromatic amino acids. For staining, the gel was placed at room temperature in a covered container with Coomassie stain solution consisting of 0.1% (w/v) Coomassie blue R350, 30% (v/v) ethanol, and 10% (v/v) acetic acid for 2-3 hr. Once the gel had stained an homogeneous dark blue it was destained for approximately 12 to 15 hours at room temperature on a rocker, or until all background stain had disappeared, leaving the stained protein bands. The destaining solution consisted of 30% (v/v) ethanol and 10% (v/v) acetic acid and was replaced three to four times during the destaining process.

4.2.5. Western Blot analysis

Except for the transfer step, all incubations and washes involved in the protein immunoblot, or western blot analysis, were performed at room temperature. The protein on the the poly-acrylimide gel was transferred to a Nitrobind Cast, 0.22 micron, Pure Nitrocellulose membrane (GE Water and Process Technologies). Transfer to the membrane was performed using the 422 Electro-Eluter (BioRad Technologies) at a constant 300 milliamps for 90 minutes in Tris-glycine buffer

containing 25 mM Trizma base and 192 mM glycine. A transfer "sandwich" was prepared as presented in Figure 4-6. Firstly, the nitrocellulose membrane and two pieces of Whatman paper, the same dimensions as the gel, and sponges were prepared by soaking in Tris-glycine buffer before applying to the sandwich. One sponge was placed on the negative electrode (black) of the transfer cassette and Whatman paper placed over it. The gel was carefully placed back to front on the Whatman paper and rolled to remove all bubbles from between the gel and the Whatman paper. The membrane was carefully placed on top of the gel and again rolled to remove the bubbles. A further piece of Whatman paper was added and the bubbles removed before placing the final sponge over the Whatman paper. The cassette was closed and secured before placing in the transfer tank containing cold Tris-glycine buffer with ice blocks. Alternatively, the tank filled with cold buffer was placed in an ice bath, after securing the lid, and protein transfer was commenced.

Following transfer, the membrane was initially "blocked" for approximately one hour. This was done using 5% skim milk powder in Tris buffered saline (TBS) containing 50 mM Tris-HCl and 150 mM NaCl, to reduce background binding of the antibodies to the membrane. The membrane was washed once with TBS and incubated overnight with 1 µl 0.5 mg per ml 6xHis monoclonal mouse antibody (Clontech) in 5 ml 1% skim milk powder in TBS. The membrane was given 3 x 5 mins washes in TBS and incubated for 2 hours in 1:30,000 alkaline phosphatase conjugated goat anti-mouse antibody in 20 ml 1% skim milk powder in TBS. A further wash of 3 x 5 mins with TBS was performed. Finally the nitrocellulose membrane was incubated in 1 ml Western Blue Stabilised Substrate for Alkaline Phosphatase (Promega). The reaction was stopped by rinsing with water and the membrane allowed to dry between two sheets of Whatman paper.

4.2.6. Viscometric cellulase assay

Use of viscometric analysis to assay endo-(1,4)-β-glucanase activity is based on two premises. The first is that endo-(1,4)-β-glucanase digestion of the cellulose substrate reduces the viscosity of the substrate solution. Secondly, a known volume of a liquid of a particular viscosity will pass more quickly through a capillary tube than one of higher viscosity. The time taken for the liquid to pass through the capillary tube is measured and converted to a relative measure of enzyme activity. Activity of the expressed endo-(1,4)-β-glucanase protein was measured by viscometric analysis using the CMC-4M cellulose (Megazyme) substrate. CMC-4M cellulose is a carboxymethyl substituted cellulose preparation with approximately four carboxymethyl substitutions per 10 glucose units. The cellulose was prepared as a 1% solution by heating 1 gm of CMC-4M with 90 ml of sterile filtered water at 90°C for one hour. After the solution had cooled to room temperature, 5 ml 2 M sodium acetate buffer at pH 5.0 was added with a few drops of toluene as an antimicrobial agent and the volume made up to 100 ml using filtered sterile water. The analysis was performed using an Otswald viscometer (Figure 4-7).

Analysis involved measuring the reduction in viscosity of the CMC-4M cellulose substrate that occurs when subjected to digestion by the expressed endo-(1,4)-β-glucanase. The assay was performed at 37°C in a water bath. The substrate was diluted to 0.25% using sodium acetate buffer at pH 5 and allowed to stand in the water bath for at least two hours to equilibrate. The viscometer was also placed in the water bath and allowed to equilibrate before adding 500 µl of the prepared substrate. To this was added calcium chloride to a concentration of 30 mM and 100 µl purified protein. The solution was mixed by gently blowing air through the viscometer after which the fluid level was raised above the upper mark by suction. The time taken for the bottom of the meniscus to pass between the upper mark and the lower mark was measured.

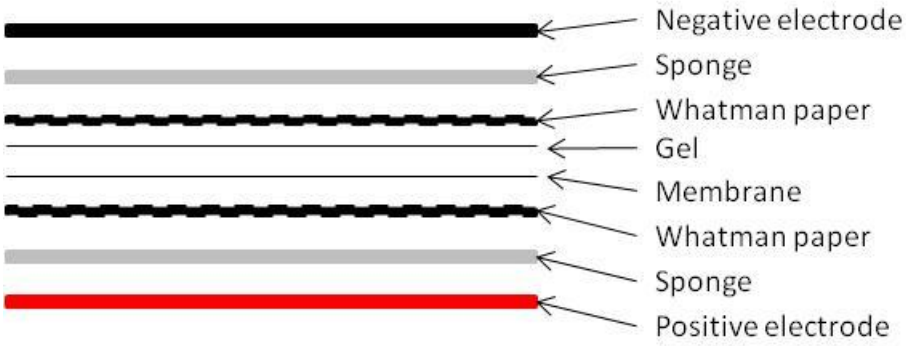


Figure 4-6: The “transfer sandwich”.

The “sandwich” used to transfer proteins from the gel to the nitrocellulose membrane consisted of sponges used as padding and Whatman paper, situated on each side of the gel and membrane

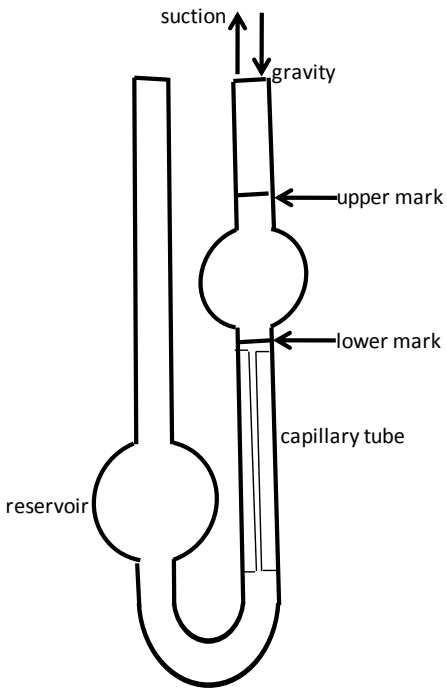


Figure 4-7: Ostwald viscometer used to measure relative viscosity of liquids.

The liquid is sucked from the reservoir on the left so the upper level of fluid is above the upper mark. The liquid is allowed to flow under the influence of gravity from the upper mark to the lower mark and the time taken for this to occur is measured. More viscous liquids will flow more slowly through the capillary tube due to their higher surface tension.

4.3. RESULTS

4.3.1. Heterologous expression in *E. coli*

4.3.1.1. Expression in the pET-DEST42 expression vector

Heterologous expression was performed in a 20 ml culture of *E. coli* BL21 star cells carrying *HvCEL1* cloned into the pET-DEST42 expression vector. Induction using 1 mM IPTG was performed on three different clones using cells transformed with empty pGEM-T vector as a control. Cells transformed with the empty pET-DEST42 vector cannot be used as a control as the ccdB gene contained within the cloning site of the vector is lethal to these *E. coli* cells. The cells were induced for 4 hours at 37°C and processed for SDS PAGE. The SDS PAGE was performed using Tris-glycine buffer and the gel stained with Coomassie stain. The results indicated that only insoluble protein was produced. The incubation was repeated with the induction carried out at room temperature for six, 12 and 18 hours. The results are presented in Figure 4-8. Coomassie staining of the gel and western blot analysis were performed. The 6xHis antibody did not attach to protein from the pGEM-T control cells, indicating that antibody only bound to the expressed HvCEL1 protein. There appeared to be only small differences between the results of the three treatments. Although a large amount of protein was produced, as indicated by size evaluation and 6xHis antibody binding, it was determined that the expressed HvCEL1 protein was largely insoluble.

4.3.1.2. Expression in the pTOOL7 vector

Five constructs in total were cloned into the pTOOL7 expression vector and Origami (D23) cells were transformed with these clones. This vector contains the recombinant fusion thioredoxin tag which aids folding of the expressed protein. The Origami (DE3) cells contain mutations that enable the formation of disulphide bridges. It is not established if such bridges exist in the barley endo-(1,4)-β-glucanases. However, the structure of the termite endo-(1,4)-β-glucanase that shares most of the conserved sites with other members of the GH9 family

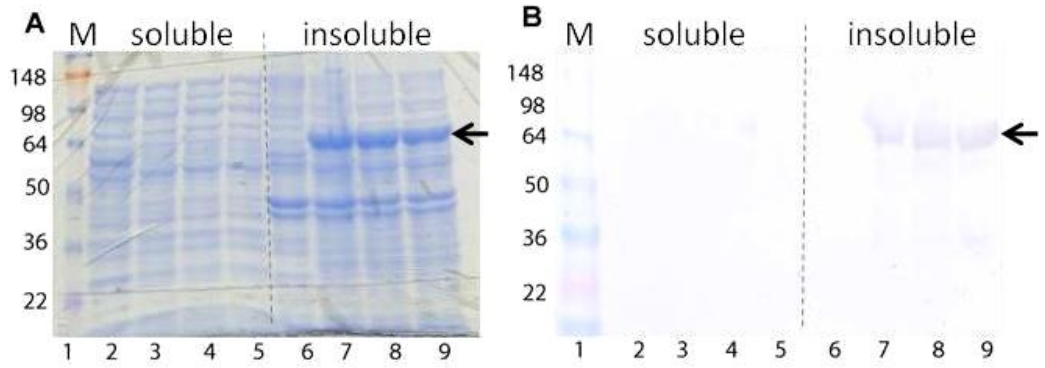


Figure 4-8: Coomassie stained gel and western blot for HvCEL1 in the pET-DEST42.

Expression was performed in BL21 star *E.coli* cells at 20°C for 6, 12 or 18 hours. The resultant crude extracts were subjected to (A) Coomassie staining of SDS PAGE and (B) western blot using the 6xHis tag antibody. The predicted size of the expressed protein is 61kD. For both A and B, Lane 1: SeeBlue2 Marker, Lane 2 and 6: soluble and insoluble fraction from empty pGEM-T vector, Lane 3, 4 and 5: soluble fraction 6, 12 and 18 hours of induction HvCEL1/pET-DEST42, Lane 7, 8 and 9: insoluble fraction 6, 12 and 18 hours of induction HvCEL1/pET-DEST42. The numbers along the left hand side indicate molecular weights of the SeeBlue Plus 2 (Invitrogen) marker protein bands in kDaltons using Tris-glycine buffer during electrophoresis. The arrows show the expected position of the expressed protein. A negative band on the western blot for the pGEM-T protein indicates that the 6xHis antibody is only binding to the expressed HvCEL1 protein.

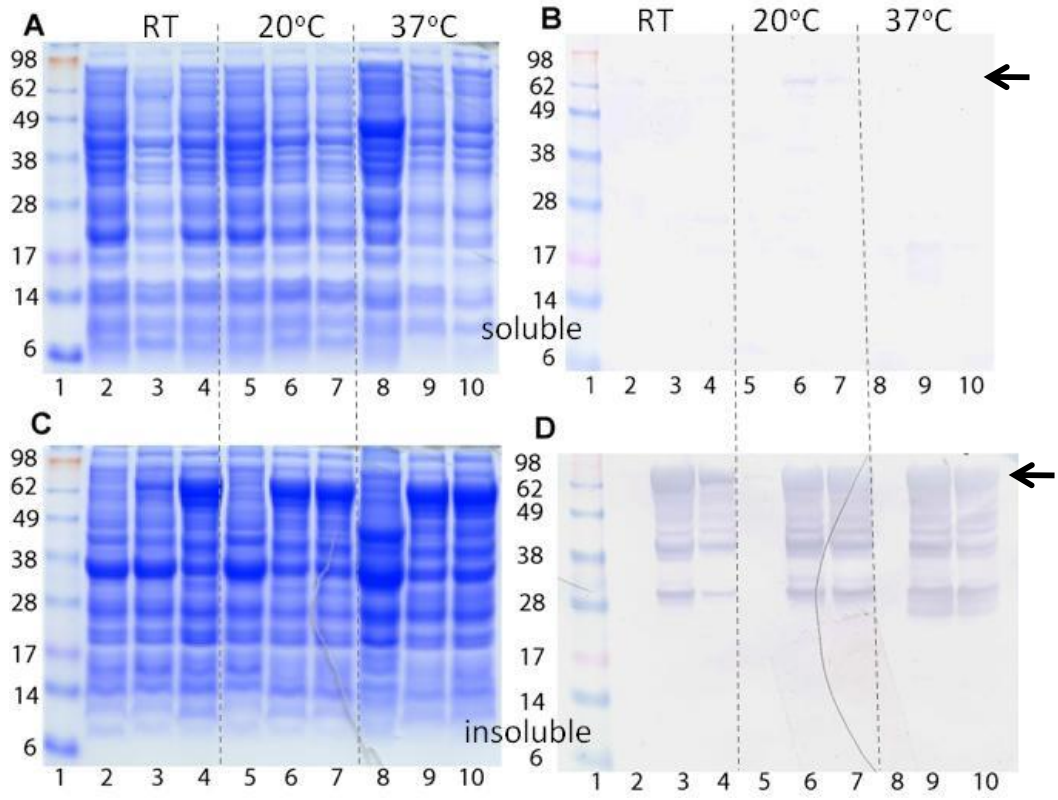


Figure 4-9: Coomassie stained gels and western blot analysis of HvCEL1 in the pTOOL7 vector in Origami (DE3) cells.

Western blot was performed using the 6xHis tag. The separation was performed using MES buffer and the expected size of the expressed protein is 68.5kD. Incubations occurred at room temperature (RT) (Lanes 2, 3 and 4), 20°C (Lanes 5, 6 and 7) and 37°C (Lanes 8, 9 and 10). Cultures were induced with IPTG using 0 mM (Lanes 2, 5 and 8), 0.05mM (Lanes 3, 6 and 9) and 0.1mM (Lanes 4, 7 and 10). Panels A (Coomassie stained gel) and B (western blot) show soluble protein extracts and C (Coomassie stained gel) and D (western blot) represent the insoluble fraction. The arrows indicate the predicted position of the expressed protein.

has a non-conserved disulphide bridge (Khademi et al., 2002). Being non-conserved does not discount the possibility that disulphide bonds exist between cysteines in the plant endo-(1,4)-β-glucanases. It was hoped that improved folding and stabilisation may result in the production of some soluble protein.

4.3.1.3. HvCEL1 in pTOOL7

Three induction temperatures and three different levels of IPTG were attempted. Concentrations of IPTG used were 0mM, 0.05mM and 0.1mM with incubations at room temperature for 8 hours, 20°C for 5 hours and 37°C for 3 hours. These results are presented in Figure 4-9. There was no detectable protein expression without induction of the promoter by IPTG indicating that no "leaky" expression was occurring. Again, large amounts of insoluble protein were expressed with much degradation and a faint band in the soluble fraction that was just detectable by western blot analysis indicated that there may have been some soluble expression. It was also possible that the band detected by western blot was contamination from the insoluble protein fraction during processing. Further incubations of HvCEL1, as well as HvCEL3 and HvCEL11 were performed and are described in the next section.

4.3.1.4. HvCEL1, HvCEL3 and HvCEL11 in pTOOL7

HvCEL3 and three *HvCEL11* cDNA constructs were cloned into pTOOL7. The *HvCEL11* constructs consisted of a full length sequence without the signal peptide of 73 kD (*HvCEL11_dSP*), a catalytic domain of 60.5 kD (*HvCEL11_CD*) and a 23 kD construct with just the carbohydrate binding domain (*HvCEL11_CBD*) (Table 4-2). These and the *HvCEL1* construct were used to transform Origami B (DE3) cells and expression was induced using autoinduction media. The cultures were initially incubated in LB media overnight. The next morning a 400 µl aliquot was placed into 3.6 ml autoinduction media and incubated at 23°C for a total of 30 hours. The results are presented in Figure 4-10. Again a large amount of insoluble protein was

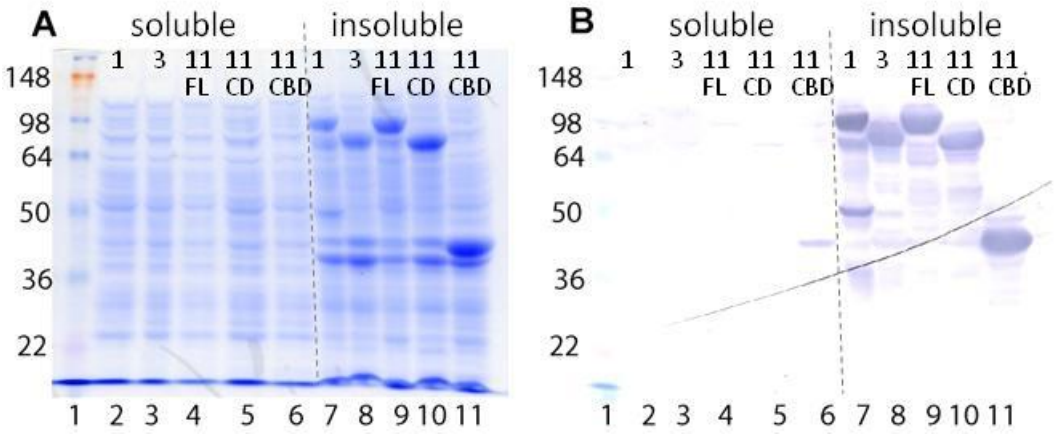


Figure 4-10: Coomassie stained gel and Western blot of HvCEL1, 3 and 11 in pTOOL7.

The cells were incubated in autoinduction media to induce expression. A Tris-glycine separating gel buffer was used. Panel A represents the Coomassie stained polyacrylamide gel and Panel B shows the results of the Western blot for HvCEL1 (Lanes 2 and 7), HvCEL3 (Lanes 3 and 8), HvCEL11_FL (Lanes 4 and 9), HvCEL11_CD (Lanes 5 and 10) and HvCEL11_CBD. For both panels, Lanes 2-6 contain soluble extracts and Lanes 7–11 contain the insoluble component. Expected size of the expressed proteins are HvCEL1 68.5kD, HvCEL3 63.3kD, HvCEL11_FL 73kD, HvCEL11_CD 60.5kD and HvCEL11_CBD 23kD

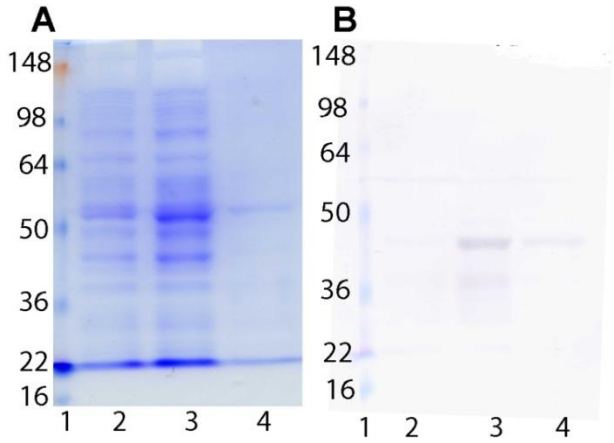


Figure 4-11: The Coomassie stain and Western blot of HvCEL11_CBD/pTOOL7 in *E. coli*.

The soluble protein crude extract was purified using Ni-NTA agarose and imidazole. Panel A is the Coomassie stained gel and Panel B is the Western blot. Lane 1 shows the SeeBlue2 marker, Lane 2 shows the flow through resulting from the 10mM imidazole wash, Lane 3 is the crude extract prior to binding of the protein to Ni-NTA agarose and Lane 4 shows the purified protein after washing with 250mM imidazole. The expected size of the expressed protein is 23kD.

expressed by all constructs. A distinctive positive band was seen in the soluble fraction of the HvCEL11_CBD by western blot, though this was not evident in the soluble fraction of the Coomassie stained gel. The mobility of this protein was however, that of a protein about twice the size of the expected 23kD and may have been a dimer. The remainder of the soluble fraction for HvCEL11_CBD was further purified using Ni-NTA agarose and imidazole. The gel in Figure 4-11 shows that many of the contaminants in the crude extract were removed but the partially purified protein was larger than predicted. Expression of the HvCEL11_CBD indicated that expression of the endo-(1,4)-β-glucanases may be possible using the prokaryotic system, however, expression of the CBD was not directly relevant to the original aim of the heterologous expression experiment. Therefore in the interests of time, induction of its expression with the aim of increasing quantity of protein was not continued. As time was becoming a limiting factor it was also decided that attempts at expression of HvCEL11_FL and HvCEL11_CD should also cease. Furthermore with the probability that at least the plant GH9A endo-(1,4)-β-glucanases may require glycosylation (Molhoj et al., 2001; Master et al., 2004), subsequent protein expression was performed using the eukaryotic *Pichia pastoris* expression system instead.

4.3.2. Protein expression in *P. pastoris*

HvCEL1 and *HvCEL3* were expressed in the *P. pastoris* system using the pPICZαA and B vectors (Invitrogen). *HvCEL1* was cloned into the pPICZαA vector in frame with the 6xHis tag. For *HvCEL3*, pPICZαB was used as several of the restriction enzyme sites in the pPICZαA vector were also in the *HvCEL3* cDNA sequence. Successfully cloned constructs were transformed into *P. pichia* by electroporation and selected with Zeocin. Initially, 12 colonies were cultured in 5.0 ml BMMY and BMGY in 50 ml tubes for further analysis. Once induction had commenced, 0.5 ml aliquots were removed at 2, 4 and 6 days and dot blots performed on all supernatants to determine those clones producing the greatest

levels of expressed protein. No heterologously expressed protein was evident after 2 days of incubation, at 4 days HvCEL1 was expressing at low levels and at 6 days HvCEL1 had four clones expressing at low levels and HvCEL3 had one clone expressing protein. A further 40 colonies for each gene were selected and grown in 2 ml cultures in 10 ml tubes with 200 µl aliquots removed at 2, 4 and 6 days of induction. Dot blots were performed and five clones for HvCEL1 and three clones for HvCEL3 were weakly positive. The supernatants of these clones were subjected to SDS PAGE and western blot. No protein was evident. The supernatants were concentrated by precipitation with trichloracetic acid. Again the Coomassie stain and western blot did not indicate that any protein was present in the cultures although it is possible that the trichloracetic acid had degraded the protein. As the dot blot indicated the presence of protein, it was concentrated and purified using Ni-NTA agarose and eluted with 250 mM imidazole. A dot blot was performed indicating that protein was present, however, a Coomassie stained gel and western blot were both negative. A viscometric cellulase activity assay was negative.

Since the amount of protein was putatively too low to be detected by western blot, Coomassie stain or activity assay, 500 ml cultures were set up with a view to increasing the quantity of protein produced for subsequent purification and concentration. Cultures were induced for 6 days, the cells removed by centrifugation and protein in the supernatant was precipitated with 70% ammonium sulphate. The precipitate was dissolved in 3 ml of phosphate buffered saline and desalted on a PD-10 column. Expressed protein was purified using Ni-NTA agarose columns and eluted with a total of 1 ml 250 mM imidazole. SDS PAGE was performed and gels either stained with Coomassie or transferred to nitrocellulose membrane for western blot. The results are presented in Figure 4-12. Protein expression was not visible by Coomassie staining but the western blot indicated a quantity of degraded HvCEL1 protein and for both HvCEL1 and HvCEL3, a high molecular weight protein of approximately 160 kD was detected. It was possible that the high molecular

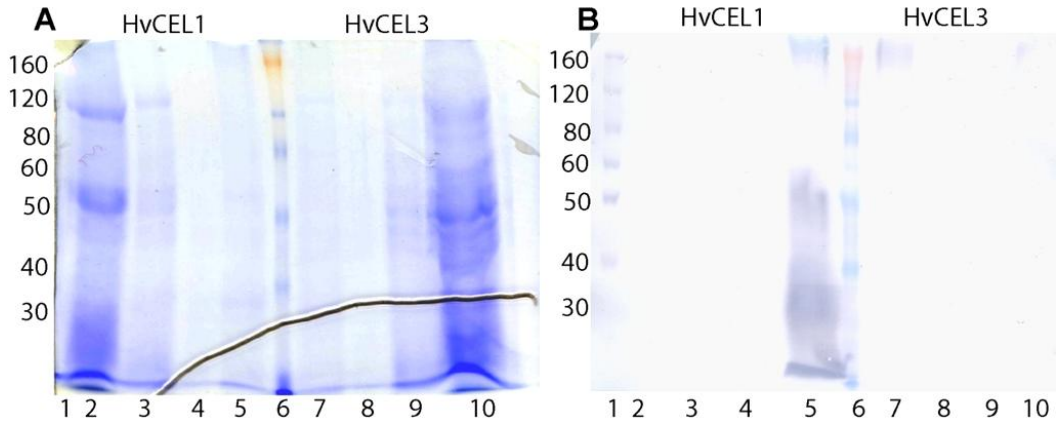


Figure 4-12: Coomassie stain and Western blot expression of HvCEL1 and HvCEL3 in *Pichia pastoris* for six days.

Panel A shows the Coomassie stained acrylamide gel and Panel B is the Western blot. Lane 1 is the 6xHis tag marker lane, the SeeBlue2 marker in Lane 6 separates HvCEL1 on the left from HvCEL3 on the right. Lanes 2 and 10 are the ammonium sulphate precipitate after desalting on the PD-10 column, Lanes 3 and 9 show the eluate of the 10mM imidazole wash, Lanes 4 and 8 show the resin after 250mM wash and Lanes 5 and 7 are the eluted protein after the 250mM imidazole wash. The expected size of the expressed protein is 68.5kD for HvCEL1 and 58.3kD for HvCEL3.

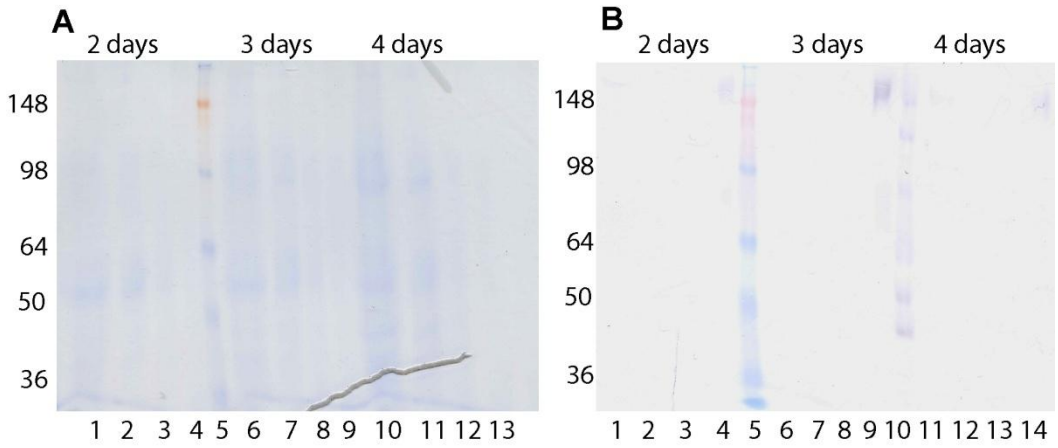


Figure 4-13: Coomassie stain and western blot for expression of HvCEL1 in *Pichia pastoris*

Panel A shows an SDS PAGE and Panel B the Western blot of expression of HvCEL1 in *Pichia*. Lane 5 is the SeeBlue2 marker and Lane 10 in Panel B is the 6xHistag marker. Lanes 1-4 after 2 days of induction, Lanes 6-9 after 3 days induction and Lanes 11-14 (Panel A) and 10-13 (Panel B) after 4 days induction. For each induction period the first lane (Lanes 1, 6 and 11) is fraction before desalting through the PD-10 column, the second lane (Lanes 2, 7 and 12) is the fraction eluted from the PD-10 column, the third lane (Lanes 3, 8 and 13) is the fraction collected after elution of the Ni-NTA agarose with 10mM imidazole and the fourth lane (Lanes 4, 9 and 13 Panel A/14 Panel B) is the fraction eluted from the Ni-NTA agarose with 250mM imidazole

weight was the result of glycosylation (Molhoj et al., 2001; Master et al., 2004) or that the proteins were forming oligomers or binding to other proteins. Viscometric enzyme activity assays in both cases were negative.

Based on the large amount of putatively degraded protein at day six, a second set of incubations of HvCEL1 only was performed. Three incubations of 250 ml each were performed with one induced for two days, the second for three and the third for four days. For all three incubations, protein was concentrated by ammonium sulphate precipitation, desalted on a PD-10 column and purified using Ni-NTA agarose. The results of the Coomassie stain and western blot are presented in Figure 4-13. Again, no protein activity was detected and only a small amount of protein, approximately 160kD in weight, was detected by western blot.

4.4. DISCUSSION

As outlined in the introductory section of the Chapter, successful expression of endo-(1,4)-β-glucanases in both prokaryotic and eukaryotic systems has been reported in the literature. It was therefore determined that it may be possible to obtain enzymatically active protein from at least one system. Heterologous expression in a prokaryote system is logistically simpler and requires less time than other heterologous expression systems such as yeast (Gräslund et al., 2008). Moreover, as eluded to in the introduction, others have produced enzymically active endo-(1,4)-β-glucanases using a prokaryote system. This was performed with three different genes and five constructs.

The cDNAs used for heterologous expression originated from three different clades as outlined on the phylogenetic tree described in Figure 2.6, Chapter 2. These three proteins were selected because they represented the three sub-families of the endo-(1,4)-β-glucanase family (Urbanowicz et al., 2007a). HvCEL1 is a KORRIGAN from the GH9A sub-family, but the primers used for amplifying the sequence were designed to exclude the transmembrane and cytosolic domains (Figure 1.15, Chapter 1). HvCEL3 is a soluble protein that co-transcribes with HvCEL1 as determined in the Chapter relating to transcriptional activity of the endo-(1,4)-β-glucanases. It is categorised as a sub-family GH9B protein because it lacks both a transmembrane domain and a carbohydrate binding domain however, from intron-exon and sequence analyses it is phylogenetically closer to the GH9A sub-family as described in Chapter 2. When compared with the other barley endo-(1,4)-β-glucanases, HvCEL3 had the highest level of transcript in stem tissues. From transcriptional analyses, HvCEL11 has very low transcript levels in stem but is highly transcribed in floral tissues and early developing caryopsis. HvCEL11 is from the GH9C sub-family and contains a carbohydrate binding domain. To avoid reported instability of the full-length protein when expressed heterologously (Urbanowicz et al., 2007b) or extracted from fresh tissue (Yoshida et al., 2006), a protein with just the catalytic

domain and a construct consisting of only the carbohydrate binding domain were prepared as well as the full-length protein with both the catalytic and carbohydrate domains intact. All cDNAs were amplified such that the expressed protein lacked a signal peptide.

In using the prokaryotic expression system, several strategies were implemented in an attempt to increase the amount of soluble protein. These included reducing the induction temperature, use of a vector with the thioredoxin fusion protein containing an easily refolded tag recognised by the bacterial chaperones, the use of *E. coli* strains with mutations that enable disulphide bond formation and both IPTG induction and autoinduction as a means of inducing transcription of the T7 RNA polymerase.

As a strategy to increase levels of soluble protein, lowering the induction temperature slows the rate of transcription (Sahdev et al., 2008), and also slows the rate of translation and hence allows more time for folding (Vera et al., 2007). Reducing induction temperature may also increase the amount of soluble protein by increasing the quality of the protein in the inclusion bodies. This strategy was unsuccessful in producing more soluble protein, but the quality of the protein in the inclusion bodies was not investigated.

Use of different vectors and host strains was also employed to try and improve levels of soluble protein. Two different vectors, pET-DEST42 and pTOOL7, were used for expression of protein. After the failure of pET-DEST42, it was determined that pTOOL7 may produce some soluble protein as it contained the thioredoxin tag to facilitate folding. Different host strains were used to reduce protease degradation, improve folding strategies and control the rate of transcription.

The autoinduction method makes use of the T7 promoter system, whereby *lac*UV5 is inhibited by glucose (Studier, 2005). The glucose blocks the promotion of T7 mRNA polymerase and once removed, the lactose is able to

promote the transcription of the T7 RNA polymerase and induce production of the target protein. Autoinduction removes the need to check the density of the cell culture before induction and also enables the use of higher densities of cells and therefore higher concentrations of target proteins are expressed. It is usually used to promote uniformity of expression between different cell cultures, but may also increase the amount of soluble protein when incubations are performed at 20°C compared to the use of IPTG (Studier, 2005). The use of very low levels of IPTG, to control the rate of induction, was tried but this proved unsuccessful.

Using the pTOOL7 vector with autoinduction and the Origami B (DE3) cells was successful in producing soluble protein for HvCEL11_CBD, however, as this was not directly related to the project, increasing quantity and purification were not undertaken.

Unfortunately none of these attempts produced soluble protein for the proteins with catalytic domains. These proteins have a larger and more complex structure than the HvCEL11_CBD, which increases the difficulties inherent in heterologously expressing proteins. However, there are other options that were not implemented here that have the potential to provide a better outcome. Varying the 3' and 5' ends of the cDNA, refolding of the insoluble protein sequestered in the inclusion bodies, codon optimisation of the cDNA, screening of a very large number of transformants, use of a less active RNA polymerase and coexpression of folding chaperones are strategies that have been used successfully to enable an increase in the amount of soluble protein produced (Varenne et al., 1984; Gräslund et al., 2008; Burgess et al., 2009).

Gräslund et al. (2008) advise an initial expression trial with at least 10 different cDNA constructs, using three different primers for the 5' and 3' ends of the cDNA as well as the full length construct. The reason for this is that small differences in the NH-terminal and COOH-terminal sequence can affect the solubility of the protein. Also, altering the position of the 6xHis tag may affect solubility of the expressed protein.

Most of the protein was present in inclusion bodies as the insoluble fraction. Inclusion bodies result from misfolding or partial folding of the protein. This produces a "sticky" aggregate that forms into dense insoluble inclusion bodies approximately 0.2 µm to 0.5µm in diameter (Burgess et al., 2009). It may take from seconds to minutes to fold a protein, depending on whether there are prolines in the amino acid sequence that slow down the process. Correct folding of eukaryotic proteins in prokaryotic systems is in the order of 10% (Gräslund et al., 2008). Fortunately, proteins contained in inclusion bodies are usually safe from proteolytic enzyme activity, therefore refolding misfolded or partially folded proteins is an option. Indeed, the Coomassie stain and Western blot of HvCEL11_FL that was sequestered in the inclusion bodies indicated that proteolysis was minimal. This was in contrast to the experiences of Urbanowicz et al. (2007) and Yoshida et al. (2006). Burgess et al. (2009) suggest that the majority of proteins sequestered in inclusion bodies could be salvaged under suitable denaturation and refolding conditions. Salvaging of partially folded proteins may be possible by simply subjecting the proteins to mild conditions that allow for complete folding. Otherwise, solubilising of proteins contained in inclusion bodies requires initial denaturation and then refolding. Refolding involves the systematic trialling of varying conditions, including controlled temperature, salt concentration, pH, reducing conditions and the presence of divalent cations (Burgess et al., 2009). Denaturants include SDS, Sarkosyl and urea, after which dilution of the protein is performed using a suitable refolding buffer. The refolded protein is cleaned, concentrated, and finally reoxidised to ensure that disulphide bonds are able to form. Successful refolding of proteins is, perhaps, a controversial issue as it is also advised that refolding is generally unsuccessful and should only be performed as a last resort (Gräslund et al., 2008).

Codon usage is known to vary markedly between species (Grantham et al., 1981). It is possible that a codon commonly used in one species is little used in another. The result may be a marked slowing in the rate of translation (Varenne et al.,

1984). These authors suggest that the search for the correct tRNA is the rate limiting step in protein chain elongation. Unfortunately, pausing translation may lead to its abortion, particularly if the uncommon codons occur within the first 15-25 codons of the cDNA (Eyre-Walker and Bulmer, 1993; Valdivia and Isaksson, 2005). With the availability of commercially produced, synthetic cDNA services, it is now possible to adjust codon usage to comply with the expression system to be used. Alternatively, tRNA supplemented strains of host bacteria may also be used to successfully overcome codon bias (You et al., 1999).

An option for increasing production of soluble protein is to screen more colonies and select those that have higher levels of protein expression. When expression was first attempted in the prokaryote system here, three different colonies were tested, however, there did not appear to be any difference in expression rates and none produced soluble protein, so no further screening was performed. In hindsight, with the lack of success using other means, a large scale screening of colonies may have been successful in finding a clone producing high levels of target protein.

Lack of success in producing significant amounts of soluble protein using a prokaryotic system and indications in the literature that glycosylation may be a requirement for successful production of enzymically active endo-(1,4)-β-glucanases (Molhoj et al., 2001; Master et al., 2004) prompted a trial in a eukaryotic expression system using the yeast *Pichia pastoris* as the host. Unlike protein expression in the prokaryote system, *Pichia pastoris* performs most of the post-translational modifications required by eukaryotic proteins. These include both N- and O-glycosylation, disulphide bridge formation and folding. Both N- and O-glycosylation are complex post-translational modifications that begin during translation, with further modifications occurring in the endoplasmic reticulum and Golgi (Wilson, 2002). Though N-glycosylation appears to be well studied and increasingly well understood, there is not a lot of information

regarding O-glycosylation in plants (Lerouge et al., 1998; Wilson, 2002). It would appear to be a well controlled and precise mechanism. On the other hand, the *Pichia* expression system can produce proteins that are over-glycosylated. During this project, *P. pastoris* expression produced small amounts of secreted target protein. These had a much higher than expected apparent molecular weight that may have been due to high glycosylation or the result of aggregations of the target protein with itself or other secreted proteins. HvCEL1 has two poly-proline sites in the expressed protein that may have hindered the translation of the protein as well as providing possible sites for non-covalent protein binding (Williamson, 1994; Kay et al., 2000; Master et al., 2004). Poly-proline sites are not present in HvCEL3 and the most likely cause of the apparently high molecular weight is excessive glycosylation as reported in the literature (Molhoj et al., 2001; Master et al., 2004).

Thus it was shown that protein was being secreted but the amount of purified protein was very small. It is possible that some protein is lost in the purification process, though this was not apparent from the Coomassie stained gels or the western blots. Low levels of secretion in yeast expression systems have been attributed to many factors including expression vectors, host strains, induction pH, induction temperature and cell density (Mattanovich et al., 2004). These factors often contribute to high stress levels in the yeast culture resulting in protein degradation in the endoplasmic reticulum with low levels of protein excretion.

Different strains of yeast are available for heterologous expression and Idris et al. (2010) recommends a screening process for the different modified host strains for selection of the one most suitable for the protein to be expressed. SMD1168 is a protease deficient strain and was selected because it had been used successfully in the laboratory in the past and would minimise protein degradation. Variation in the speed of processing, that is, folding, disulphide bond formation, N- and O-glycosylation and processing of the signal peptide

occur in the endoplasmic reticulum (ER) and may be a function of the strain (Idiris et al., 2010). Screening of different vector systems as well as promoters is also a feasible proposition. As outlined for prokaryote expression, small changes in the 5' or 3' ends of the sequence may affect folding and processing of the protein. There may be stalling of translation, transcription or folding leading to subsequent degradation.

Problems with protein trafficking that occur between ER and Golgi, or after Golgi processing, such as mis-sorting, result in protein degradation and reduced secretion into the media (Idiris et al., 2010). As with *E. coli* expression systems, codon usage may also be a factor in determining rates of translation (Grantham et al., 1981). This will slow the speed of processing and may induce protein degradation in the ER

One recommendation is that if buffer is not added to the media, the pH of the media drops to as low as 3. This reduces the amount of proteolysis by neutral pH proteases and while *Pichia* cultures are capable of growing at low pH (Brierley et al., 1994). Another possible solution may be to induce at a pH more in keeping with the optimum pH for enzyme activity. Khademi et al. (2002) indicated that at pH 5.6 the termite endo-(1,4)-β-glucanase was in its active state. There is also a report that incubating *Pichia* at pH 6, with 1% casamino acid included in the media, inhibited proteases (Clare et al., 1991). This would have the added benefit of expressing protein close to the pH at which the endo-(1,4)-β-glucanases are reported to be most active.

Another heterologous expression system uses insect cells and a baculovirus expression vector (Smith et al., 1983). Production of the recombinant virus is a complex and time consuming activity compared to *E. coli* and *P. pastoris*, which makes it a less attractive proposition. There are also reports of heterologous expression in tobacco leaves using tobacco mosaic virus as the vector (Shivprasad et al., 1999). These authors experienced some problems with translation but were able to achieve expression levels of the target protein of up to 10% of total protein

in the leaf. Though these systems look promising, a lot of time and a greater knowledge of the systems is required before they could be used successfully.

4.5. SUMMARY

Heterologous expression of the endo-(1,4)-β-glucanases in both *E.coli* and *P. pastoris* was unsuccessful. Three different genes and a total of five cDNAs were expressed using *E. coli* as the host cell. Along with several vectors and host strains, a number of different incubation parameters were trialled. Large amounts of heterologously expressed protein were produced but the majority was secreted as insoluble protein in the inclusion bodies. Further attempts at expression were therefore performed with the eukaryotic yeast host *P. pastoris*. Two cDNA constructs were cloned into the pPICZα vector designed to secrete the protein into the media. Transformants were screened and several clones were selected for protein expression. Purification of the protein, however, showed only low levels of secreted protein that was inactiveusing viscometric analysis. Lack of time prevented further troubleshooting, however, some suggestions were offered for improving the levels of expression of soluble protein.

In the following Chapter is described the transformation of barley with two GH9A cDNAs. One of these transformations was carried through to the second generation and the results indicate that while some plants continues to overexpress the transgene, in other plants transgene inactivation was occurring.

Chapter 5

Transformation of Barley with Endo-(1,4)-β-glucanase cDNAs

5.1. INTRODUCTION

In Chapters 2 and 3 the phylogeny and transcription of the endo-(1,4)-β-glucanase family of genes are described. Although the members of this gene family have similar sequence identities, the transcript abundance, both temporally and spatially, shows a wide range of variation, which suggests that different genes in the family have a variety of functions in various tissues and at different stages of plant development.

Experiments to increase expression of the endo-(1,4)-β-glucanases have been performed to study the functions of endo-(1,4)-β-glucanases. Overexpression of the endo-(1,4)-β-glucanases has been reported in dicotyledonous plants such as *Arabidopsis*, poplar (*Populus alba*), sengon (*Paraserianthes falcataria*) and hybrid aspen (*Populus tremula* x *tremuloides*). These experiments show that the influence of overexpression on the plant phenotype is dependent on the endo-(1,4)-β-glucanase in question. Overexpression of the soluble GH9B *Arabidopsis* endo-(1,4)-β-glucanase, CEL1 (AtGH9B1), in poplar produces a taller plant with larger leaves (Shani et al., 2004). This follows earlier reports associating this endo-(1,4)-β-glucanase with cell elongation (Shani et al., 1997). The poplar PaPopCEL1 (PtCel9B1), an orthologue of AtGH9B1, when overexpressed in *Arabidopsis* also produces a plant with larger leaves that result from an increase in cell size (Park et al., 2003). On the other hand, *Arabidopsis* transformed with the hybrid aspen PttCel9A1, the orthologue of *Arabidopsis* KOR1, is reported to produce transgenic plants that are phenotypically slightly smaller than their wildtype controls in both height and leaf size (Takahashi et al., 2009). Gross anatomy of the plant cells appeared to be normal in this case.

The effect of overexpression of endo-(1,4)-β-glucanases on cellulose levels also produces some varying results. Shani et al. (2004) record a 10% increase in cellulose content of basal stem material in poplar when overexpressing the *Arabidopsis* CEL1 (Shani et al., 2004). This is in contrast to the PttCel9A1 endo-(1,4)-β-glucanase, which, when overexpressed in *Arabidopsis*, produced total cellulose levels that were similar between transgenics and wildtype plants

(Takahashi et al., 2009). However, a decrease in cellulose crystallinity was reported in the transgenic plants.

When the poplar PaPopCEL1 (PtCel9B1) gene was overexpressed in sengon (*Paraserianthes falcataria*), similar amounts of cellulose to the wildtype were observed but with markedly reduced levels of xyloglucan (Hartati et al., 2008). This same construct when overexpressed in *Arabidopsis* again produced crystalline cellulose levels similar to that of the wildtype, and reduced xyloglucan concentrations (Park et al., 2003). This led the authors to speculate that the poplar endo-(1,4)-β-glucanase may be hydrolysing the xyloglucan in the cell wall. These plants also showed an increase in creep capacity per load applied and this reinforced the possibility of an increase in hydrolytic activity in the cell wall.

The endo-(1,4)-β-glucanases are hydrolytic enzymes, but their cellulose synthesis role is believed to be of major importance (Nicol et al., 1998; His et al., 2001; Lane et al., 2001; Sato et al., 2001; Szyjanowicz et al., 2004; Zhou et al., 2006) The importance of cellulose in imparting strength to the cell wall is briefly described in Chapters 1 and 3 as are the other wall components (Heuchert et al., 1983; Appenzeller et al., 2004). Cellulose alone is able to withstand very high tensile forces, or stress levels, but has very low strain capacity (Chanliaud, 2002). Stress is the force per unit area that produces a strain on a physical body. Strain is the amount of change in size and shape of a physical body under the action of applied forces. Therefore a cell wall comprising just cellulose is strong but inelastic.

The non-cellulosic components influence the alignment of the cellulose within the cell wall so changing the relative composition of the wall may have a dramatic effect on its mechanical properties (Chanliaud, 2004). For example, increasing the relative concentration of cellulose to xyloglucan produces a stronger but less elastic network. When these ratios are reversed, the cell wall becomes more pliant but less able to withstand stress forces. While removing the xyloglucan noncovalent interactions will produce less ordered cellulose

microfibril architecture with increased creep capacity, this would not compromise strength. Addition of non-cellulosic components, while reducing the wall's ability to withstand stress, would provide the wall with elastic properties and thus increase the strain capacity of the cell wall (Chanliaud, 2004). It is this combination of flexibility and strength that makes the cell wall an ideal composite for support of the plant. So is it possible to increase the amount of cellulose in the cell wall by, for example, increasing the levels of endo-(1,4)-β-glucanase expression, yet still retain sufficient flexibility to maintain strain capacity of the stem?

Barley full length endo-(1,4)-β-glucanase cDNAs driven by the 35s promoter were transformed into barley. The analysis of the first (T_0) and second (T_1) generations of plants grown to maturity is described. Experiments on T_0 plants include a transcriptional analysis of the transgene, Southern blot analysis of TDNA insertion as well as mature stem and leaf crystalline cellulose estimations and stem strength measurements. Several of the T_0 generation plants were selected for analysis at the T_1 stage. These plants were subjected to transcript analysis, and stem strength and cellulose analyses. Microscopy was also performed on the stems of selected T_1 plants.

5.2. METHODS

5.2.1. Amplification and cloning into the pMDC32 vector

5.2.1.1. Amplification of the barley cDNA sequence

Amplification of the open reading frame (ORF) of the *HvCEL1* endo-(1,4)-β-glucanase gene was achieved using leaf and root cDNA templates that were prepared as described in Chapter 3, section 3.2.4.6. The forward primer, CTCCCCCCTCTCCGTCCTCCTCCTCCG, was located 5' of the start codon and the reverse primer, GGTTTCCAAGCTCCCTCCATTTGATGC, was located in the 3' untranslated region (UTR). The *HvCEL5* endo-(1,4)-β-glucanase ORF was amplified using a cDNA template that contained a "cocktail" of RNA extracted from various tissues as described in Chapter 2, section 2.2.2.1. The RNA included caryopsis at 6 days after pollination (DAP), developing embryo, rachis at 12 DAP and stem. The forward primer, GCCACCATGCACTCCCAGAACCAC spanned the start site of the cDNA, whilst the reverse primer CAAGAAACATATACAGTAGTACAGGC was located in the 3'UTR

The Invitrogen polymerase mix, Elongase, with accompanying Buffers A and B were used in the ORF amplification mixes. The Elongase polymerase mix contains a *Taq* polymerase and a proofreading exonuclease enzyme to allow correction of wrongly incorporated nucleotides and ensure a more reliable sequence copy. A matrix of four different reaction mixtures used for amplification of the cDNA sequences is shown in Table 5-1. The four reactions were amplified using annealing temperatures of between 50°C and 64°C. The cycling conditions were as follows: Step 1: 2 mins at 94°C, Step 2: 30 secs at 94°C, Step 3: 30 secs at 55-64°C and Step 4: 30 secs at 68°C. Steps 2 to 4 were repeated for 39 cycles.

Table 5-1**:** Matrix for amplification of *HvCEL1* and *HvCEL5* cDNAs.

Amplification of the open reading frame of the *HvCEL1* and *HvCEL5* cDNAs was performed using the Invitrogen polymerase mix, Elongase, with several different concentrations of buffer and DMSO. The matrix shown in the table indicates four different PCR mixes used with annealing temperatures of between 50°C and 64°C.

	A	B	C	D
water (ul)	14	12.5	14	12.5
Buffer A (ul)	2.5	2.5	3.5	3.5
Buffer B (ul)	2.5	2.5	1.5	1.5
DMSO (ul)	1	2.5	1	2.5
5 mM dNTP (u	1	1	1	1
Template (ul)	1	1	1	1
4 uM F Primer	1	1	1	1
4 uM R Primer	1	1	1	1
Elongase (ul)	1	1	1	1
TOTAL (ul)	25	25	25	25

5.2.1.2. TA Cloning into the pCR8/GW/TOPO vector

As described in Chapter 2, section 2.2.3, successfully amplified sequences were separated on a 1% gel to remove unwanted amplification products and the appropriate DNA band was cut out of the gel and cleaned using the Macherey-Nagel Nucleospin kit as described in Chapter 2, section 2.2.5. Incorporation of an exonuclease into the Elongase (Invitrogen) polymerse mix reduces the number of DNA molecules that have been A-tailed. Therefore, if the amplification reaction was more than 24 hours old, it was A-tailed as described in Chapter 2, section 2.2.7.1 prior to cloning into the pCR8/GW/TOPO entry vector.

The double stranded pCR8/GW/TOPO entry vector (Invitrogen) contains several important features (Figure 5-1). The 3' thymidine overhangs accommodate TA cloning (Invitrogen, 2006). It contains a topoisomerase with a high energy covalent bond at the cloning site. This covalent bond provides the energy for efficient cloning of the insert into the vector, with subsequent release of the topoisomerase. The M13 forward and reverse primer sites are situated approximately 100 nucleotides from the insert cloning site and were used for sequencing. The vector also has spectinomycin resistance for *E. coli* selection. Finally, the pCR8/GW/TOPO vector contains the lethal ccdB gene located between the attL1 and attL2 sites employed for the Gateway recombination reactions (Invitrogen, 2004-2005). Successful cloning involves cleaving of the ccdB gene and its replacement with the gene of interest. The ccdB gene is removed when a DNA fragment is successfully inserted between the attL1 and attL2 sites, so that vectors not containing the insert are lethal to the host bacteria.

The reaction mixture used for cloning the amplified DNA sequence into pCR8/GW/TOPO contained 1 µl cleaned amplified DNA fragment, 0.5 µl pCR8/GW/TOPO vector mix (Invitrogen), 1 µl salt containing 1.2M NaCl and 0.06M $MgCl_2$ and 3.5 µl of filtered sterile water. The reaction mixture was incubated at room temperature for 3 hours.

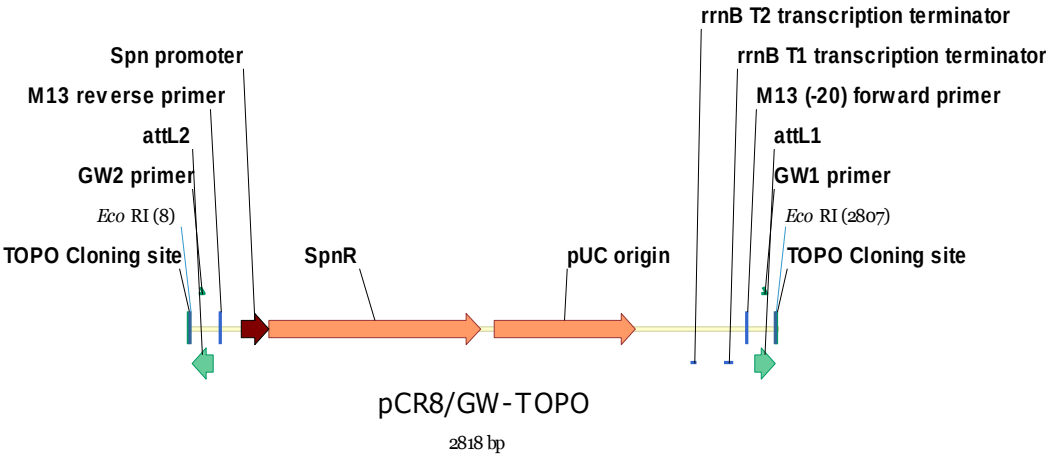


Figure 5-1: The linearised pCR8/GW/TOPO vector prior to cloning.

The diagram of the pCR8/GW/TOPO vector shows the topoisomerase (TOPO) cloning sites for AT cloning, the location of the M13 forward and reverse primer sites, *EcoR*I restriction enzyme sites and the GATEWAY recombination sites attL1 and attL2. SpnR indicates the position of the spectinomycin resistance gene for antibiotic selection. Not shown is the *ccdB* gene located between the attL1 and attL2 recombination sites in the circularised vector**.**

5.2.1.3. Transformation of DH5α E. coli cells

Heat shock transformation of competent DH5α *E. coli* cells was performed as described in Chapter 2, section 2.2.7.3. The heat shocked cells were incubated at 37°C in LB for one hour, spread onto LB agar plates containing 50 μg per ml spectinomycin (Sigma) and incubated at 37°C overnight. The transformed vector, or plasmid, was extracted from the *E. coli* cells by alkaline lysis, an *Eco*RI digest was performed to check insert size and the insert sequenced by AGRF using M13 forward and reverse primers as well as primers designed within the sequence. All of these activities are as described in Chapter 2, section 2.2.7.5.

5.2.1.4. Recombination from pCR8/GW/TOPO into pMDC32

Transfer of the insert from pCR8/GW/TOPO to the pMDC32 binary expression vector (Figure 5-2) was performed using Gateway (Invitrogen) technology (Curtis and Grossniklaus, 2003). As with the pCR8/GW/TOPO vector, the cloning site of the pMDC32 vector contains the lethal *ccdB* gene flanked by the attR1 and attR2 recombination sites. The pMDC32 vector also contains two selectable markers; an antibiotic resistance gene, kanamycin, for bacterial selection and hygromycin, for selection of plantlets that have been transformed with the pMDC32 vector.

The recombination reaction is mediated by the LR Clonase II enzyme mix that contains an enzyme to excise the insert at the attR1 and attR2 sites and a second enzyme to facilitate integration of the insert into the destination vector.

The reaction mix contained 150 ng of the pMDC32 (0.5 ul) vector, 150 ng pCR8/GW/TOPO vector with insert (1.5 μl), 2 μl Buffer enzyme mix and TE buffer to 10 ul. The mixture was incubated at 25°C for 5 hours or overnight. At the completion of the incubation step, 2 μl Proteinase K were added and the mixture incubated for a further 15 mins at 37°C. The cloning mixture was used to transform heat shock competent One Shot Mach 1 (Invitrogen) *E. coli* cells as

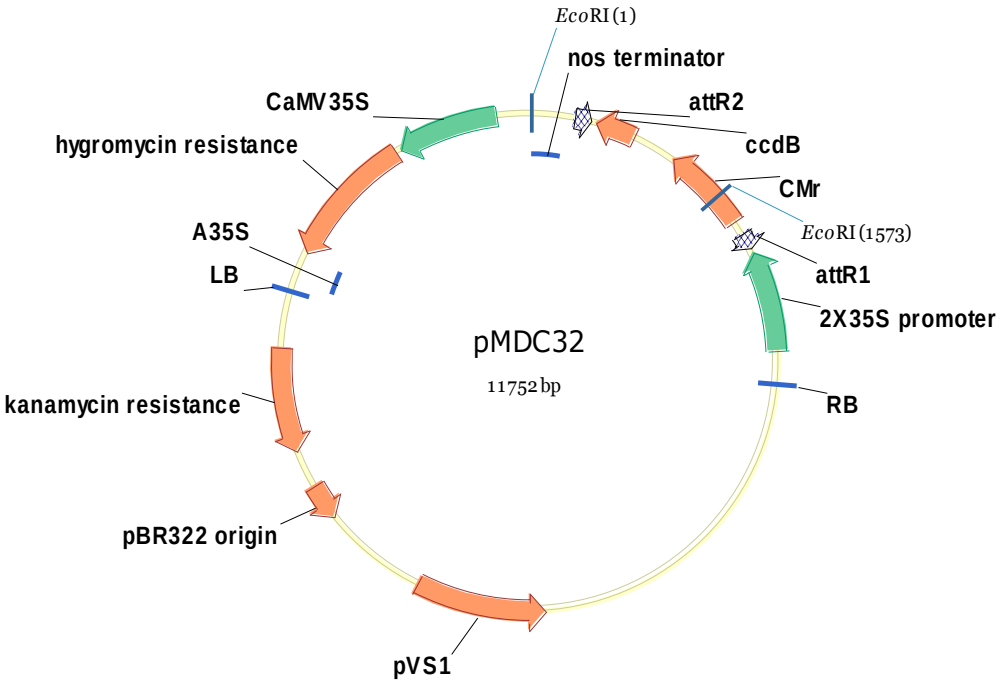


Figure 5-2: The pMDC32 vector.

The GATEWAY enabled vector uses the attR1 and attR2 sites for transfer of the target gene from the entry vector into the expression vector. The lethal ccdB gene and chloramphenicol resistance gene are located between the attR sites and are removed upon successful cloning. Other features important for its use in plant transformation include the 2 X 35S promoter, hygromycin resistance gene for plant selection, the kanamycin resistance gene for bacterial selection and the nopaline synthase (nos) terminator sequence (Curtis and Grossniklaus, 2003).

described in Chapter 2, section 2.2.7.3 and the transformed cells spread onto LB agarose plates containing 25 µg per µl Kanamycin (Sigma).

Cultures seeded with successfully growing colonies were grown on a shaker overnight at 37°C. Plasmid was extracted using the Qiagen miniprep kit where 2 ml culture was centrifuged at 10,000 x g for 30 secs, the supernatant was removed and remaining cells resuspended by vortexing in 250 µl Buffer P1 containing RNaseA and a further 250 µl of Buffer P2 was added. The tube was immediately inverted 4 to 6 times and 350 µl Buffer N3 was added and the tube inverted 4 to 6 times. The tube was centrifuged for 10 mins at 16,000 x g. The supernatant was applied to the Qiagen mini-prep spin column, centrifuged for 30-60 secs and the flow-through discarded. The spin column was washed with 750 µl Buffer PE and recentrifuged for 30 to 60 secs. The flow-through was again discarded and the tube centrifuged for an additional 60 secs to ensure removal of all Buffer PE. The column was placed in a fresh microfuge tube and the contents eluted with 50 µl Buffer EB, the tube allowed to stand at room temperature for 60 secs and centrifuged for 60 secs at 16,000 x g.

*EcoR*1 digests as described in Chapter 2, section 2.2.7.5 were performed on the prepared pMDC32/insert plasmids to ensure the insert was of the correct size.

5.2.1.5. Transformation of barley embryos with the pMDC32 vector

The barley transformations using the *HvCEL1* and *HvCEL5* cDNAs were performed by Dr Rohan Singh of the Plant Transformation Group at the Australian Centre for Plant Functional Genomics, University of Adelaide using the barley variety, Golden Promise. Transformation using the *Agrobacterium*-mediated system uses a binary vector system, whereby the vector can be replicated in both *E. coli* and *Agrobacterium* (Curtis and Grossniklaus, 2003). The *Agrobacterium* cells, transformed with the binary vector, infect the plant cell using its tumour inducing T1 plasmid that has been modified for avirulence. The *Agrobacterium tumefaciens* strain, AGL-1, was used for transformation of the barley embryos. Transformation of barley was effected by washing the immature embryos with the *Agrobacterium* cell culture containing the binary vector. This

was followed by callus induction, and plant production using hygromycin selection of successfully transformed embryos.

5.2.2. Growth of the transformed plants

Transgenic plantlets that survived hygromycin selection in tissue culture were placed into soil and transferred to the glasshouse. The plants were grown at approximately 23°C daytime temperature, down to a minimum of 13°C at night. Day length was set at 12 to 13 hours. Once the plants had flowered they were transferred to air-conditioned rooms where conditions were more variable with day time temperatures up to 25°C.

T_1 generation grain collected from six different T_0 plants was initially grown under growth room conditions. Two grains per pot were germinated and for the first six weeks the plants were grown on a short day cycle of 10 hours light and 14 hours dark with a temperature range of 18°C night and 22°C day-time temperatures. During this time leaf material was collected for DNA extraction and plants were thinned to one plant per pot. After six weeks, day length was adjusted to 12 hours retaining the same temperature range. At flowering the plants were transferred to the air-conditioned rooms in the glasshouse.

5.2.3. Genomic DNA preparation

5.2.3.1. Phenol/chloroform/iso-amyl alcohol extraction

Frozen leaf tissue was ground in a 2 ml microfuge tube and 600 µl DNA extraction buffer containing 1% sarkosyl, 100 mM Tris-HCl, 100 mM NaCl, 10 mM EDTA, pH 8.5 was added. Tubes were placed on ice until all samples had been ground and 600 µl phenol:chloproform:iso-amyl alcohol in a 25:24:1 ratio was added, mixed thoroughly by hand for 5 secs and again placed on ice until all samples had been processed. Samples were placed on an orbital mixer at room temperature for at least 10 mins and centrifuged for 10 mins at 10,000 x g. The upper layer was transferred to a fresh tube, care being taken not to disturb the interphase layer and both 60 µl 3M sodium acetate and 600 µl isopropanol were

added and the tube gently inverted to mix. DNA was allowed to precipitate at room temperature for two mins and the tube centrifuged for 10 mins at 10,000 x g. The supernatant was removed and the pellet washed in 70% ethanol, the tube centrifuged and the supernatant discarded. After drying, the pellet was resuspended in 30 µl RNase40, left overnight at 4°C and stored at -20°C.

5.2.3.2. Extract-N-Amp

The Extract-N-Amp (Sigma-Aldrich) DNA extraction kit was used as a rapid means of genomic DNA extraction. Leaf material approximately 0.7 cm in diameter with cut edges on all sides was placed in 20 µl Extraction Solution and incubated at 95°C for 10 minutes. The tubes were placed on ice and 20 µl Dilution Solution was added to the tube.

PCR was performed using 5 µl of REDExtract-N-Amp PCR Ready Mix (Sigma-Aldrich), 1 µl DNA solution (as extracted above), 1 µl 4 mM forward and reverse primers and 2 µl filtered sterile water.

5.2.4. Southern Blotting

A restriction enzyme digest was prepared with 3-6 µg genomic DNA, 10 Units *EcoR*1 (NEB) restriction enzyme, 1 µl *EcoR*1 buffer (NEB) and sterile filtered water added to a total volume of 10 µl. The reaction was incubated at 37°C for five hours.

A 1% agarose gel was prepared using Ultrapure Agarose (Invitrogen) dissolved in TAE buffer. The total 10 µl volume of previously *EcoR*1 digested genomic DNA was loaded onto the gel and products separated overnight at 30 volts. The gel was viewed under UV light, after washing for 20 mins in an ethidium bromide solution, to ensure the DNA had been adequately digested.

Transfer of DNA to a Hybond N^+ membrane (GE Healthcare Life Sciences) was performed as follows. A sponge soaked in 0.4 M NaOH was placed in the bottom of a shallow tray and a piece of Whatman paper cut to the same size as the membrane was soaked in 0.4 M NaOH and placed on top of the sponge. Air

bubbles were removed by gently rolling a cylinder across the top of the Whatman paper. A plastic open frame screen was place over the Whatman paper. The gel was placed upside down inside the frame of the screen and air bubbles removed as before. The membrane was soaked for 30 secs in 0.4 M NaOH and aligned over the gel and the stack again rolled to remove air bubbles. Another piece of 0.4 M NaOH soaked Whatman paper was placed over the membrane and this was topped with a 10 cm wad of dry paper towels with a glass plate placed on top of the paper towel as a weight. The tray was filled with 0.4 M NaOH and transfer allowed to occur for at least 4 hours before dismantling the stack. The membrane was rinsed in 2X SSC (Chapter 2, section 2.2.9.2) and blotted dry with paper towel in preparation for probing the membrane.

The 814 bp hygromycin probe template was prepared by PCR using the GTCGATCGACAGATCCGGTC forward primer and GGGAGTTTAGCGAGAGCCTG reverse primer. Hygromycin probes were ^{32}P labelled as described for probe preparation in Chapter 2, section 2.2.9.2. Hybridisation of the membranes and autoradiography also proceeded as described in section Chapter 2, section 2.2.9.2.

5.2.5. Quantitative Real Time PCR

Preparation of cDNA for quantitative PCR was prepared as described in Chapter 3, section 3.2.4.6 using RNA extracted from leaf material. QPCR was performed as described in Chapter 3, section 3.2.5, with the difference that only one control gene, *HvGAPDH*, was used for normalisation of transcript numbers. For the *HvCEL1* transgene the forward and reverse primer sequences were GCATCCAAATGGAGGGAGCTTGGAAACC and GTTTGAACGATCG GGGAAATTC, respectively, producing a 131 bp fragment. Forward and reverse primers for the endogeneous gene were as used for QPCR in Chapter 3 and annotated in Appendix A, Table 3. QPCR of *HvCEL5* employed primers used for QPCR in Chapter 3 (Appendix A, Table 3). For QPCR of the hygromycin cDNA the forward primer was GCCGTGGTTGGCTTGTATG and the reverse

primer sequence was GGGGCGTCGGTTTCCACTAT, which produced a 291 bp fragment.

5.2.6. Stem strength analysis

Mechanical properties of stem and leaf material were measured using an Instron machine (Instron Pty Ltd) (Figure 5-3) and data analysed using Bluehills2 materials testing software. For leaf analysis, five of the most mature stems were selected and the flag leaves of these stems were harvested, interleaved with absorbent paper to ensure they remained flat and allowed to dry at room temperature. The middle 10 cm of the leaf was isolated and width and breadth measured using digital calipers. Tensile force required to break the leaves was measured by applying the leaf section to the pneumatic grips as illustrated in Figure 5-3 A and the grips "jogged" apart until the applied load was zero Newtons. The gauge length was reset to zero for extension, and testing commenced.

The stem strength analysis involved selection of the five tallest stems that were harvested after the seed had matured. Ideally, stems should be from the primary tillers, however, flowering heads on several plants had aborted and so this was not always possible. For the T_0 plants, the top three internodes below the peduncle were selected for testing. Stem diameter measurements were made using digital callipers and because barley stems are eliptoid in shape, an average of the smallest and largest diameter at the centre of the internode was taken. Total length and weight information of the three internodes was collected prior to measurement of strength at the centre point of each internode. For the T_1 plants, only the second internode below the peduncle was tested because previous statistical analysis (data not shown) indicated that testing of this internode alone yielded representative information for the entire stem. The three point flexure test was used for stem strength analysis with the stem balanced on the lower anvils of the Instron as illustrated in Figure 5-3 B. Using the specimen protect mode, a load of 0.5N was applied to the centre of the internode before the gauge was reset to zero prior to measurement to ensure that the gauge was resting on the

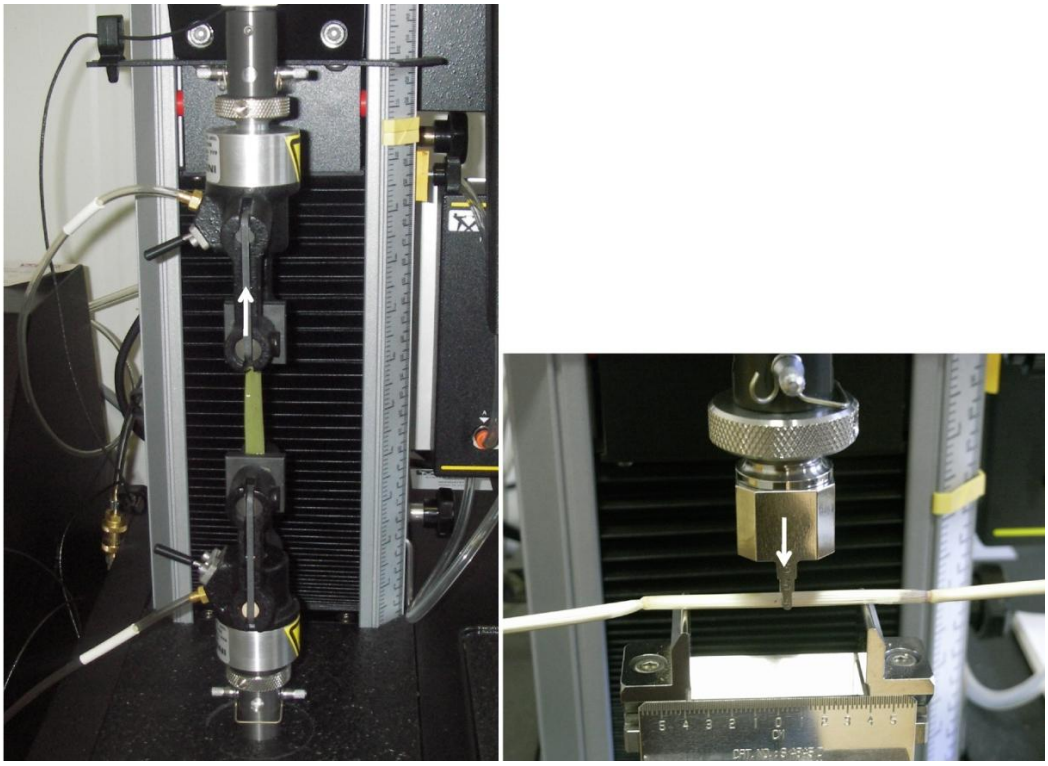


Figure 5-3: Measuring leaf and stem material using an Instron machine.

Two methods were used for measuring the strength of plant material using the Instron (Instron Pty Ltd) as indicated by the stem bending and leaf pull analyses. A: leaf pull analysis for strength testing of leaf material was performed by clasping the 10cm central portion of the leaf between pneumatic clamps and the force required to break the leaf was measured. The direction of force is indicated by the white arrow. B: stem bending for stem strength was performed by calculating the maximum force required to bend the stem internode at its centre. The stem was balanced upon the lower anvils and the maximum force required by the crosshead to bend the stem was measured. The white arrow indicates the downward direction of force.

internode but not placing any pressure on the stem. This ensured that measurements between internodes could be accurately compared. Testing commenced as the crosshead was lowered onto the stem to breaking point. Flexural load to break was measured in Newtons.

5.2.7. Cellulose analysis

Leaf material was cut into 0.5 to 1 cm pieces and ground to a coarse powder with ball bearings in a Retsch mill. Stem material was cut into 0.5-1 cm pieces and ground in a similar fashion using either a Capmix (ESPE GmbH and Company) or a GenoGrinder 2000 (Biotechnology Training and Consulting, Incorporated) to produce a coarse powder.

Leaf material and stem tissue that had been ground using the Capmix was dried overnight at 40°C in a vacuum oven prior to analysis. Stem tissue was freeze dried prior to grinding in the GenoGrinder and assumed to be dry because it had been subsequently stored in an airtight container with silica dessicant. Following both preparation methods, material was allowed to equilibrate to room temperature before accurately weighing between 15 and 20 mg of tissue into a 2 ml Eppendorf tube. The balance was tared with the tube and the material subsequently weighed into the tube. Acetic/nitric acid reagent containing eight parts acetic acid, one part nitric acid and two parts water was prepared and either 1 ml or 1.5 ml was added to the weighed material, sealed with microtube sealers and placed in an oven or heat block at 100°C for 90 minutes. The tubes were mixed gently 2 to 3 times during the incubation. After removal from the heat, the tubes were allowed to equilibrate to room temperature and centrifuged at 16,000 x g for five minutes. The supernatant was removed, care being taken not to disturb the friable pellet and the pellet was washed several times. Washes included an initial water wash, two 80% ethanol washes and a final acetone wash. After each wash, the sample was centrifuged at 16,000 x g for 5 min and the supernatant carefully removed. After the acetone wash, the pellet was allowed to air dry completely. Tubes were placed in the vacuum oven overnight at 40°C

with silica dessicant. The pellet was accurately weighed and the percent crystalline cellulose per mg dry tissue was calculated.

5.2.8. Microscopy

Samples of stem internode from selected T_1 plants were harvested at approximately 1-2 weeks after flowering. The stem was mature but still green. Different plants showing a range of phenotypes were selected for microscopy and included those with an unusual phenotype, several with a relatively normal phenotype, together with several control plants. As the internodes of the main tillers were reserved for stem strength and cellulose analysis, the stems selected for microscopy were from secondary tillers. The aim of microscopic analysis of the internodes was to examine the gross structure of the cells of the different internode tissues to determine if the stem was developing in a normal fashion.

The second internode below the flag leaf was initially selected for examination, and in selected cases the fourth internode was also analysed. Stem sections approximately 0.5 cm in length, obtained from the middle of the internode, were placed in 2 ml Farmer's fixative containing three parts ethanol and one part glacial acetic acid in 2 ml Eppendorf tubes. Sections were left in Farmer's fixative for two to four days, before being transferred to 80% ethanol for storage.

Stem sections were cut using the Leica VT1200 vibrating blade microtome (Vibratome). Initially, stem sections were secured to the cutting platform with a very thin layer of Selley's "Fix'N'Go Supa glue" (cyanoacrylate glue) and covered with filtered water. The blade was set to cut sections of 30 μm, proceeding through the section at a speed of between 0.8 and 1.0 mm per second with a blade amplitude of 0.8 - 1.0 mm. When cutting older, tougher tissues the slower speed and shorter amplitude parameters were more effective. Blade angle was adjusted for each piece of stem to obtain the thinnest possible sections.

After sectioning, the tissues were stained in toluidine blue containing 0.01% toluidine blue and 0.1% sodium tetraborate. Sections were stained for two minutes in the toluidine blue, washed in filtered sterilised water seven times to

remove excess stain, mounted on slides in 90% glycerol, covered with a slide cover slip and sealed with clear nail varnish. The prepared slides were viewed under bright field with the Leica AS LMD microscope using the Kohler illumination system.

Further slides produced from super glue impressions of the epidermal cells of the internodes were also prepared in order to calculate the average length of the epidermal cells in the different lines of the T_1 plants. Longitudinal pieces of internode approximately 3-4 mm in width and 1.5-2 cm in length were sectioned from the centre of the internode. A glass slide was thoroughly cleaned with 100% ethanol and a very thin layer of Selley's "Fix'N'Go Supa glue" (cyanoacrylate glue) was applied to the slide in a longitudinal strip. The stem section was carefully placed onto the glue, epidermal side down, and sandwiched with a second slide. The "slide-stem sandwich" was stabilised with a medium-sized bull dog clamp and left for exactly three minutes. The sandwich was dissembled by removing the clamp, carefully removing the second slide and gently easing the stem section from the glue. This was best achieved by grasping the end of the stem with forceps and pulling upwards but retaining an obtuse angle between the top of the stem and the slide. An imprint of the epidermal structures associated with the stem was left on the glue attached to the slide. The glue imprint was viewed under the Leica AS LMD microscope using Nomarski or differential interference contrast (DIC) microscopy. Using DIC microscopy enabled greater contrast between structures embedded in the glue imprint and provided a 3-D image. The number of cells spanning eight fields of view at 200X magnification was counted with eight replicates for each stem. The average number of cells across that span was calculated, the length of the span over which the cells were counted was calculated and divided by the average number of cells along that span to give the average length of each cell.

5.2.9. Statistical Analysis

All statistical analyses were performed using the Microsoft Excel statistical functions.

5.3. RESULTS

5.3.1. Cloning into pMDC32 and transformation into barley

5.3.1.1. HvCEL1

The full length ORF sequence for *HvCEL1* was successfully amplified using cDNA from leaf tissue. The PCR amplification mix, from Table 5-1, included 4% DMSO, 3.5 µl Buffer A (Invitrogen), 1.5 µl Buffer B (Invitrogen), and an annealing temperature of 60°C. Nucleotide sequence of the insert as cloned into the pCR8/GW/TOPO entry vector is presented in Appendix J. Two divergences from the published sequence were found. The published sequence has a proline at position 149 and a phenylalanine at 172. The sequence isolated here has a serine and tyrosine, respectively, in these positions which correspond with the wheat orthologue sequence. It was therefore concluded that the published sequence may be incorrect or that sequence variations are associated with the ORF of the gene or varietal differences. After recombination into the pMDC32 vector, the size of the excised fragment after *EcoR*1 digest indicated that the gene had been inserted so the plasmid was subsequently transformed into *Agrobacterium*. The *Agrobacterium* transformation of barley produced 15 plants, called G235-1 to G235-15, that survived hygromycin selection.

5.3.1.2. HvCEL5

A full length *HvCEL5* ORF was successfully amplified using the reaction mix D (Table 5-1), which included 10% DMSO, 3.5 µl Buffer A and 1.5 µl Buffer B at an annealing temperature of 57.6^{o}C. Nucleotide sequence of the insert as cloned into the pCR8/GW/TOPO entry vector is presented in Appendix J. At amino acid 255, the published sequence contains a histidine, however the sequence of the cloned insert contained a glutamine at this site, which is common to other barley endo-(1,4)-β-glucanases and the maize orthologue. As with HvCEL1, it was concluded that sequence variations may exist between different varieties of barley or the published sequence is incorrect. Cloning into pMDC32 and plant

transformation was performed as for *HvCEL1*. There were 19 plants that survived hygromycin selection and were labeled 234-1 to 234-19.

5.3.2. Analysis of T_0 transgenic barley plants

5.3.2.1. Phenotype

The phenotype of the T_0 plants may be difficult to interpret as the effects of the transformation process cannot be discounted. In general, the plants were of normal growth habit and all produced tillers and flower heads. Some plants did not produce viable seed while others had only two or three heads with viable seed. These plant types produced large numbers of tillers. In contrast, some plants were tall with broad leaves and low tillering and produced five or six viable heads. It was not possible to ascertain whether these effects were due to the transformation effects, or phenotype resulting from over expression of the genes.

5.3.2.2. Southern Blotting

A Southern blot was performed on all plants transformed with *HvCEL1* and *HvCEL5*, using a hygromycin probe. For *HvCEL1* it was clear that most of the plants had integrated the transgene at multiple loci, with plants 8, 10, 12 and 14 containing only one locus (Figure 5-4 and Table 5-2). There was some uncertainty around the exact number of TDNA loci for plants 6, 11 and 13. The Southern blot for *HvCEL5* (data not shown) was less clear due to problems associated with insufficient genomic DNA and poor digestion of DNA.

5.3.2.3. Transcript Analysis

For *HvCEL1* (G235), QPCR was performed using primers specific for the transgene as presented in Table 5-3. The forward primer was designed from the 3' UTR of the gene and the reverse primer was located at the 5' end of the nopaline synthase (NOS) terminator of the vector. Therefore the amplified product is specific for the transgene. This configuration relies on transcription of

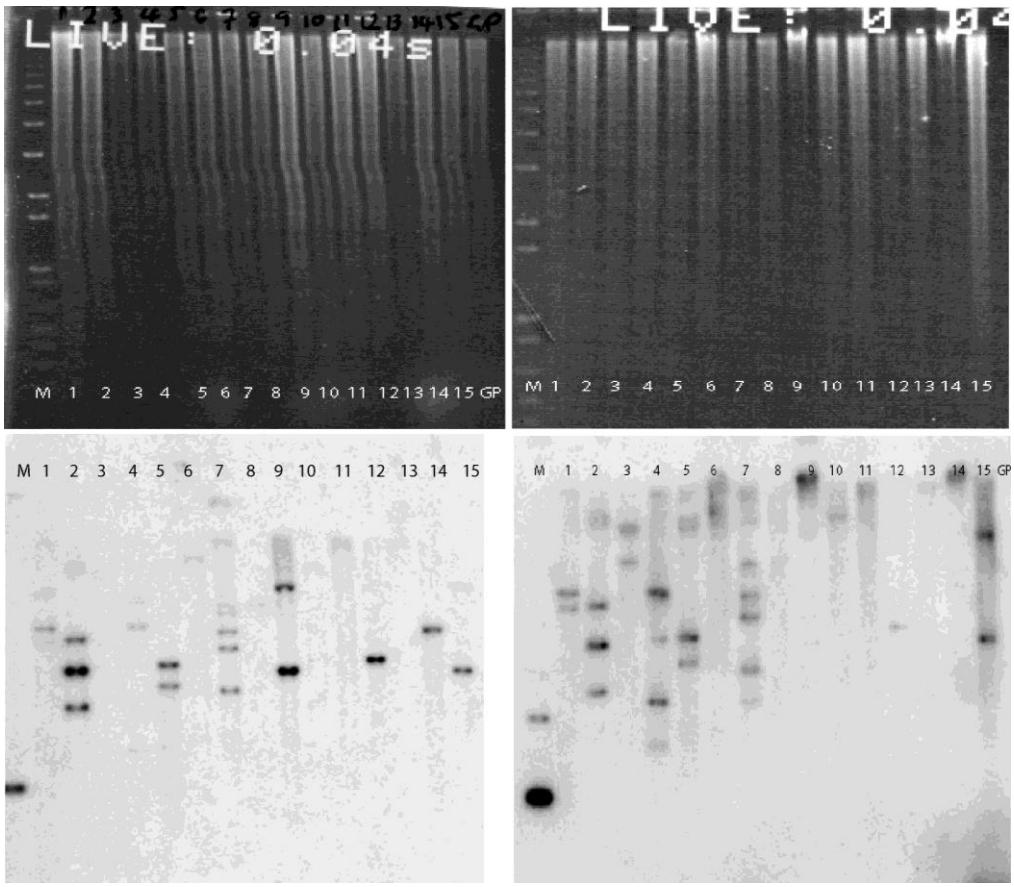


Figure 5-4: Southern blot for *HvCEL1* (G235) T_0 transformed barley plants.

Two gels (top) and membranes (bottom) were processed for the 15 plants that survived hygromycin selection. Plant identification numbers are at the bottom of the gel and top of the membrane. Membranes were screened using the hygromycin gene probe and show the number of loci in the genomes of the plants into which the hygromycin TDNA had integrated.

Table 5-2: A summary of the Southern blot analysis for *HvCEL1*.

The number of loci at which the transgene had integrated into the genome for each transgenic plant is shown. A question mark next to the insert number indicates uncertainty regarding the number of loci in that line. GP- wildtype Golden Promise control

HvCEL1	
plant no.	**No. of loci transgene integration**
G235-1	2
G235-2	3
G235-3	2
G235-4	4
G235-5	2
G235-6	?
G235-7	5
G235-8	?1
G235-9	2
G235-10	?1
G235-11	?1
G235-12	1
G235-13	?
G235-14	1
G235-15	2
GP	0

Table 5-3: Primers for QPCR.

QPCR for HvCEL1, HvCEL5, transgenes (Tx), hygromycin, and endogenous genes were performed using the primers as indicated in the following table. Length indicates the expected size of the amplified product.

Gene name	Forward primer	Reverse primer	Length (bps)
HvCEL1	CAAATGGAGGGAGCTTGGAAACC	TGATCCGTCATACCGATACAGAAGTAG	169
HvCEL1Tx	CAAATGGAGGGAGCTTGGAAACC	GTTTGAACGATCGGGGAAATTC	138
HvCEL5 and HvCEL5Tx	CGGTGTTGAGCAGATCCATCAGATG	CAAGAAACATCTACAGTAGTACAGG	175
Hygromycin	GCCGTGGTTGGGCTTGTATG	GGGGCGTCGGTTTCCACTAT	291

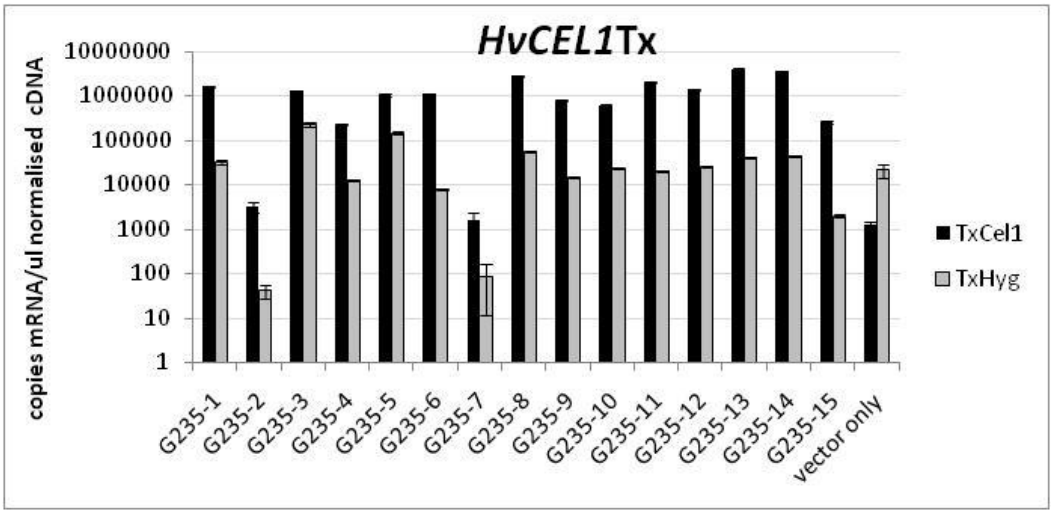


Figure 5-5: QPCR on *HvCEL1* (G235) transformed barley lines.

Relative transcript levels of the *HvCEL1* transgene (black bars) and hygromycin cDNA (grey bars) are shown here. The primers for the transgene span a region between the 3' UTR of the inserted cDNA and the NOS terminator sequence of the vector. Both G235-2 and G235-7 showed relatively low levels of transcript for both transgene and hygromycin. Vector only indicates control plant transformed with the uncloned pMDC32 vector.

the NOS terminator in the plant cell. QPCR using primers that amplify part of the hygromycin DNA was also performed (Figure 5-5). For plants 2 and 7 the hygromycin transcript was very low, however, the Southern blot analyses indicated that the DNA had inserted at multiple loci for both plants (3 and 5 loci, respectively Table 5-2). This suggests that either transcription of the TDNA in these plants is being suppressed or the transcript is being rapidly degraded, although hygromycin resistance was adequate for survival during the selection process. Experience in transformation of barley indicates that suppression of transcription sometimes occurs when several inserts are present in the genome (reviewed in Flavell, 1994; Vaucheret et al., 1998). Plants 4 and 15 also had multiple copies of the TDNA in the genome and had reduced levels of the transgene and hygromycin transcripts.

A total of 19 *HvCEL5* (G234) plants were analysed by QPCR using the QPCR forward and reverse primers as listed for *HvCEL5* in Table 5-3. Two sets of primers designed specifically for the transgene were unsuccessful in amplifying a product when either the pMDC32 vector carrying the *HvCEL5* cDNA or the cDNA prepared from leaf material of these plants was used as a template. Primers designed from the *HvCEL5* 3' UTR amplified a product of the correct size using the pMDC32-*HvCEL5* vector as template so it was expected that plants transformed with *HvCEL5* would show increased relative levels of transcript compared to endogenous transcripts. The results of the QPCR, however, as presented in Figure 5-6, show that levels of transgene transcript were similar, or less, than those found in the Golden Promise wildtype control. To ascertain if the vector was present in the plants, a benchtop PCR was performed using genomic DNA from the *HvCEL5* (G234) plants and the hygromycin primers used to amplify the hygromycin probe sequence. The agarose separation gel of this PCR is presented in Figure 5-7. Plants 1, 2, 12 and 19 were negative and so assumed not to be transgenic. However, the remainder of the plants contained the hygromycin gene and therefore also contain the *HvCEL5* insert. It was possible that either secondary folding of the transgene mRNA was

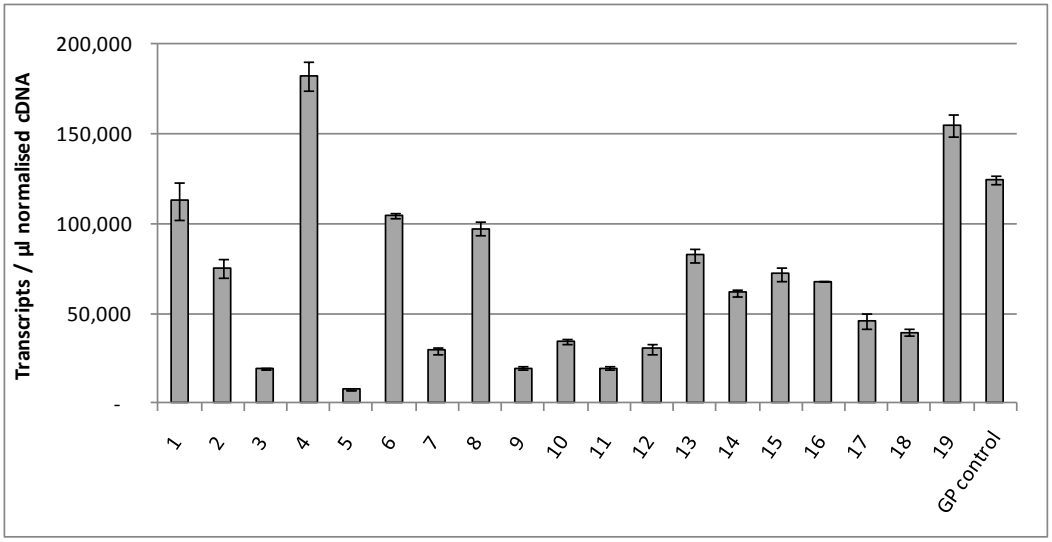


Figure 5-6**:** Relative levels of endogenous and transgene transcripts for *HvCEL5* (G234).

Primers from the 3'UTR of the HvCEL5 gene were used to determine levels of HvCEL5 endogenous and transgene transcripts. Much greater relative transcript levels were expected if they were to include both endogenous and transgene transcripts. GP indicates wildtype Golden Promise control

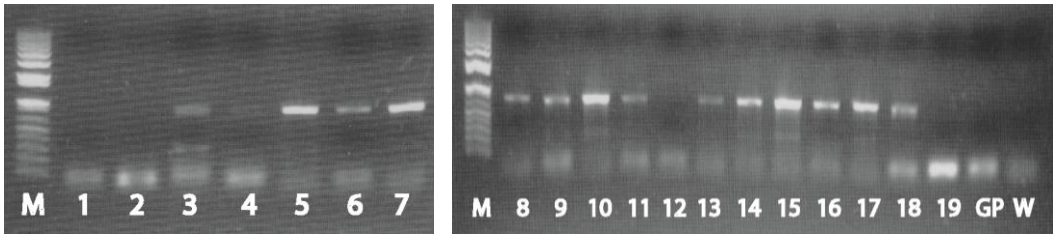


Figure 5-7: Agarose separation gel following amplification of the hygromycin gene to confirm the presence of the *HvCEL5* (G234) TDNA.

Hygromycin primers amplified a product of 814 bases from a genomic DNA template to confirm that the HvCEL5 (G234) TDNA was inserted into the genome of the plants. The products of the PCR were separated on a 1% agarose gel. Product was visible for all plants except 1, 2, 12 and 19 leading to the conclusion that these were not transgenic. GP- untransformed Golden Promise control, W- water control.

occurring to prevent binding of the primers or that gene silencing was occurring.

5.3.2.4. Stem and leaf strength and cellulose analyses

Stems were harvested after grain ripening and stored at room temperature until testing. They were subjected to flexural strength analysis and levels of cellulose of the ground stem material were measured. The results of these analyses are presented in Figure 5-8. Only eight of 15 plants were analysed fully for the *HvCEL5* (G234) transformation. G234-7 and G234-11 had the highest stem strengths at 5 N each and G234-9 and G234-14 had the greatest cellulose contents at 4.2 and 4.4 mg per cm stem, respectively. The vector control plants had also been through the transformation process and had lower stem strength and cellulose contents than the majority of the transgenics.

A wide range of stem strengths with variable cellulose contents were measured in the *HvCEL1* (G235) transgenic lines. All overexpression transgenic plants measured had flexural strengths similar to, if not greater than, the vector only control plants. Percent cellulose was variable and ranged between 12% in plant G235-5 to the highest at just over 45% in plant G235-6. Percent cellulose was not strongly correlated with stem strength, whereas mg cellulose per cm stem showed a strong correlation with stem strength with $r^2 = 0.71$ (Figure 5-9). These results are similar to those seen in maize by Appenzeller et al. (2004), despite the difference in anatomy between the hollow barley stem and that of the solid maize stem. Correlation with dry tissue weight was also noted and was similar to that observed by Appenzeller et al. (2004). Plants G235-12 and G235-14 had the highest levels of cellulose with an average of 5.4 and 5.8 mg cellulose per cm stem and the strongest stems with a flexural strength of 5.8 and 6.3 N, respectively. These two plants did not have the highest levels of cellulose measured as percent. That stem strength and cellulose per cm stem are correlated may indicate that cellulose *per se* may not be responsible for the strength of the stem, but other factors influencing total cellulose levels, such as the number of cells across the width of the stem wall may also be important.

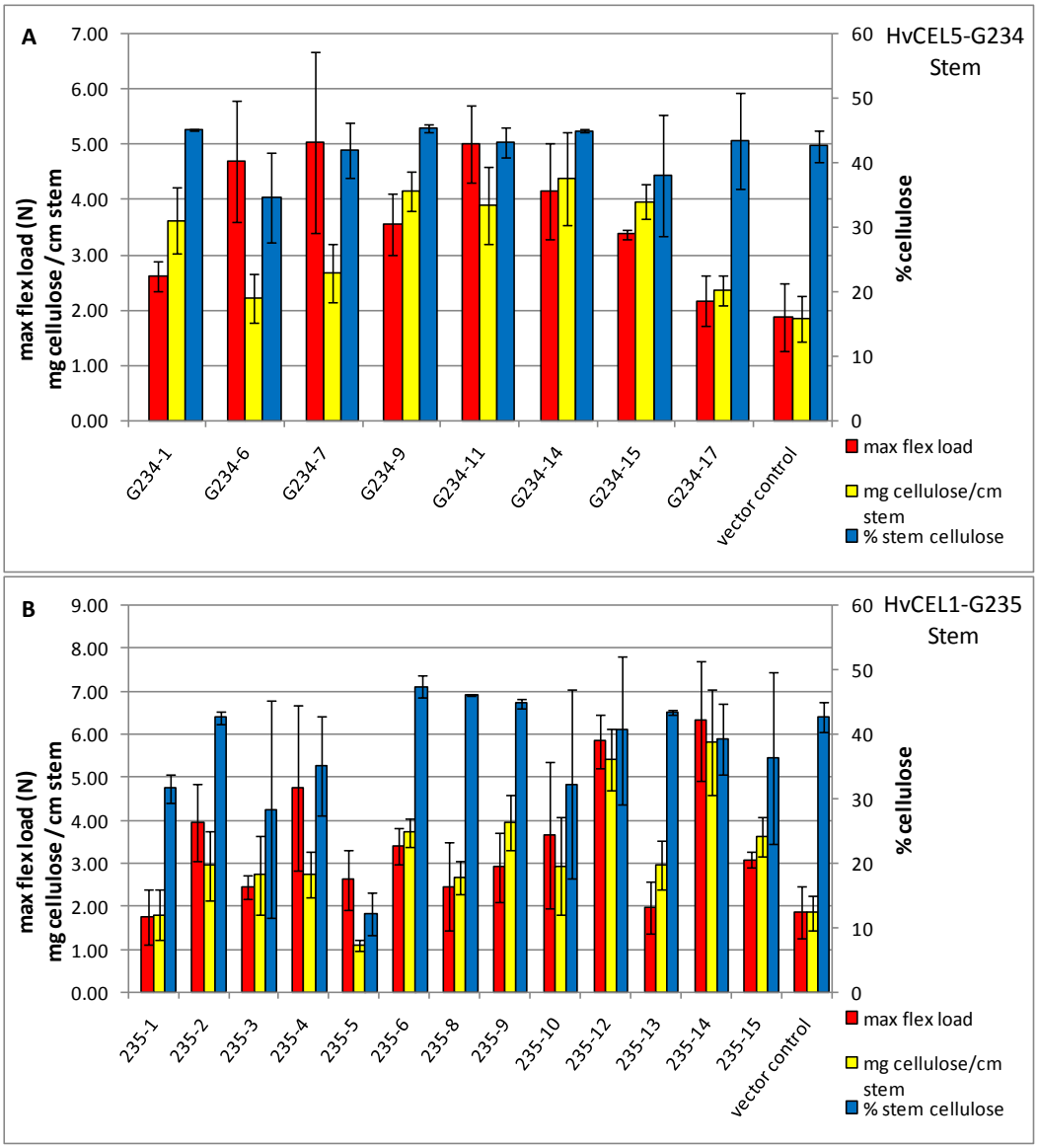


Figure 5-8: Combined results of the stem strength and stem cellulose content from *HvCEL1* and *HvCEL5* transgenic plants.

A: *HvCEL5* G234 and B: *HvCEL1* G235. For each plant five stems were measured for stem strength and the average for the three strongest stems on the plant were recorded. Graphs show maximum flexural load (Newtons) (red bar) and mg cellulose per cm stem (yellow bar) on the left axis and % cellulose (blue bar) on the right axis. Vector control indicates control plant transformed with the uncloned pMDC32 vector.

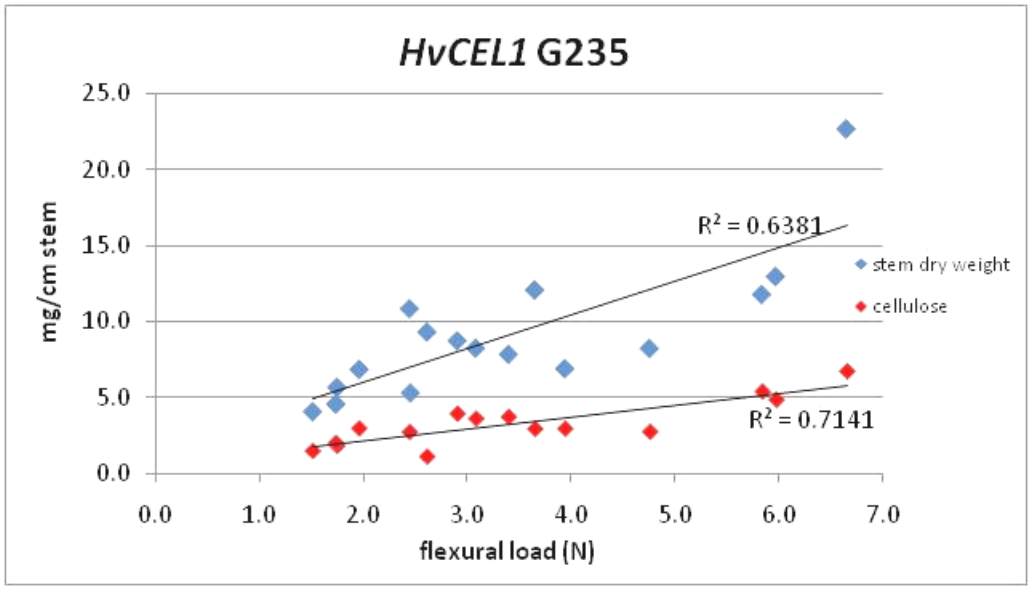


Figure 5-9: Correlation between flexural load and both stem dry weight and cellulose levels for *HvCEL1* (G235).

Taking the average for each plant, stem dry weight (blue) and cellulose (red) as mg per cm stem (vertical axis) were plotted against flexural load (N) (horizontal axis). Correlations are indicated for each set of data. These correlations are similar to those reported in maize (Appenzeller et al., 2004).

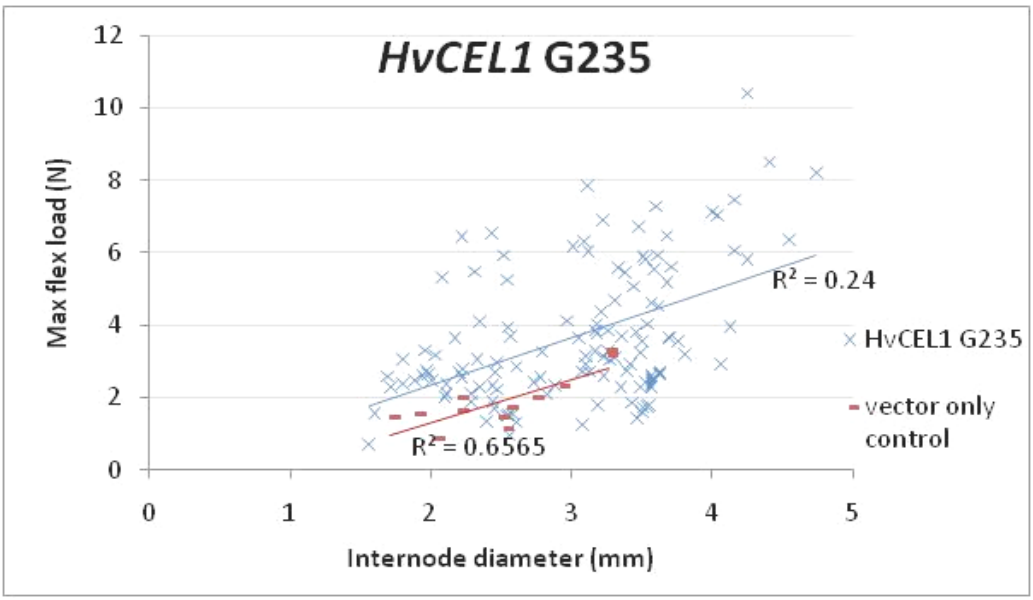


Figure 5-10: A graph of the correlation between flexural load and internode diameter.

The low correlation between stem diameter and flexural load to break for the transgenic plants, shown in blue, may indicate a dysfunction of cell wall synthesis. The vector-only control plants show a much higher correlation between stem diameter and stem strength as described by Berry et al. (2006).

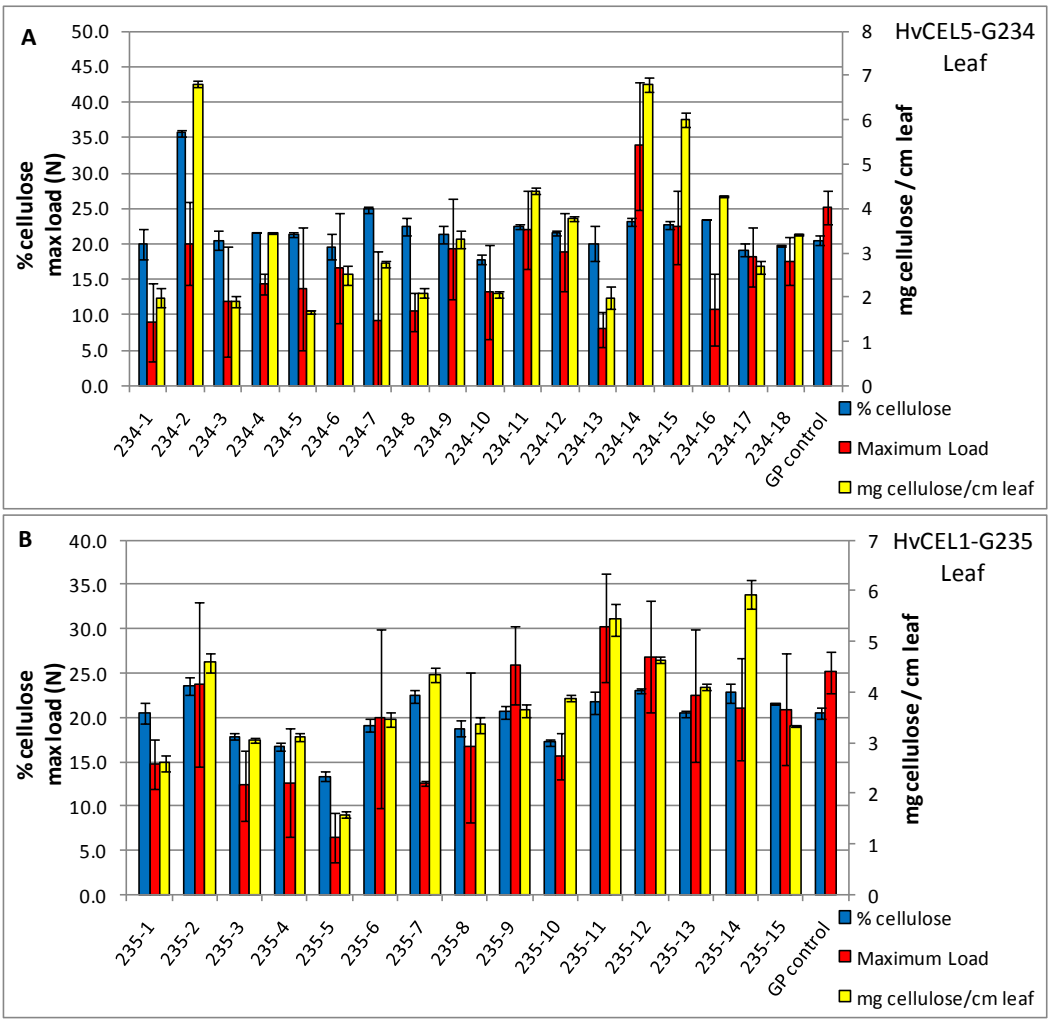


Figure 5-11: The results of the average leaf pulls for HvCEL5 (G234) and HvCEL1 (G235).

The left hand vertical axis represents % cellulose (blue bars) and the average maximum load to break (red bars) while the right hand vertical axis is mg of cellulose per cm leaf material.(yellow bars). A: *HvCEL5* G234 and B: *HvCEL1* G235. GP - Golden Promise wildtype control

Barley is more likely to lodge at the internodes in the middle of the stem and increasing stem diameter reduces potential for lodging (Berry et al., 2006). The correlation between stem diameter and stem strength in the T_0 transgenic plants as measured on the Instron here was not significant at $r^2 = 0.24$ (Figure 5-10). On the other hand, the control plants showed a correlation of 0.66. However, as mentioned previously, the effects of the transformation process often confound the appearance of a phenotype in the initial regenerant plants since transformation and tissue culture are stressful procedures.

Of the five flag leaves from each plant selected for testing, the results from the three strongest were averaged and graphed. Glasshouse grown, wildtype Golden Promise barley leaves were used as controls because no vector only control had been harvested. Results are shown in Figure 5-11 and indicate a wide range of leaf strengths not only between plants, but also within each plant. The *HvCEL5* transformation had 3 plants, 234-2, 234-14 and 234-15 that were high in at least one parameter measured. For *HvCEL1*, the plants with high leaf strength were plants 235-9 and 235-11 and those with high mg cellulose per cm leaf were plants 235-11 and 235-14. However, standard deviations were also very large.

Due to the difficulties associated with transcript and Southern blot analysis of the *HvCEL5* transgenic plants as well as restrictions in glasshouse space, no further analyses were carried out on these plants. Selected plant lines from the *HvCEL1* transformation were subjected to further analysis.

5.3.3. Analysis of the *HvCEL1* T_1 barley plants

A total of 12 plants from six different lines of *HvCEL1*-G235 were planted in 6" pots, to provide 72 plants in all. It was expected that the plants from parents carrying a single locus would be exhibiting Mendelian inheritance and provide null segregants to be used as control plants. Seed from lines G235-4, G235-9, G235-10, G235-11, G235-12 and G235-14 was planted. Selection of these second generation lines was not straightforward due to the large variation exhibited by the parental T_0 lines (Figure 5-8 and Figure 5-11). Plant G235-4 had smooth rounded stems that were the strongest stems relative to their

diameter, although percent cellulose and cellulose per mg stem were not high. G235-9 was selected on the basis that percent cellulose was high, as was mg cellulose per cm stem. The G235-10 line had only one transgenic locus and although it showed a low average maximum flexural load and cellulose content, the main stem was amongst the strongest of the G235 lines with above average levels of cellulose per cm stem. The T_0 G235-11 stems were not tested but leaf material indicated that the cellulose per cm leaf was high and the leaf strength was among the highest sampled. G235-12 and G235-14 lines contained only one locus at which the vector had inserted and also had higher than average mg cellulose per cm stem and high maximum flexural load to break. Of the 72 seeds planted only one failed to germinate, indicating good seed viability. Analysis of the T_1 generation of plants included an initial PCR analysis to determine those plants that did not contain the vector insert to be used as null segregant controls. Further analyses included a transcript analysis, measurements of strength and cellulose content of the stems from all plants as well as microscopy on selected plants.

5.3.3.1. Selection of null segregants and QPCR

Genomic DNA was extracted from leaf material using the Extract-N-Amp extraction and PCR kit (Sigma-Aldrich). Hygromycin primers were used for PCR to determine the presence or absence of the vector DNA, and plants that recorded three negative PCRs were deemed to be null segregants. These were retained as null segregant controls and included plants 9-1, 11-6, 12-1, 12-3, 12-5 and 12-6.

Transcript analysis of the selected null segregants suggested that 11-6 and 12-3 were null segregants because both hygromycin and transgene transcript were at background levels (Table 5-4; Figure 5-12). On the other hand, although levels for hygromycin were background only, levels of transgene transcript were slightly higher than background; 9-1 (6115 transcripts), 12-1 (3490 transcripts), 12-5 (2662 transcripts) and 12-6 (5427 transcripts) suggesting the presence of the

Table 5-4: Results of QPCR analysis on T_1 generation of the HvCEL1 plants.

Particularly outstanding are the Line 4 plants and the plants designated as TDNA suppressed that have relatively low levels of Hygromycin and transgene transcripts. These data are shown graphically in Figure 5-12.

Plant	Hygromycin	TxCel1	HvCel1
4-1	4770	79777	10378
4-2	593	7486	4220
4-3	960	12266	14061
4-4	221	2651	6050
4-5	327	8476	13103
4-6	405	7977	19118
9-2	246800	10676874	27728
9-3	157512	14061598	14428
9-4	190702	807788	1263
9-5	242715	1374264	2205
9-6	154689	9725299	31717
10-1	44004	750090	1571
10-2	98395	580526	2012
10-3	220367	1609320	1644
10-4	126131	2504927	1030
10-5	54617	625675	893
10-6	79848	979218	861
11-1	105424	6151305	19832
11-3	85258	5269122	10830
11-5	77840	2200140	6163
12-2	85197	5570319	19110
12-4	218347	1071680	757
14-1	19322	3015368	4070
14-2	22062	2682123	4153
14-3	32054	4272406	2251
14-4	38352	7571467	1665
14-5	37406	7811199	11500
14-6	25468	3420548	10290
control 11-6	25	47	10123
control 12-3	37	184	8088
TDNA suppr 9-1	382	6115	4395
TDNA suppr 12-1	84	3490	13105
TDNA suppr 12-5	87	2662	11351
TDNA suppr 12-6	117	5427	7042

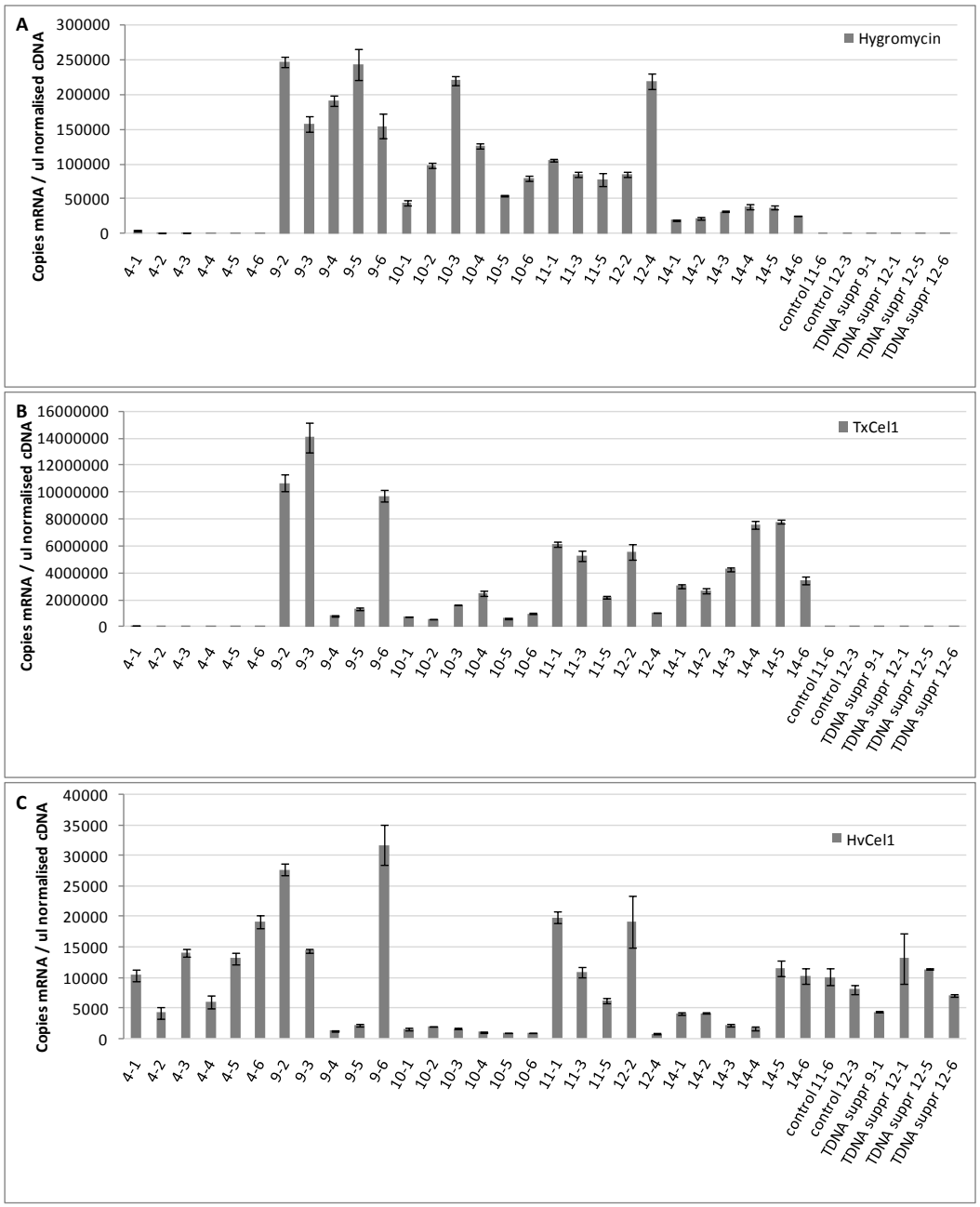


Figure 5-12**:** QPCR results for *HvCEL1* T_1 plants.

QPCR was performed using primers specific for Hygromycin, the *HvCEL1* transgene (*HvCEL1* Tx) and endogenous gene (*HvCEL1*) to provide information regarding transcription levels in the plants. This analysis revealed that plants 9-1, 12-1, 12-5 and 12-6, which were negative for hygromycin and selected on that basis to be used as null segregant control plants, contained low levels of transgene transcript (Table 5-4) and, therefore, were not null segregants.

vector DNA in these plants. All of these plants showed a relatively normal phenotype, but because of the possibility of gene suppression, which may also be affecting the endogenous gene transcripts, these plants were not suitable for use as null segregant controls. The results of the analyses on these plants were included in as a separate group.

QPCR analysis for the remaining plants showed that the transgene was transcribed at very high levels in all of the putative transgenic progeny except those in line 4 (Table 5-4;Figure 5-12). The highest level of transgene transcript was produced by 9-3 with over 14,000,000 transcripts per µl normalised cDNA. The lowest, with 580,562 transcripts per µl cDNA was from 10-2. Some of the transgenic plants showed low levels of endogenous *HvCEL1* transcript as well, which was particularly noticeable in 9-4 (1263 transcripts) and 9-5 (2205 transcripts), all of the line 10 plants (1571, 2012, 1644, 1030, 893 and 861 transcripts), 12-4 (757 transcripts), 14-3 (2251 transcripts) and 14-4 (1165 transcripts). It was noted that all plants with low endogenous transcripts also had the lowest levels of transgene transcript (less than 2.5 million), except 14-4 (7.5 million). All progeny of the line 4 plants showed relatively low levels of transgene transcripts, between 79,777 and 2,651 copies, and much reduced hygromycin transcription, with 221 - 4770 transcripts, which was markedly less than the other transgenic plants. These plants did not have reduced levels of endogenous *HvCEL1*.

5.3.3.2. Phenotype of the T_1 plants

Prior to stem elongation, except for two obviously small plants from line 11, the T_1 generation plants showed relatively normal development as indicated by the photographs in Figure 5-13. Excluding three smaller plants 11-2 (24 cm) and 11-4 (21.4 cm) and 10-3 (37 cm), and four larger plants (9-2, 14-3, 14-5 and 14-6) that were 50 cm or taller, plant height at two months of age ranged from 40-49 cm (Figure 5-14; Table 5-5). Figure 5-15 shows plant height at 2 months separating the plants into their various phenotypes. This graph reinforces the observation that the phenotypes could not be easily distinguished at this age. Although not manifest early in development, once stem elongation commenced

Table 5-5: A summary table of the results of the analysis of the T1 HvCEL1 plants.

Plants that died are highlighted in pink and null segregant control plants in yellow. Plants highlighted in blue (TDNA suppr) were deemed to be null segregants by PCR analysis of genomic DNA but were found to contain small levels of HvCEL1 Tx transcript and therefore could not be confidently used as null segregant control plants. N/A – not analysed, N/S – null segregant.

Plant	Phenotype	Height at 2 mths (cm)	Internode length (cm)	Stem strength (N)	% cellulose	mg cellulose/ cm stem	Diameter (mm)
4-1	normal	46	9.8	1.4	38.9	2.2	2.19
4-2	normal	46	7.3	2.5	41.8	3.2	2.38
4-3	normal	47	10.0	1.3	41.0	1.7	2.36
4-4	normal	47	8.9	2.5	45.7	2.3	2.36
4-5	normal	45	7.2	2.5	51.0	3.9	2.58
4-6	normal	47	7.9	1.8	45.3	2.3	2.58
9-6	intermediate	43	7.0	2.0	49.4	2.8	2.53
9-2	intermediate	51	6.4	1.5	48.0	2.6	2.74
9-3	normal	46	5.4	1.7	52.9	3.2	2.41
9-4	intermediate	40	5.6	0.7	16.9	1.0	2.05
9-5	dwarf	46	6.7	0.7	23.7	1.2	2.31
10-1	dwarf	42	5.4	2.1	24.1	2.6	2.44
10-2	intermediate	42	6.7	2.0	32.1	2.2	2.39
10-3	dwarf	43	4.3	0.5	11.4	2.3	2.07
10-4	dwarf	37	3.2	0.5	15.0	1.5	1.86
10-5	intermediate	42	6.2	2.0	33.5	1.9	2.19
10-6	dwarf	43	died	N/A	N/A	N/A	N/A
11-1	normal	48	7.8	1.4	52.7	3.8	2.87
11-2	dwarf	24	died	N/A	N/A	N/A	N/A
11-3	normal	48	8.3	1.4	49.1	2.3	2.56
11-4	dwarf	21.5	died	N/A	N/A	N/A	N/A
11-5	normal	42	7.6	2.4	51.0	4.5	3.08
12-2	normal	49	7.7	1.6	45.1	2.0	2.17
12-4	intermediate	48	7.7	1.9	21.1	1.8	2.49
14-1	intermediate	42	4.5	1.9	22.9	3.6	2.09
14-2	dwarf	46	2.9	1.5	18.7	2.7	2.00
14-3	intermediate	57	5.1	4.9	29.1	7.5	3.46
14-4	dwarf	47	died	N/A	N/A	N/A	N/A
14-5	intermediate	50	4.8	4.5	36.4	4.1	2.70
14-6	intermediate	54	7.2	4.5	34.3	4.3	2.47
11-6	N/S control	N/A	6.6	3.4	51.5	5.0	3.00
12-3	N/S control	46	8.4	1.9	38.3	3.9	2.44
9-1	TDNA suppr	46	8.2	2.1	27.8	2.7	2.76
12-1	TDNA suppr	44	10.8	3.4	39.7	3.1	2.71
12-5	TDNA suppr	44	6.4	4.4	38.6	2.7	2.79
12-6	TDNA suppr	49	9.1	4.4	31.4	4.6	2.95



Figure 5-13: Photographs of HvCEL1 T_1 plants at 2 months.

A: Control plants, B: G-235-4 with similar growth to the control plants, C: G235-9- similar growth to the control plants, D: G235-10- similar to the control plants, E: G235-11- the two smaller plants died, F: G235-12- similar to controls and G: G235-14- show reduced tillering. Apart from the two small plants from G235-11, there were no significant differences in the height of the transgenic plants compared to the null segregants, however, some plants showed a less upright phenotype. It should be noted that the QPCR analysis showed that only two of the six plants selected as null segregant control plants by PCR of genomic DNA were, in fact, null segregants.

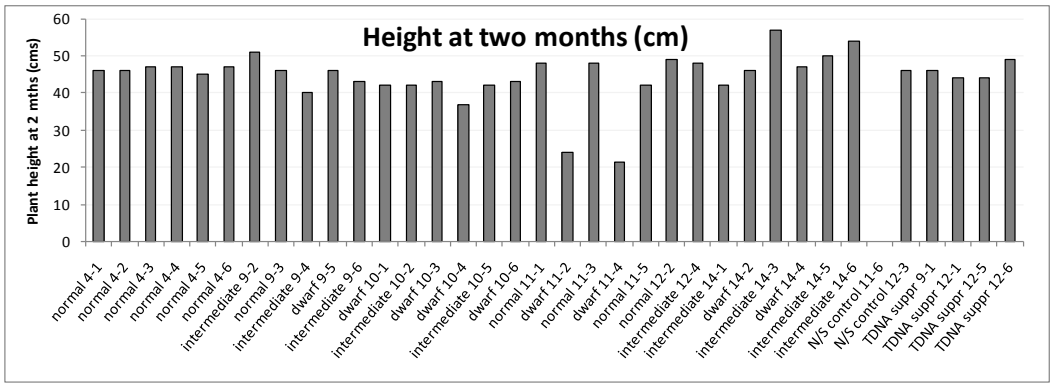


Figure 5-14: Height of the *HvCEL1* T_1 plants at 2 months.

These measurements were taken when the plants were two months of age (Figure 5-13). The majority of plants varied between 40 and 50 cm in height, with three plants, 9-2, 14-3 and 14-6, greater than 50 cm, and 10-4, 11-2 and 11-4 less than 40 cm.

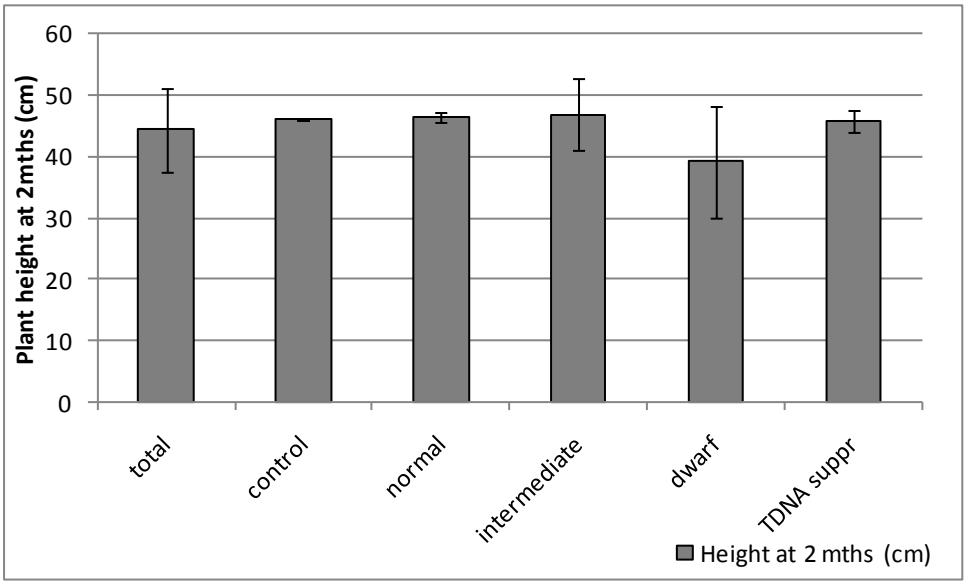


Figure 5-15: Average leaf height of *HvCEL1* T_1 plants at 2 months.

The average heights of the different phenotypes of the T_1 plants at two months of age prior to stem elongation are presented in this graph. The standard deviation for the average plant heights of the different phenotypes indicates that the differences between the phenotypes are not significant.

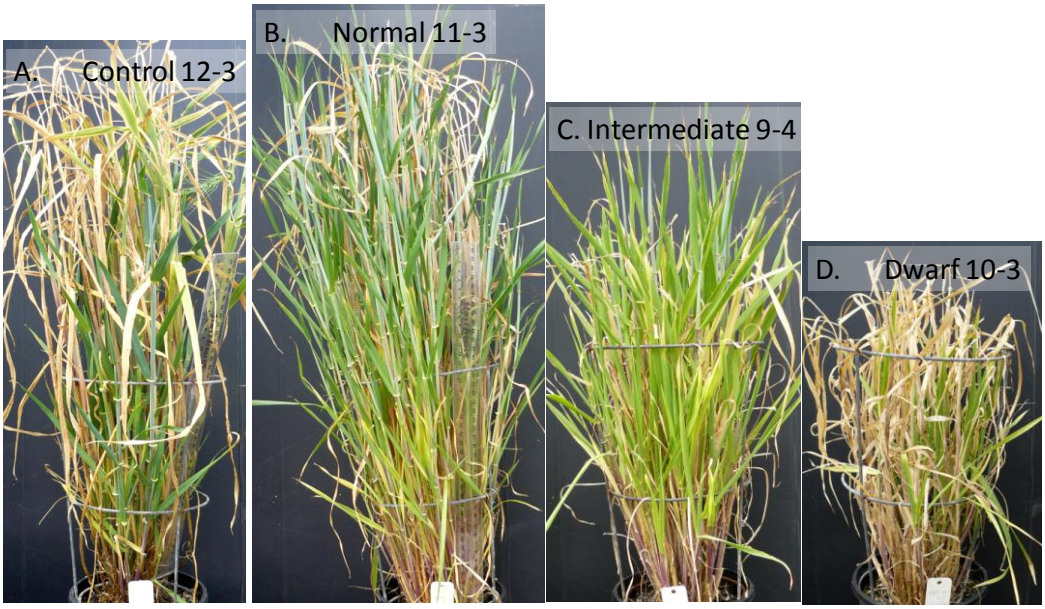


Figure 5-16: T_1 plants at 5 months.

Photographs of representative plants illustrating the variation in phenotype shown by the *HvCEL1* T_1 plants. A: null segregant control; plants selected as negative for hygromycin and transgene transcripts, B: normal phenotype with high levels of transgene and hygromycin transcripts and normal levels of endogenous *HvCEL1* transcripts, C: phenotype intermediate between normal and dwarf, D: dwarf phenotype plant with low levels of transgene and endogenous *HvCEL1*, and hygromycin transcripts.

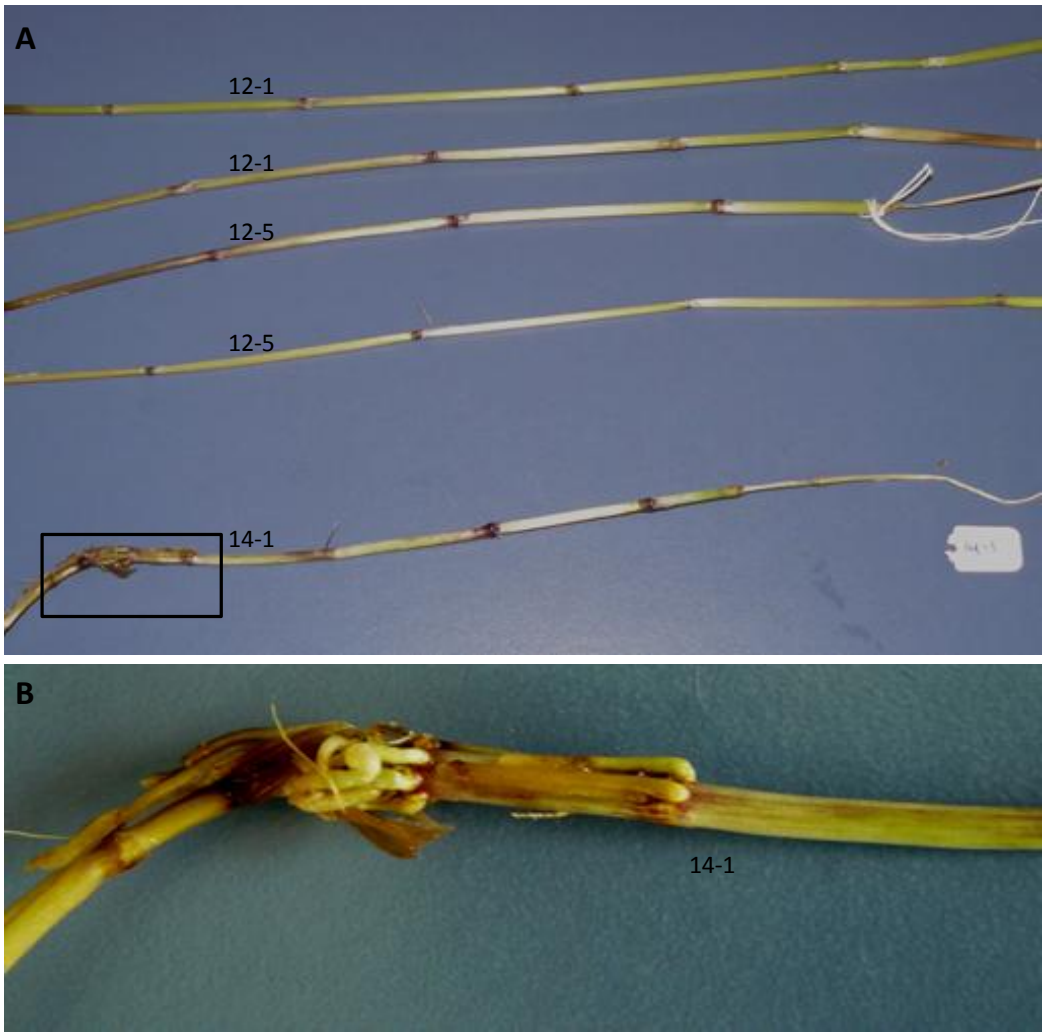


Figure 5-17: Adventitious root production.

A: shows four stems as indicated and 14-1 a dwarf plant with short internodes. B: shows the adventitious nodal root production that was observed in all of the transgenic dwarf plants.

some plants failed to grow as expected, producing a continuum of phenotypes. A subjective nomenclature for categorising the plants was derived, based on plant phenotype, as illustrated in Figure 5-16. Photographs of all plants can be found in Appendix G. All of the transgenic plants with the normal phenotype had an upright appearance and were of a similar height to the control plants. Some appeared bushier with large numbers of tillers, while others were less bushy than the control plants. At the other end of the spectrum were the dwarf plants that were categorised by their inability to grow much beyond the top of the support cage in the pot. In general, the dwarf plants were less upright in appearance and their stems were shorter than the controls and normal plants. The intermediate phenotype plants had a taller, more upright, bushy appearance that resulted from the production of large numbers of tillers. Exceptions were 14-3 and 14-5 that had the less upright untidy appearance of the dwarf plants but were of intermediate height. Although not quantified, it was observed that the dwarf and several of the intermediate phenotype plants produced tillers and associated stems that senesced soon after booting and were quickly replaced by another tiller and stem. Both of these phenotypes produced plants with very few or no seed heads.

Four plants died, all of which possessed the dwarf phenotype. The previously mentioned small plants, 11-2 and 11-4, died at three months of age, before being moved to the glasshouse. Although these plants were obviously dwarfed, they initially looked healthy, however, as the other plants grew and flourished, these plants stopped growing and failed to tiller or produce stems and finally succumbed to fungal disease. Dwarf plants 10-6 and 14-4 also died prematurely, but these plants showed a similar phenotype to the dwarf plant in Figure 5-16.

All of the dwarf plants produced adventitious roots from several nodes (Figure 5-17). This was not evident on any of the other transgenic plants or the controls. Also noted, but not quantified, was the increased predisposition to, and level of, aphid infestation seen on the dwarfed and some of the intermediate plants compared to the plants with a normal growth habit.

5.3.3.3. Stem strength and cellulose analysis of the T_1 plants

Results for the T_1 plants for internode length, average stem strength, percent cellulose, mg cellulose per cm stem and stem diameter are presented in Table 5-5. The four graphs that comprise Figure 5-18 show data derived from the stems of the mature plants. Figure 5-18 A illustrates flexural load to break (N) of the second internode below the peduncle and is the average of three stems. Most plant stems recorded average flexural loads to break of < 3 N. The average levels were greatest for plants 14-3, 14-5, 14-6, 12-5 and 12-6, all of which were > 4 N and 11-6 and 12-1, which recorded average flexural loads of > 3 N. Figure 5-18 B shows cellulose levels as a percent of total stem tissue. A large amount of variation was noted between different plants with eight plants (9-4, 9-5, 10-1, 10-3, 10-4, 12-4, 14-1 and 14-2) recording levels of less than 25% cellulose, while levels of > 45% cellulose were recorded by 11 plants (4-4, 4-5, 4-6, 9-2, 9-3, 9-6, 11-1, 11-3, 11-5, 11-6 and 12-2). Figure 5-18 C shows mg cellulose per cm stem and Figure 5-18 D graphs stem diameter for each plant.

As with the T_0 plants, correlation coefficients measured between several parameters were determined. Similarly to the T_0 plants, the correlation coefficient between percent cellulose and stem strength was low (r^2 = 0.04) (Figure 5-19 A). For the T_0 plants a strong correlation was observed between stem strength and mg cellulose per cm stem, but for the T_1 plants the correlation was less significant (r^2 = 0.53) (Figure 5-19 B) although stronger than that for percent cellulose. Several reasons for this are feasible. Because several of these plants failed to produce grain heads, or heads were not formed on the primary tiller stems, an increase in variability is likely and has been observed in the past (pers. comm. Dr Filomena Pettolino, University of Melbourne). Furthermore, it is also possible that because the cell wall synthesis machinery may have been perturbed, cell wall production is most likely dysfunctional, particularly in the dwarf plants. Figure 5-19 C illustrates a poor correlation between percent cellulose and internode length (r^2 = 0.24). Figure 5-19 D illustrates a moderate correlation of r^2 = 0.45 between internode diameter and stem strength. A correlation of r^2 = 0.57 was observed between internode

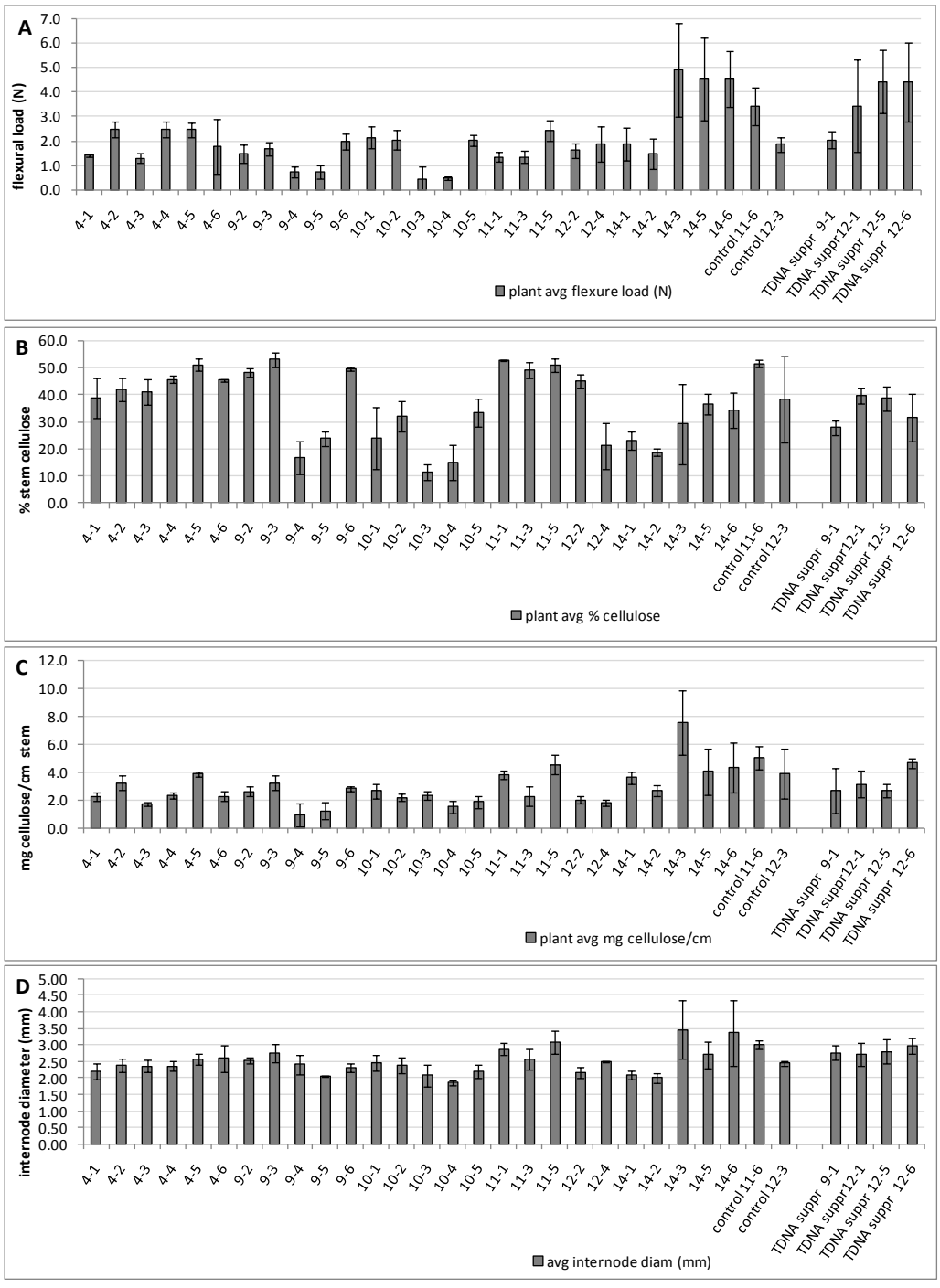
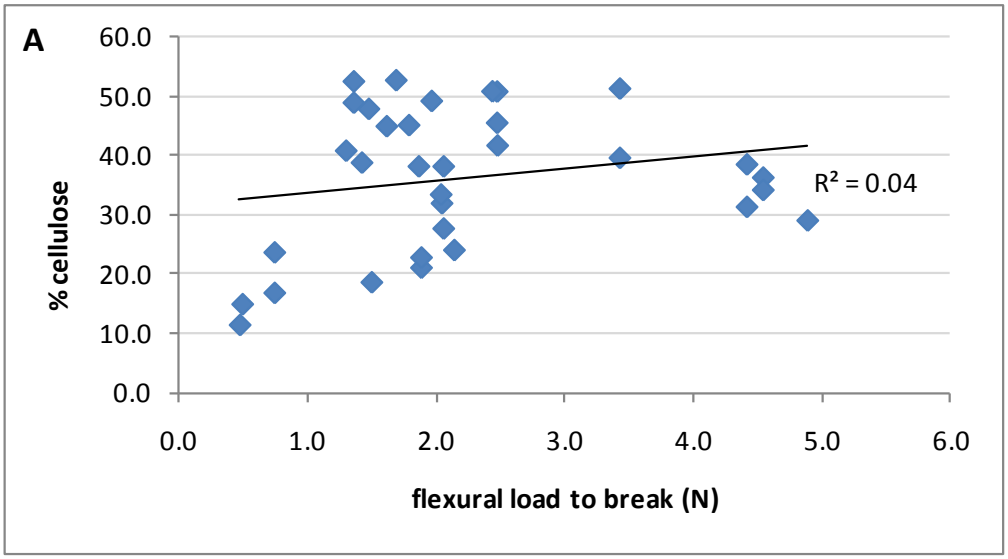
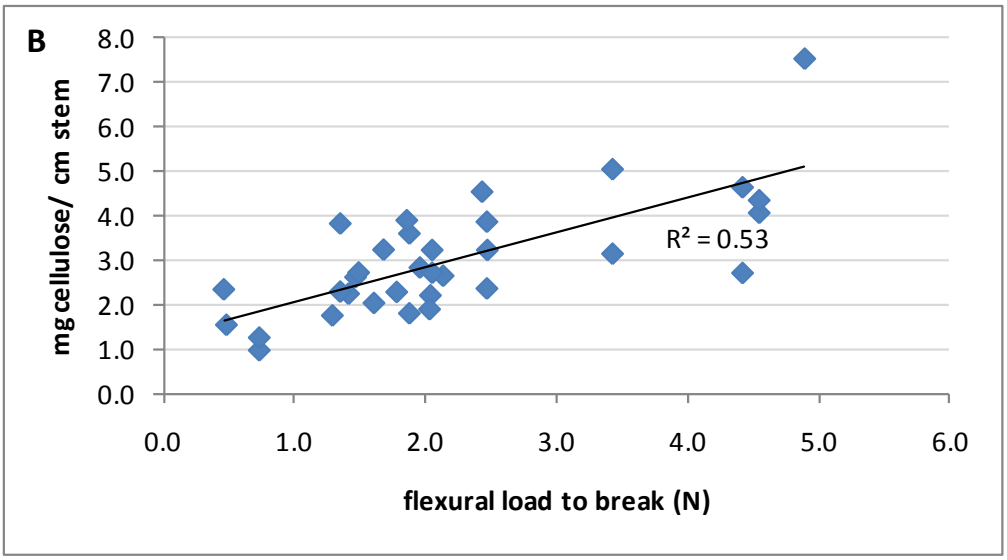
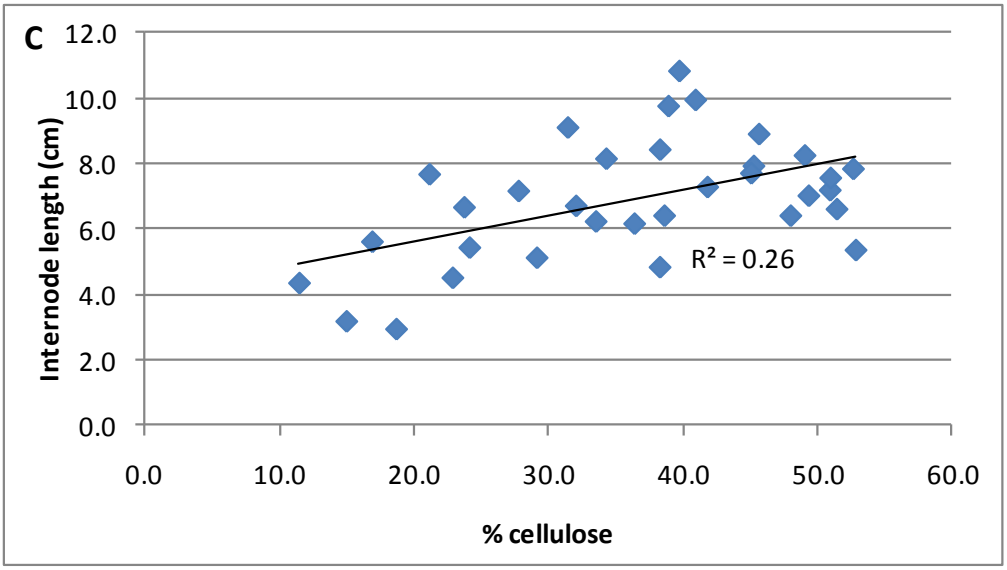


Figure 5-18: Graphs of HvCEL1 T_1 plants showing flexural load to break, % cellulose, mg cellulose per cm stem and stem diameter.

A: Flexural load to break (N), **B:** % cellulose **C:** mg cellulose per cm stem and **D:** stem diameter (mm). These data are the average results for the three tallest stems in each plant. The measurements were conducted on mature stem material taken from the second internode below the peduncle. Control - null segregants, TDNA suppr - hygromycin gene -ve, *HvCEL1* transgene +ve


A
% cellulose
flexural load to break (N)
R² = 0.04
0.0
10.0
20.0
30.0
40.0
50.0
60.0
0.0
1.0
2.0
3.0
4.0
5.0
6.0


B
mg cellulose/ cm stem
flexural load to break (N)
R² = 0.53
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
0.0
1.0
2.0
3.0
4.0
5.0
6.0


C
Internode length (cm)
% cellulose
R² = 0.26
0.0
2.0
4.0
6.0
8.0
10.0
12.0
0.0
10.0
20.0
30.0
40.0
50.0
60.0

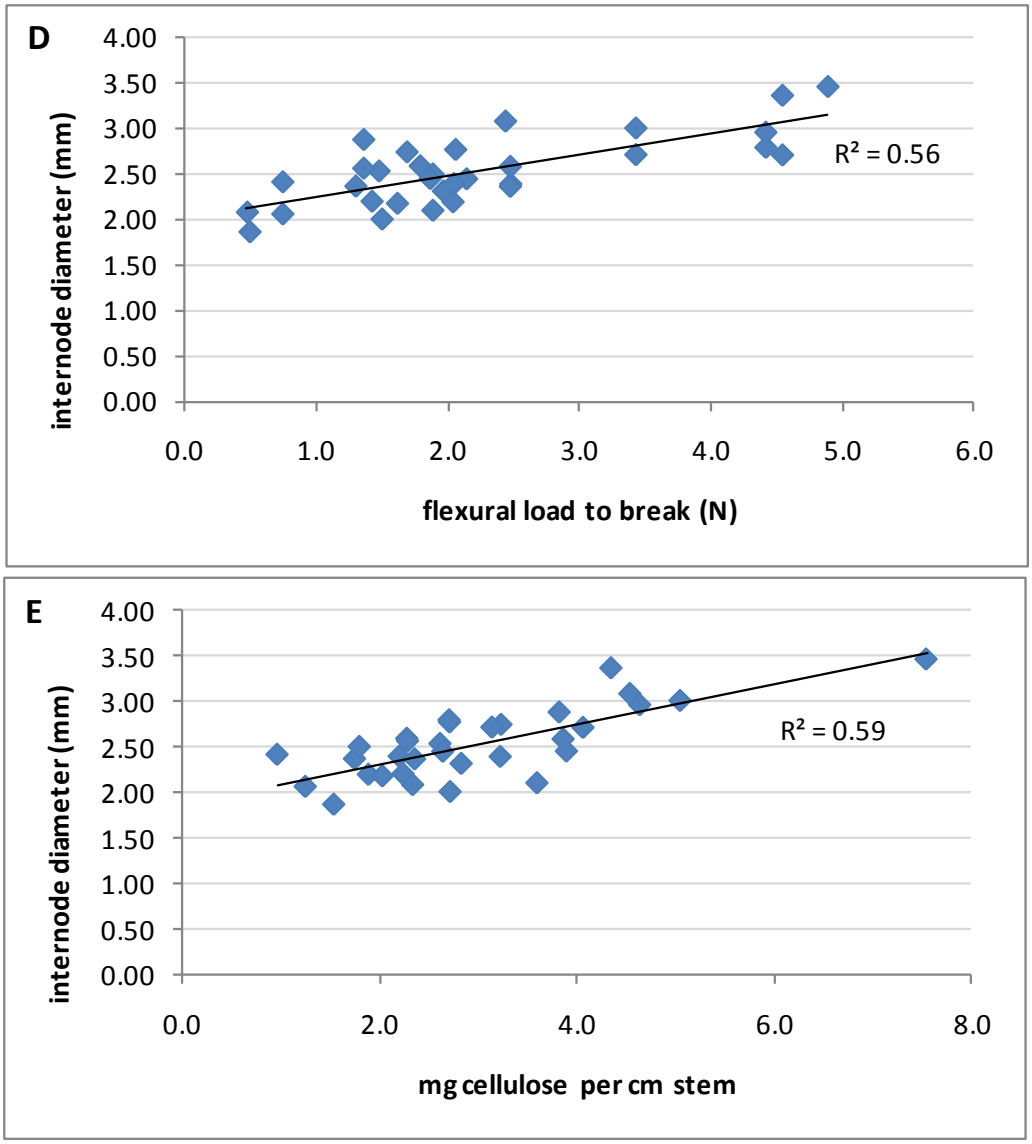


Figure 5-19: Correlations between various stem parameters measured on the HvCEL1 T_1 transgenic plants.

Correlations between A: average flexural load to break and % cellulose, B: average flexural load to break and mg cellulose per cm of stem, C: % cellulose and internode length, D: flexural load to break and stem diameter, E: mg cellulose per cm stem and internode diameter. A poor correlation exists between stem strength and % cellulose. The correlation of stem strength with mg of cellulose per cm stem is much stronger however not as good as that seen with the T0 plants. The correlation between internode length and % cellulose is low, although a positive trend is evident. Stem diameter correlates similarly with stem diameter and mg cellulose per cm stem. The outlier in B and E is plant 14-3

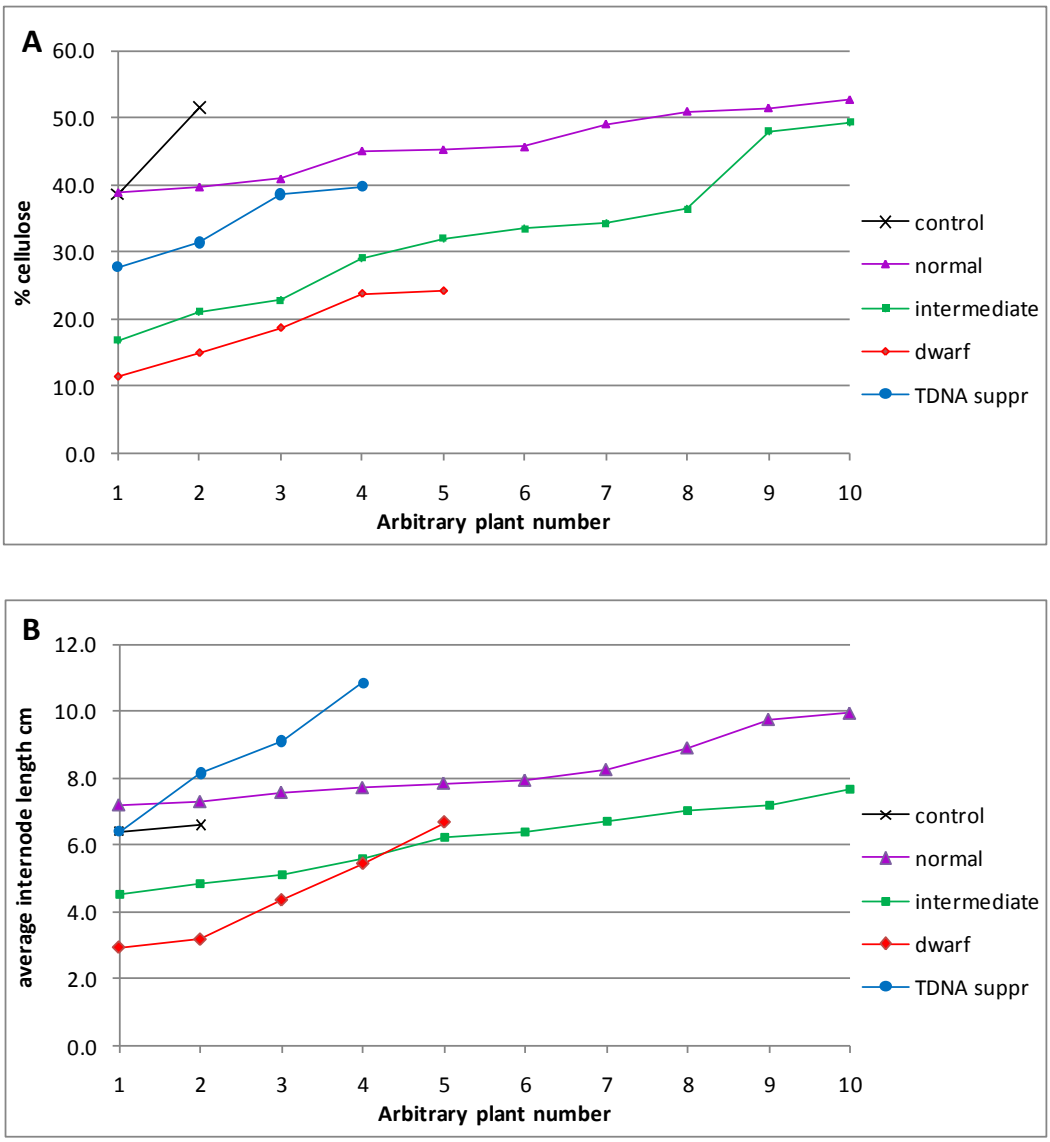


Figure 5-20: *HvCEL1* G235 T_1 plants showing cellulose levels and internode length with respect to phenotype.

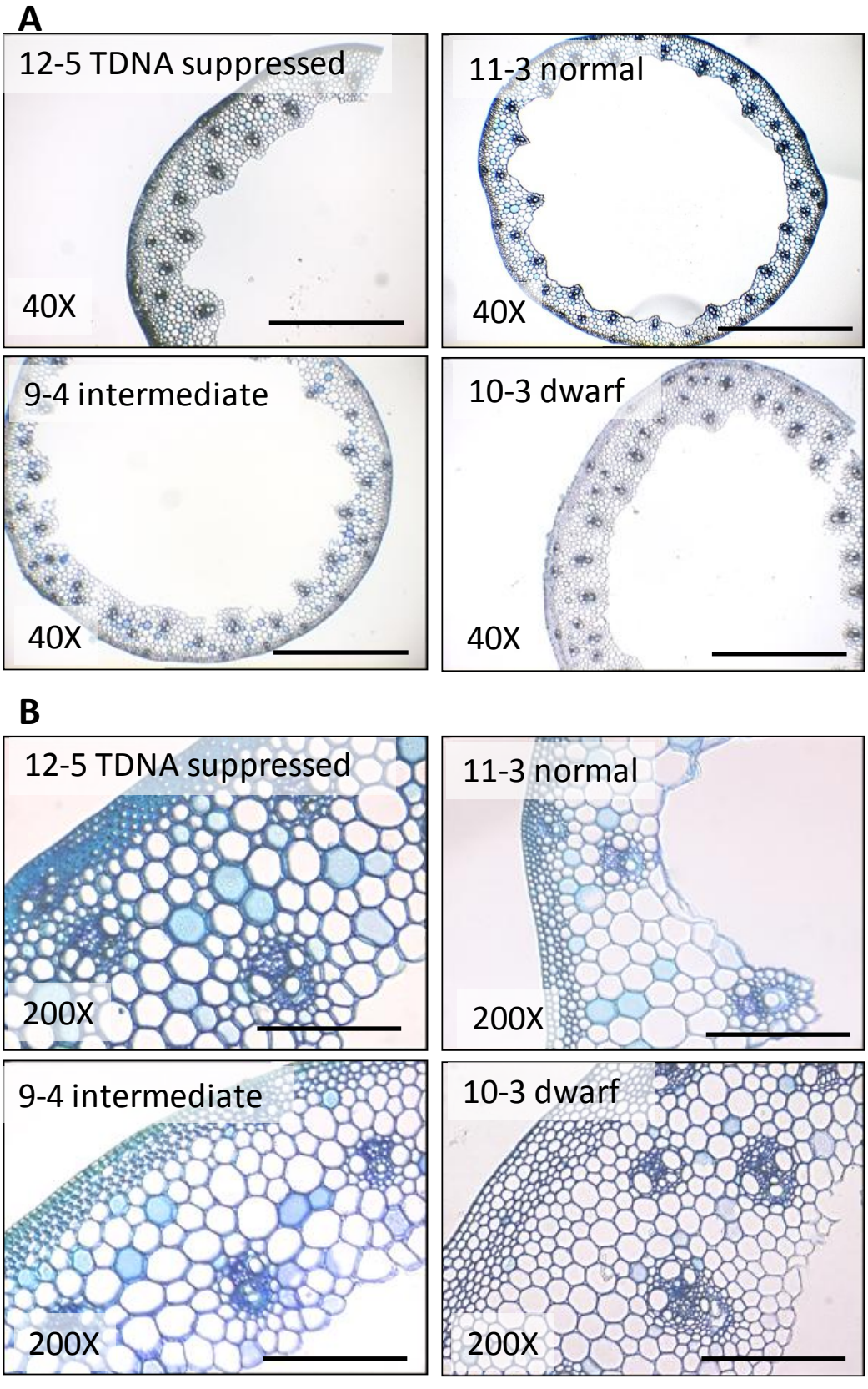
The A: % cellulose and B: average internode length has been sorted according to phenotype. The graphs highlight that differences are observed between the different phenotypes but also show that there is overlap of the phenotypes. Cellulose levels clearly separate the normal, TDNA suppressed and control phenotypes from the dwarf plants with the intermediate phenotype spanning the gap between them. Average internode length less clearly illustrates a similar trend, with the intermediate phenotype plants spanning the gap between the dwarf plants and the controls, TDNA suppressed and normal phenotype plants.

diameter and mg cellulose per cm stem (Figure 5-19 E). Because plant height and percent cellulose was a strong indicator of phenotype, the data was regrouped according to phenotype for these parameters (Figure 5-20). These graphs illustrate both the separation and overlap of the phenotypes. Although clearer in Graph A showing the cellulose levels, in both cases the dwarf phenotype was clearly separate from the control, normal and transgene suppressed phenotypes with the two groups spanned by the intermediate phenotype plants.

On the basis of the phenotype variation, and the relationship of phenotype with cellulose levels several plants were selected for microscopic analysis of their stems to ascertain if there were any obvious changes in cell morphology.

5.3.3.4. Bright field microscopy of stems of T_1 plants

Microscopy was performed on 12 plants including 12-1, 12-5, 4-5, 9-4, 9-5, 10-2, 10-3, 10-4, 11-1, 14-1, 14-3 and 14-5. This analysis does not include any of the control plants as at this time it was not known that four of the six putative null segregants were producing transgene transcript and probably contained the TDNA. As a result, the analysis actually contains two of the suppressed plants, 12-1 and 12-5. Cross sections of the second internode below the flag leaf of the stems were examined after sectioning and staining with toluidine blue. A representative selection of the results are presented in Figure 5-21 A, B and C. The results of all plants that underwent stem cross-sectioning and staining with toluidine blue are presented in Appendix H. The toluidine blue staining indicated that cells, including xylem vessels (Figure 5-21C), appeared to have a normal anatomy, although it is not possible from this study to determine if the chemical structure of the cell wall was normal. Some differences existed between stems with regard to the number of cells in the cross- section of the stem. The dwarf and intermediate plants had more cells across the radial diameter of the stem (Figure 5-21 B). The stems of the TDNA suppressed plants had 13-15 cells in cross-section, while the dwarf plant, 10-3, had 23 cells across the radius of the stem.


A
12-5 TDNA suppressed
40X
11-3 normal
40X
9-4 intermediate
40X
10-3 dwarf
40X
B
12-5 TDNA suppressed
200X
11-3 normal
200X
9-4 intermediate
200X
10-3 dwarf
200X

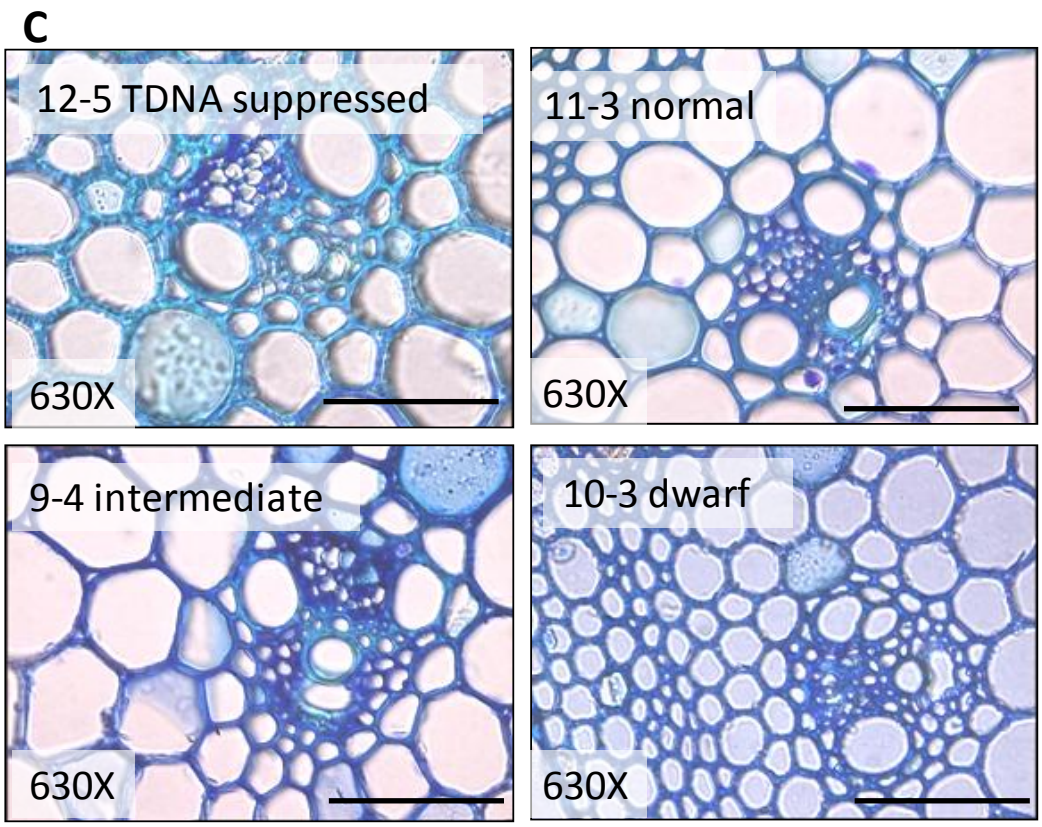


Figure 5-21: Bright field microscopy of Toluidine Blue stained cross-sections of the second internode below the peduncle of the T_1 barley plants.

(Pages 16 and 17) A: X40 magnification of cross-section of whole stems, B: X200 magnification of cross-section of whole stems, C: X630 magnification of cross-section of vascular bundles and surrounding parenchyma cells. In panels A and B differences were noted in the width of the stem wall where the dwarf stems contained a larger number of smaller cells in both the epidermis and parenchyma compared with the other phenotypes or the wildtype (Figure 5-22). The normal transgenic plant had fewer cells across the width of the stem, which was narrower than the wildtype plant. Compared to the wildtype stems, the stems of the transgenic lines all appeared to have a normal morphological structure. The bar at the bottom of the field of view represents as follows: 40x=1mm, 200x=0.2mm, 630x=63.5 µm

The intermediate plant, 9-4 had 12-13 cells across the radius and normal phenotype plant, 11-1, had only 8-9 cells. The fourth internode below the flag leaf, which is more mature than the second internode, was also examined in some plants to determine whether the walls of the transgenic plants were maturing in a normal fashion and showing secondary cell wall thickening. The sclerenchyma cells situated immediately below the epidermis were the most obvious cell type affected in the T1 plants (Figure 5-23). The TDNA suppressed plant, 12-5, had markedly thickened sclerenchyma cell walls, whereas the dwarf plant, 10-3, showed very little thickening of the cell wall in the sclerenchyma layer. The intermediate plant, 9-4 showed some secondary thickening and the normal transgenic plant, 11-1, showed thickening of the sclerenchyma, similar to that of the TDNA suppressed plant.

5.3.3.5. Epidermal glue impressions

A closer examination of the external surface of the plant stems was undertaken to further explore the variation in internode length. The aim was to investigate whether reduced internode lengths in the dwarf plants were the result of fewer cells, or shorter cells along the length of the internode. For this purpose, glue impressions of the outer surface of the stems were prepared. Figure 5-24 shows a sample of the resultant epidermal impressions obtained by this method. A summary of the results of this analysis are presented in Table 5-6 and Figure 5-25 A, B and C. It was not possible to draw clear conclusions with regard to this data, due to the small numbers of plants analysed. However, it was possible to provide some broad generalisations on correlative trends, with plant 14-3 being an exception. Plant 14-3 was somewhat anomalous in that it had epidermal cells that were 50% longer than the control cells and with less than half the number of cells per internode. This plant was also measured as having high stem strength and mg cellulose per cm stem. Pearson correlation coefficient (r) analyses were performed on these results and are found in Table 5-7. In these plants, a strong relationship was found for percent cellulose with internode length (r = 0.86) and

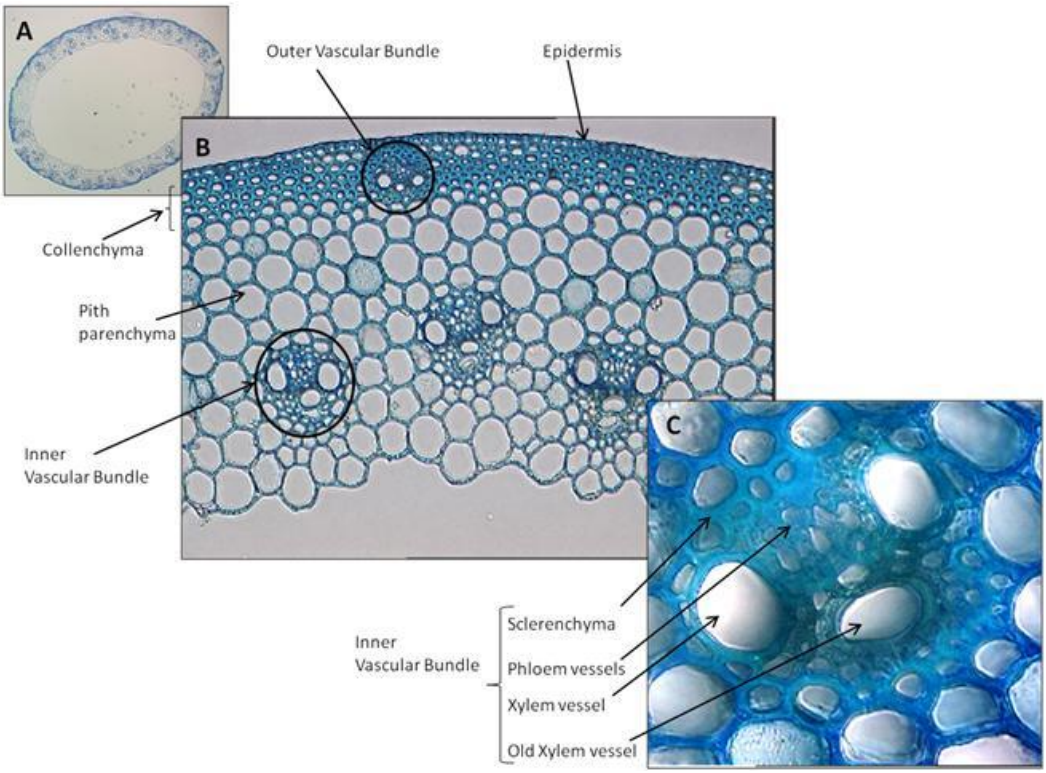


Figure 5-22: Toluidine blue staining of a barley stem.

As presented in Chapter 1, this figure illustrates the salient features of the barley stem at magnification A. X40, B. X200 and C. X640.

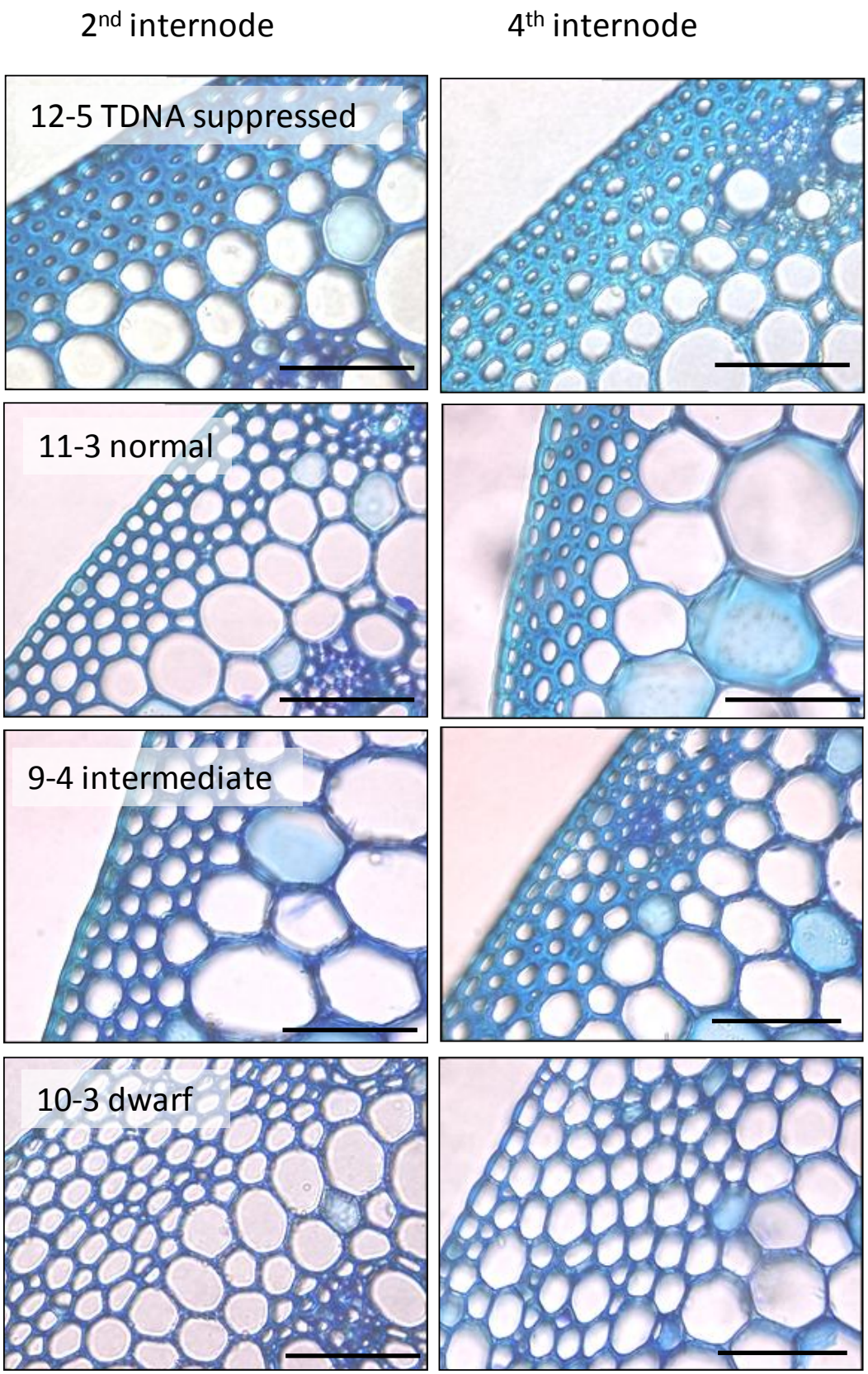

2nd internode
4th internode
12-5 TDNA suppressed
11-3 normal
9-4 intermediate
10-3 dwarf

Figure 5-23: Toluidine blue staining of stem cross-sections highlighting sclerenchymal tissues in the T_1 barley plants.

Cross sections were taken from the 2nd and 4th internodes below the peduncle of the same stem. The stems of the different plants were not necessarily at the same stage of development, however, the aim was to determine if secondary thickening was occurring over time in the transgenic plants, and in particular the dwarf and intermediate phenotypes. The dwarf plant (10-3) failed to produce noticeable secondary thickening in the sclerenchymal cell walls, whereas the normal (11-3) and intermediate (9-4) phenotypes showed secondary thickening similar to the TDNA suppressed plant (12-5). The bar at the bottom of the field of view represents 63.5 µm

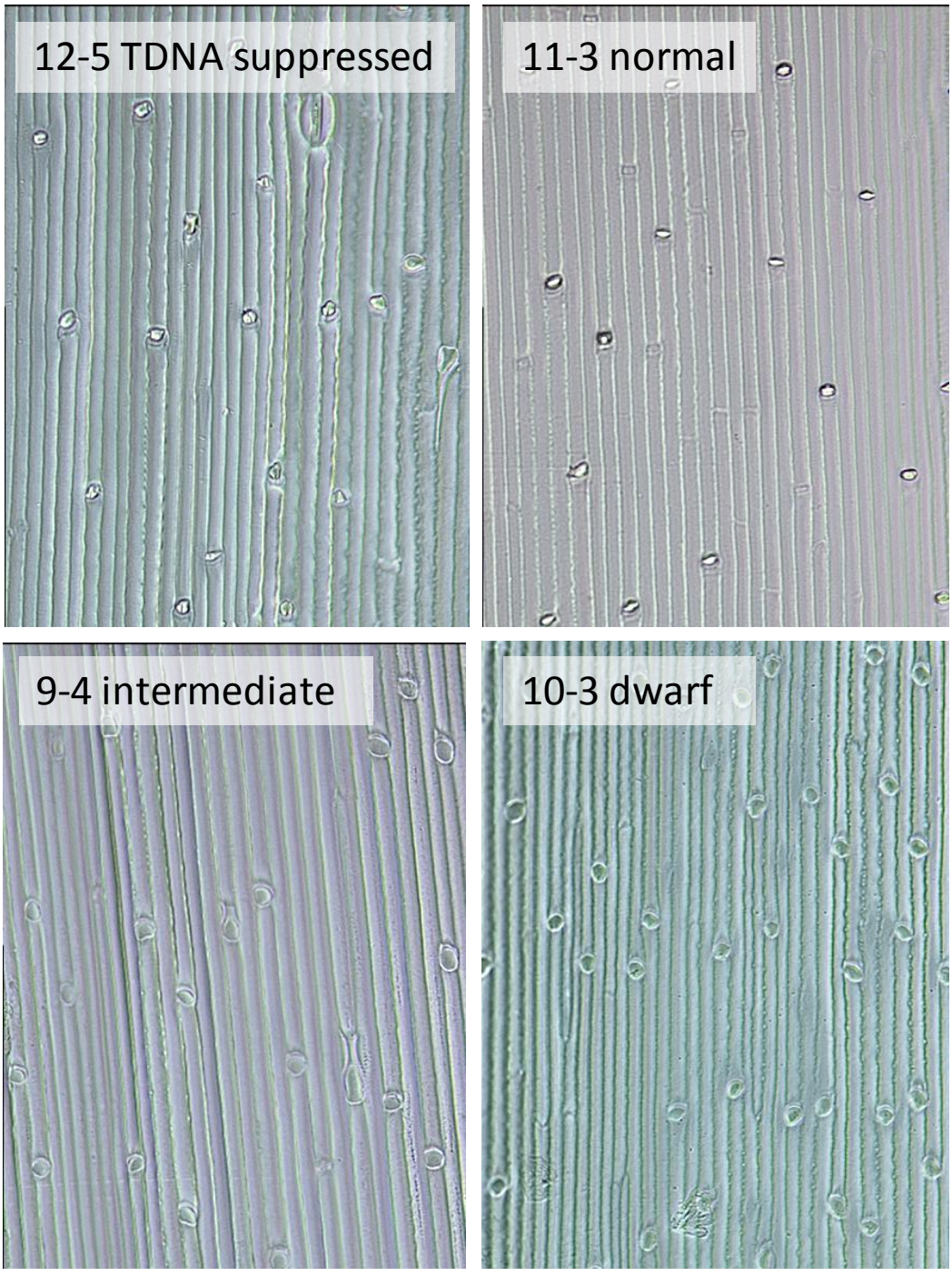


Figure 5-24: Examples of super-glue impressions of the external surface of stem internode of T_1 barley plants.

These images were taken using differential interference contrast microscopy. 12-4 represents a plant in which the TDNA was suppressed but had a normal phenotype, 11-3 displayed a normal phenotype, 9-4 showed an intermediate phenotype and 10-3 was typical of the dwarf phenotype.

Table 5-6: Epidermal cell and internode parameters of selected *HvCEL1* G235 T_1 plants.

The table presents data derived from analysis of the super-glue impressions of the stems of the second internode below the peduncle. It includes % cellulose, flexural load and mg cellulose/cm stem derived from the three strongest stems and presented in Table 5-5 and Figure 5-18. The average length and width of epidermal cells in the internode and the estimated number of epidermal cells required to span the internode longitudinally, as well as the length of the internode are included. 12-1 and 12-5 were TDNA suppressed plants, 4-5 and 11-3 are phenotypically normal transgenic plants, 9-4, 10-2, 14-1, 14-3 and 14-5 are representative of plants with an intermediate phenotype and 9-5, 10-3 and 10-4 are dwarf plants. These results indicate that both the average cell length and the number of cells spanning the internode are reduced in the plants with short internodes. Plant 14-3 was atypical in that although it had an intermediate phenotype, it had a 50% greater epidermal cell length and less than half the number of cells along the internode compared to the control plant

.

plant	phenotype	% cellulose	average cell length (µm)	internode length (cm)	cells per internode	avg flexure load (N)	mg cellulose per cm	cell width (µm)
12-1	TDNA suppr	39.1	661	9	1407	3.4	3.1	15
12-5	TDNA suppr	38.6	696	8	12113	4.4	2.7	15
9-5	dwarf	23.7	352	3	9884	0.7	1.2	15
10-3	dwarf	11.4	366	3	8477	0.5	2.3	13
10-4	dwarf	15.0	480	3	6945	0.5	1.5	19
9-4	intermediate	16.9	611	6	9178	2.0	1.4	17
10-2	intermediate	32.1	575	7	11671	2.0	2.2	15
14-1	intermediate	22.9	580	4	7713	1.9	3.6	19
14-3	intermediate	29.1	1012	5	5344	4.9	7.5	21
14-5	intermediate	38.3	677	7	9769	4.5	3.2	19
4-5	normal	51.0	496	7	14507	2.5	3.9	14
11-3	normal	49.1	580	9	16016	1.4	2.3	15

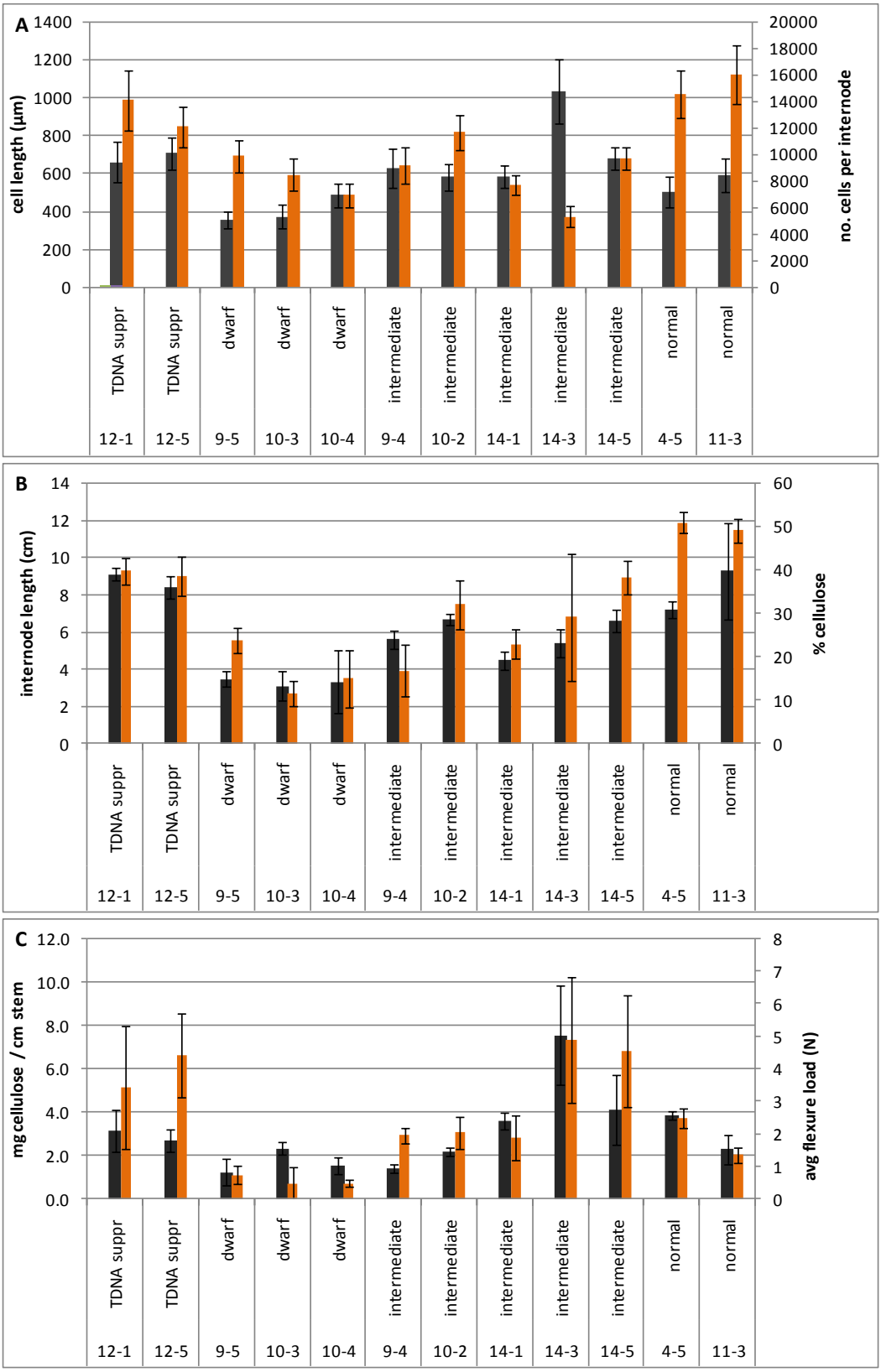

A
cell length (µm)
no. cells per internode
B
internode length (cm)
% cellulose
C
mg cellulose / cm stem
avg flexure load (N)
TDNA suppr
dwarf
intermediate
normal
12-1
12-5
9-5
10-3
10-4
9-4
10-2
14-1
14-3
14-5
4-5
11-3

Figure 5-25: Graphed results of measured parameters for *HvCEL1* T_1 plants that were selected for further analysis**.**

A: shows cell length and number of cells per length of internode. B: cells per internode and % cellulose. C: mg cellulose per cm stem with average flexural load.

cells per internode (r = 0.79), and for the average epidermal cell length with flexural load (r = 0.84) and mg cellulose per cm stem (r = 0.77). In other words, plants had longer internodes because there were more cells per internode and the stems of these plants contained more cellulose. On the other hand, the stems with longer epidermal cells were stronger and contained more cellulose per cm. Care, however, should be taken in the interpretation of these results as only a small number of plants were analysed.

Table 5-7: A selected sample of *HvCEL1* transgenic plants demonstrating the relationship between measured parameters.

The table presented shows the Pearson correlation coefficients (r) between the measured parameters. These results indicate a strong co-causal relationship between several parameters. These include % cellulose with internode length and cells per internode, as well as cell length with flexural load and mg cellulose per cm stem.

	% cellulose	*average cell length (µm)*	*internode length (cm)*	*cells per internode*	*avg flexure load (N)*	*mg cellulose per cm stem*
% cellulose	1					
average cell length (µm)	0.27	1				
internode length (cm)	0.86	0.39	1			
cells per internode	0.79	-0.21	0.80	1		
avg flexure load (N)	0.48	0.84	0.52	0.01	1	
mg cellulose per cm stem	0.28	0.77	0.13	-0.26	0.66	1

5.4. DISCUSSION

Two separate transformations of barley with two GH9A (KORRIGAN) endo-(1,4)-β-glucanase genes driven by a strong constitutive promoter were performed. Analysis of one set of plants, *HvCEL5* G234, proved difficult on several counts, so efforts were concentrated on the *HvCEL1* G235 transformants.

Examination of the T_0 plants indicated some variation in phenotype. Eight plants including G235-2, G235-4, G235-7, G235-9, G235-10, G235-11, G235-13 and G235-15 appeared similar to the vector only control. G235-1, G235-3, G235-5, G235-6 and G235-8 had a large numbers of tillers but very few seed heads at maturity with the majority of heads aborting after flowering. G235-12 and G235-14 had lower numbers of tillers and larger leaves than the control plant.

Plants were analysed by QPCR for relative transcript levels of the TDNA using primers specific for the *HvCEL1* transgene and hygromycin. Except for G235-2 and G235-7, relative levels of the transgene were between 2×10^5 and 4×10^6 per µl normalised cDNA and for hygromycin, transcript levels were between 2×10^3 and 2×10^5. Transcript levels for G235-2 and G235-7 were much lower, at less than 3×10^3 and less than 100 transcripts per µl normalised cDNA, respectively for both the transgene and hygromycin. The Southern blot analysis indicated that all but four plants, G235-8, G235-10, G235-12 and G235-14, contained vector inserts at more than one locus. For example, G235-2 had three and G235-7 at least five, TDNA loci inserted their genomes. As mentioned previously, both of these plants failed to produce more than background levels of transgene transcript, which may have been the result of suppression of transcription.

Analysis of the T_0 plants for stem cellulose levels and stem strength showed that there was a good correlation between stem strength and cellulose content of the stems on the basis of mg cellulose per cm stem as has been reported in maize (Appenzeller et al., 2004) but stem strength did not appear to correlate with stem diameter as shown in barley (Berry et al., 2006).

It is possible that any plant phenotypes associated with the 15 *HvCEL1* G235 T_0 plants may, in part, be the result of the stresses involved in the transformation process (Yahraus et al., 1995; reviewed in Cassells and Curry, 2001). Transformation associated stress may elicit the release of reactive oxygen species (ROS) such as hydrogen peroxide, superoxide radicals and hydroxyl radicals, which lead to damage to DNA with improper repair, release of metal ions with a subsequent cycle of continuous ROS production or damage to enzymes and cellular compartments. The harmful effects of ROS may be permanent or reparable. To reduce the possibility that plants permanently affected were not considered in the analysis, T_1 generation plants were grown. The process of selecting the six T_1 generation lines is outlined in the Results section 4.3.3.

Compared to the T_0 plants, the T_1 generation showed a marked increase in phenotypic variation. The transgenic lines produced plants with a continuum of phenotypes that varied from normal and bushy in appearance to dwarfed and sparse (Figure 5-16). Transcript levels of the *HvCEL1* transgene were also variable and ranged between 2.7×10^3 and 14×10^6 transcripts per µl normalised cDNA. A trend in the levels of cellulose was also observed between phenotypes (Figure 5-20 A). Several of the dwarfed and intermediate phenotype plants showed cellulose levels of less than one third that of the normal phenotype plants and the controls. The normal plants had average cellulose levels greater than the control plants, however, this difference was not statistically significant.

The observed differences in phenotype prompted a more detailed analysis of 12 T_1 plants representing the different phenotypes. Microscopic examination of the stems of the dwarfed plants indicated that secondary cell wall thickening of the sclerenchymal region of the stem was not progressing as effectively as that of the normal plants. The overall stem morphology of the dwarf plants, including the vascular bundles, however, appeared normal. The internode lengths of the dwarf plants were in some cases one third that of the controls. This was accounted for by the dwarf plants having average epidermal cell lengths half that of the control plants as well as a reduction in the number of epidermal cells spanning the length of the internode of between a half and a third. In other words, these plants were

dwarfed because they had fewer and shorter cells along the length of the internode compared to the control plants.

A graph illustrating variation in cellulose levels as they relate to transcript is presented in Figure 5-26. This graph shows relative transgene, hygromycin and endogenous *HvCEL1* transcript and average cellulose levels as a percent of the maximum for each parameter. In most cases and notwithstanding line 4, high levels of cellulose were associated with relatively high levels of transgene and endogenous *HvCEL1*gene transcript. On the other hand, relatively low levels of transgene and endogenous transcript were associated with very low levels of cellulose. Plants that had retained their endogenous transcript levels but had very low levels of the transgene had a normal phenotype and cellulose levels, for example, the line 4 plants, the controls and those plants designated as TDNA suppressed. Plants with an intermediate phenotype had any combination of the above.

Levels of hygromycin transcript were also quite variable. All of the plants in line 4 and the TDNA suppressed plants had low hygromycin, as did line 14. On the other hand, several plants including 9-4, 9-5, 10-3, 10-4 and 12-4 had high levels of hygromycin but low levels of transgenic and endogenous *HvCEL1* and concomitantly low levels of cellulose.

Low levels of transgene transcript in cases where the vector is present in the genome indicate that transcriptional and/or post-transcriptional gene suppression may be occurring in the dwarf and intermediate phenotype plants and perhaps several of the TDNA suppressed plants as alluded to earlier with regard to transgene transcript levels. Reference to transgene silencing was made some years after the first successful transformation of plants (Jones et al., 1987). Jones et al. (1987) found that many transgenic petunia plants lost transgene expression and that this loss was particularly common where multiple copies were inserted as inverted repeats. It has since been found that hypermethylation of the transgene and/or promoter induces transcriptional gene suppression of the

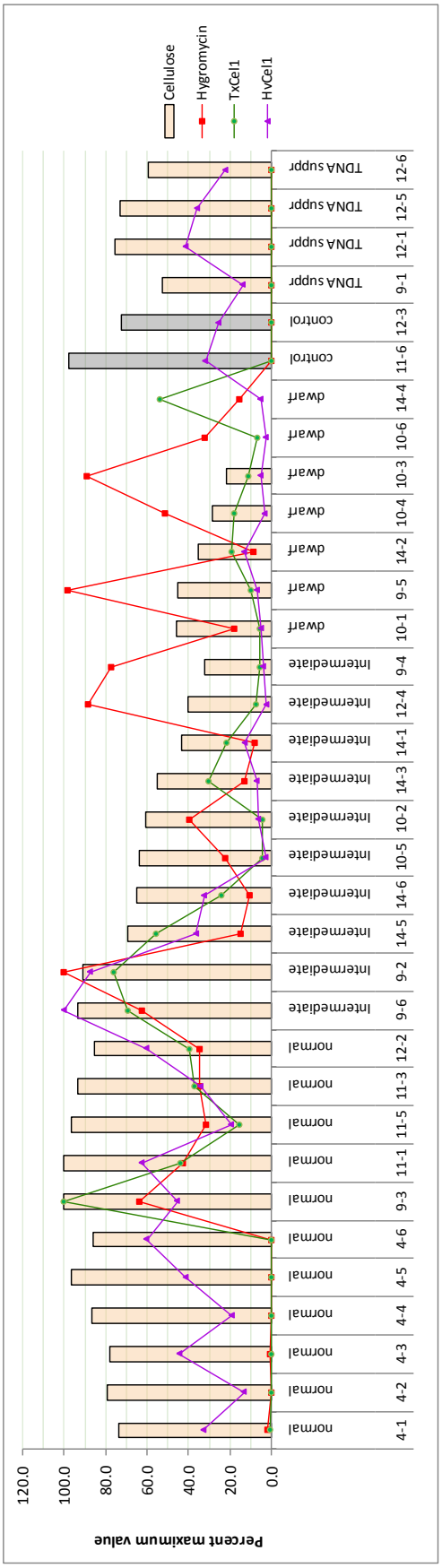

Cellulose
Hygromycin
TxCel1
HvCel1
Percent maximum value
120.0
100.0
80.0
60.0
40.0
20.0
0.0
4-1 normal
4-2 normal
4-3 normal
4-4 normal
4-5 normal
4-6 normal
9-3 normal
11-1 normal
11-5 normal
11-3 normal
12-2 normal
9-6 intermediate
9-2 intermediate
14-5 intermediate
14-6 intermediate
10-5 intermediate
10-2 intermediate
14-3 intermediate
14-1 intermediate
12-4 intermediate
9-4 intermediate
10-1 dwarf
9-5 dwarf
14-2 dwarf
10-4 dwarf
10-3 dwarf
10-6 dwarf
14-4 dwarf
11-6 control
12-3 control
9-1 TDNA suppr
12-1 TDNA suppr
12-5 TDNA suppr
12-6 TDNA suppr

Figure 5-26: Relative cellulose and transcript data for the T_1 HvCEL1 plants.

Relative transcript levels for the HvCEL1 transgene (green), endogenous HvCEL1 (purple) hygromycin (red) and cellulose (beige bars) levels have been plotted as a percent of the highest level for each parameter. Control cellulose levels are shown as grey bars.

transgene, a phenomenon commonly known as *cis*-inactivation (Meyer et al., 1993). Paramutation results because the methylation changes are retained in following generations. This epigenetic phenomenon is used by plants to control transcription of transposons (Vaucheret et al., 2001). Further research has shown that post-transcriptional suppression, which may be accompanied by loss of endogenous transcription, also occurs and is known as *trans*-inactivation (Napoli et al., 1990; Kanazawa et al., 2007). This second form of inactivation is a natural defence response of the plant to viral infections (reviewed in Dunnoyer and Voinnet, 2005; Fukunaga and Doudna, 2009). Although there are several differences between *cis*- and *trans* -inactivation of gene transcription, they have both been found to result from production of double stranded RNAs (dsRNAs). *Cis*-inactivation occurs when DNA dependent RNA polymerases produce dsRNAs (Mourrain et al., 2007; Wierzbicki et al., 2008), while RNA dependent RNA polymerases are responsible for *trans*-inactivation (Fukunaga and Doudna, 2009).

In other words, *cis*-inactivation results from production of dsRNAs from genomic DNA, whereas *trans*-inactivation results from production of dsRNAs from transcriptional RNAs. Subsequently, the dsRNAs are degraded by DICER proteins to small RNAs of 21 or 24 nts (Mourrain et al., 2007). *Cis*-inactivation results when the small RNAs associate with an ARGONAUT protein complex that also contains associated methylation proteins (Zilberman et al., 2004). The small RNAs facilitate specific binding of the ARGONAUT 4 complex to the gene and/or gene promoter in question with subsequent methylation and loss or reduction of gene transcription (Zilberman et al., 2004; Cantarel et al., 2009; Zheng et al., 2010).

Trans-inactivation occurs after transcription of the gene and results from degradation of the mRNA transcripts post-production (Mourrain et al., 2007). ARGONAUT 1 is a component of post transcriptional gene inactivation (Morel et al., 2002) as well as facilitating DNA methylation similarly to *cis*-inactivation (Morel et al., 2000). Mourrain et al. (2007) further suggest that *trans*-inactivation has a systemic component that is still to be fully elucidated.

Here, the results indicate that both *cis*-inactivation and *trans*-inactivation may have occurred. *Cis*-inactivation was suspected in plants that had reduced transcripts deriving from both the hygromycin and the *HvCEL1* transgenes without affecting the endogenous transcript levels. In other words, inactivation of the whole TDNA was probably occurring. Furthermore, the transcript reduction effect was seen in the young T_1 plants allowing for speculation that the inactivation may be epigenetic in nature. This was observed with the G235-4 line of plants. These plants all had cellulose levels similar to the control plants with low levels of transgene and hygromycin transcripts, but endogenous transcripts appeared to be unaffected. T*rans*-inactivation was suspected where endogenous transcripts were also very low. This was apparent in the T_1 progeny of plant line 10, as well as 9-4, 9-5 and 12-4. All of these plants had reduced transgene transcripts, levels of hygromycin transcript were variable and endogeneous transcript was low. These transcript levels were associated with significant reductions in stem cellulose levels. On the other hand, plants with endogenous transcript levels similar to the control plants and with higher levels of transgene transcripts generally had cellulose levels equal to or greater than the control plants. This was the case with 9-2 and 9-3, 11-1, 11-3 and 11-5 and 12-2.

It is noteworthy that in cases where the endogenous transcripts were very low, transgene transcripts were reduced but not eliminated. It has been shown that in *Arabidopsis* plants where post-transcriptional silencing was being monitored using β-glucuronidase (GUS) expression, silencing was not evident in young plants (Elmayan et al., 1998). However, as these plants matured, GUS expression was reduced, and in adult leaves was almost zero. It is likely that this phenomenon may account for the inconsistency evident between transcript levels and phenotype in the barley plants examined here.

The transcript analysis outlined in Chapter 3 indicated that the *HvCEL1* transcription was most prominent in the rapidly elongating stem internode, thus the effect of its suppression would not be evident until the stems began to elongate. It was speculated that overexpression of *HvCEL1* in barley resulted in plants with very high levels of transcript, but also plants with an inadvertent

"knockdown" phenotype that resulted from transcriptional or post-transcriptional suppression. These results support earlier claims that in *Arabidopsis* the endo-(1,4)-β-glucanase GH9A sub-family is involved in cellulose production as well as cell elongation. Therefore it is not surprising that the "knockdown" plants were dwarfed.

Dwarfing is a feature of the barley *elongation* mutants although the identity of the genes responsible has not been elucidated (Lewis et al., 2009). Several factors suggest a link between the *elongation* mutants and the *HvCEL1* dwarf plants. The *elongation* mutants all have reduced levels of cellulose with a concomitant increase in other cell wall constituents. Radial swelling of cells in the leaves and roots is seen with most of these mutants. The phenotypic changes seen with the barley *elongation* mutant plants is also a feature of several *Arabidopsis kor* endo-(1,4)-β-glucanase mutants (Nicol et al., 1998; Zuo et al., 2000; Lane et al., 2001). Radial root swelling suggests possible interruption to the transverse cellulose microfibril orientation with loss of anisotropy during cell elongation.

Anisotropic cell expansion or as is the case here, elongation, has been shown to relate to normal microtubule function (reveiwed in Bisgrove, 2008). Furthermore, microtubule orientation is closely linked to cellulose synthesis (Paredez et al., 2008). An *Arabidopsis* endo-(1,4)-β-glucanase KOR1 mutant, *kor1-3*, is implicated in loss of anisotropic cell expansion. The *kor1-3* mutant is associated with reduction in the velocity of the CESA complex along the plasma membrane as it manufactures cellulose. The *kor1-3* mutant also results in loss of microtubule orientation in affected plants. The implication is that KOR1 has a role in microtubule orientation. It is recognised that the GH9A endo-(1,4)-β-glucanase enzymes most likely span the plasma membrane in close proximity to the CESA hexamer where they may provide essential hydrolytic activity in the cellulose synthesis production line. The two polyproline rich regions located at the COOH-terminal side of the plasma membrane in the cell wall, one of which is adjacent to the transmembrane domain and the second one close to the stop codon may, as suggested by some authors, associate with other proteins or

carbohydrate moieties (Kay et al., 2000; Master et al., 2004). The KOR endo-(1,4)-β-glucanases also have a cytosolic tail. If, as is suggested by Paredez et al. (2008), cellulose synthesis is aiding in microtubule alignment, KOR endo-(1,4)-β-glucanases may provide a physical link between the cellulose synthesis machinery and the microtubule alignment proteins. Although loss of anisotropy was not apparent here, the epidermal cells of some intermediate and dwarfed plants showing both *cis*- and *trans*-inactivation of transcription did tend to have cells that were somewhat wider than the controls. This suggests that loss of *HvCEL1* function with associated reduction in elongation may have some effect on microfibril orientation.

The *elongation* mutants were also of interest as some of these mutants produced plants with an increase in the number of cell layers on the leaf surfaces, compared to the wildtype plants (Lewis et al., 2009). Here, the *HvCEL1* transgenic plants showed a relationship between phenotype and numbers of cells in the cross-sectional, or radial, orientation of the stem. Several of the dwarf and intermediate plants had significantly larger numbers of cells in a radial dimension across the stem wall, compared to the controls. This was counterbalanced by the normal phenotype plants, 4-5 and 11-1 that had fewer cells across the stem wall than the control plants. The reduction in internode length was accompanied by a reduction in the number of epidermal cells along the internode as well as these cells being shorter and somewhat wider in diameter. The reduction in cell numbers longitudinally was compensated for by more cells being stacked across the wall of the internode. This suggests that a similar number of cells were being produced by the meristem in all of the plants irrespective of their phenotype. It may be that reduction of *HvCEL1* transcript is slowing cellulose production as is suggested by Paradez et al. (2008), which may be slowing the elongation process, with resultant accumulation of the cells produced in the meristem in the stem wall.

It has been suggested that a reduction in cellulose levels results in a compensatory increase in the production of other cell wall constituents (Zuo et al., 2000; His et al., 2001; Zhou et al., 2006; Lewis et al., 2009). However, it is

also possible that other cell wall components are still being produced at normal levels and because cell wall constituents are usually measured in relative amounts, a reduction in cellulose will, arithmetically, lead to a relative increase in the levels of other cell wall components.

In this project, a full investigation of the cell wall polysaccharide composition was not undertaken. It was noted, however, that stem structure, including vascular bundles, was morphologically similar between the different phenotypes and the control plants. However, it was observed that, compared to the control plants, some of the dwarf plants failed to produce thickened secondary cell walls, particularly noticeable in the sclerenchyma layer of the stem. The intermediate phenotype plants had more substantial cell walls than the dwarves, although somewhat thinner than the controls. On the other hand, the normal phenotype plants showed similar cell wall thickness to the controls.

These results invite further investigation, and Takahashi et al. (2009) report that overexpression of the aspen *PttCel9A1*, an orthologue of the *Arabidopsis* KOR1 endo-(1,4)-β-glucanase, in *Arabidopsis* does not affect total cellulose levels in the plant but causes increased levels of non-crystalline cellulose and a concomitant reduction in crystalline cellulose. These authors conclude that activity of KOR1 "facilitates cellulose biosynthesis in a way that increases the amount of non-crystalline cellulose". Here, it was not possible to speculate on this finding because only total cellulose levels were measured.

A better understanding of the role of endo-(1,4)-β-glucanases in barley and the response of the cells to perturbation of cell wall synthesis could be effected if further analyses were performed. It may prove interesting to compare the cell walls of those plants with the overexpression transgene with those showing the knockdown phenotype. A methylation analysis of cellulose can be used to determine the relative levels of non-crystalline to crystalline cellulose (Needs and Selvendran, 1993; Burton et al., 2010). Also, a recently developed analysis tool, namely comprehensive microarray polymer profiling (CoMPP), would aid in determination of relative levels of cell wall glycans (Moller et al., 2007).

CoMPP can be used to analyse all cell wall constituents for which monoclonal antibodies or carbohydrate binding modules are available, such as arabinoxylans and (1,3;1,4)-β-D-glucans. Arabinoxylans provide a link between cellulose crystalline microfibrils and lignin, and lignification provides an important means of increasing cell wall strength. Therefore an analysis of relative lignin levels in the different phenotypes may also prove interesting.

It was not determined if the continuum observed between the different phenotypes was a reflection of protein expression levels. For this reason, it would have been greatly beneficial in the analysis of these plants if either enzyme activity or quantity of transgene protein expressed had been determined. Takahashi et al. (2009) have developed an *in muro* (1,4)-β-glucanase enzyme activity assay that releases a fluorescent resorufin anion from resorufinyl β-D-cellobioside and the result is viewed using fluorescent microscopy. This assay gives an indication of whether the enzyme has hydrolytic activity as well as showing where in the plant this activity is occurring. Extraction of protein from plant tissues and *in vitro* activity assays could also be performed using the Somogyi-Nelson photometric assay (Nelson, 1944) or viscometric analysis (Vlasenko et al., 1998). Alternatively, an analysis of the presence of enzyme could be performed by extracting the protein from the tissues, performing a western blot and probing for the presence of the protein using an anti-HvCEL1 antibody. This was not performed because no barley HvCEL1 antibody is currently available. An alternative would be to add a 6xHis tag to the transgene prior to transformation into barley. It was, however, possible that this tag may affect folding and/or protein activity.

Finally, inability to select primary tillers for stem strength and cellulose analyses in all plants may have led to some inconsistency in the results, with large standard deviations. One of the problems associated with measuring stem strength and cellulose from different plants is the necessity to harvest stems at similar stages of development, in this case, at grain maturity. Less inconsistency is also achieved if stems have undergone the same maturation process. During grain maturation, energy in the form of polysaccharides is transported from the

leaves and stems to the developing grain (Liebhardt et al., 1968; Jampatong, 2000; Baba et al., 2001; Martin, 2004). This reduces stem strength and may also alter the relative amounts of the different cell wall polysaccharides, including cellulose. Therefore those stems with mature grain heads may have relatively weaker stems and higher relative levels of cellulose due to the mobilisation of polysaccharides to the grain. Experience has shown that stems from primary tillers are stronger and contain higher levels of cellulose than their secondary counterparts possibly because they have had a longer period of time to build up levels of stored polysaccharides. In this project, primary and secondary stems were labelled at flowering. Stems were collected at similar stages of development, that is, at grain maturity, however, inconsistency in stem material was encountered because a number of grain heads aborted and failed to produce mature grain.

Some of the plants that were exhibiting overexpression of the *HvCEL1* transgene also showed levels of cellulose that were greater than the control plants. Increasing cellulose as a means of increasing stem strength had been the reasoning behind overexpression of the *HvCEL1* cDNA (Appenzeller et al., 2004). The fact that stem strength did not necessarily increase in the overexpressing plants may be related to inconsistency as outlined above or perhaps reflects the problems that may be encountered when interfering with a finely tuned and complex system such as that of cell wall synthesis. Inactivation of transcription or transcripts also may reflect the plants overarching need to retain cell wall integrity. Attempts to increase cellulose levels with overexpression of CESAs in barley have encountered similar problems including transcript inactivation and loss of stem strength (pers. comm., Ashley Tan, ACPFG, University of Adelaide). This theme was also explored recently where aspen plants overexpressing a secondary cell wall *CESA* also resulted in transgene silencing and reduced cellulose levels with concomitant reductions in stem strength (Joshi et al., 2011).

5.5. SUMMARY

Two separate transformations of barley with the *HvCEL1* and *HvCEL5* GH9A endo-(1,4)-β-glucanase gene cDNAs were performed. Whilst analysis of the *HvCEL5* (G234) transformation met with some technical difficulties, six lines from the *HvCEL1* (G235) transformation were successfully grown through to the T_1 generation. Null segregants, defined by three negative PCR reactions using genomic DNA and hygromycin primers were used as the control plants. The resultant plants showed a range of phenotypes from small dwarfed plants through intermediate sized plant to tall strong plants (normal) similar to the control plants. Stem cellulose levels also showed significant variation that related to phenotype. Dwarf plants had relatively low levels of cellulose and normal plants had high levels of cellulose, comparable or higher than the controls, and the intermediate plants showed cellulose levels that spanned the two extremes.

QPCR analysis of transgene and endogenous gene transcripts indicated that there may be *cis*-inactivation and *trans*-inactivation occurring in some of these plants that caused reduction of transgene and in some cases endogenous gene transcript levels in several plants. Several plants exhibited suppression of the TDNA and showed a phenotype similar to the normal phenotype plants. On the other hand, the dwarf phenotype appeared to be the result of *trans*-inactivation of transcription with the resultant loss of endogenous gene transcript effectively producing "knockdown" of *HvCEL1* expression. The internodes of these plants were up to three times shorter with cellulose levels, in some cases, less than a third that of the control plants. The number and length of epidermal cells in the internodes of these plants were reduced. As the number of cells across the radial axis of the internodes was in some cases double that of the controls, it was suggested that reduction in *HvCEL1* transcript levels leads to reduced cell elongation and an accumulation of cells across the diameter of the stem wall. In contrast, the transgenic plants that were showing overexpression of transgene transcript had levels of cellulose that were similar or more than the control plants with a similar or bushy appearance. However, high cellulose levels did not result

in significantly higher stem strengths, and may indicate that increasing stem strength relies on more than simply increasing cellulose levels.

Chapter 6

Summary and Future Directions

Since the "Green Revolution" that began in the 1940's, both barley and maize have been subject to stem lodging. Stem lodging has resulted from increased yields with resultant heavier grain heads, which increase the load on the stem (Liebhardt et al., 1968; Jedel, 1991; Duvick, 1999; Jezowski, 2003; Berry et al., 2006).

It has been observed that stem strength is directly related to the amount of cellulose in the plant cell wall and that, in maize, increasing cellulose levels in the wall is accompanied by increases in stem strength (Heuchert et al., 1983; Appenzeller et al., 2004). In this context the hydrolase enzyme family of endo-(1,4)-β-glucanases is of interest. Although this family hydrolyses 1,4-β linkages in carbohydrate moieties, including cellulose, somewhat counter-intuitively, it is shown to be important in cellulose synthesis as revealed in mutational studies in rice and *Arabidopsis* (His et al., 2001; Lane et al., 2001; Sato et al., 2001; Szyjanowicz et al., 2004). Loss of endo-(1,4)-β-glucanase activity due to genetic insertions or mutations leads to a reduction in levels of crystalline cellulose content of up to 30% in the plant cell wall. It follows that the endo-(1,4)-β-glucanase gene family should be investigated as a possible means of altering cellulose levels and affecting stem strength. The aim of this project was, therefore, to characterise the endo-(1,4)-β-glucanase gene family in cereals, with barley and maize as the chief species for investigation.

The first step in characterisation was the identification of gene sequences for barley, maize, rice, sorghum and *Arabidopsis*. The gene sequences from barley, maize, rice and sorghum were subsequently used for a phylogenetic analysis. A study of transcript patterns in a variety of maize and barley tissues that included the stem internode during its growth and development was also undertaken. Further experiments were performed that analysed the effects of overexpression of GH9A cDNAs in barley and heterologous expression of endo-(1,4)-β-glucanases in microbial systems.

Characterising the phylogeny of the endo-(1,4)-β-glucanase gene family has added to the current knowledge of this gene family, particularly in cereals. From

this work it was possible to nominate orthologues from barley, maize, rice and sorghum. A clade that was not present in *Arabidopsis* containing four or five genes each from barley, maize, sorghum and rice was revealed. The evolutionary analysis showed that this clade evolved after separation of the cereals from dicotyledonous plants, and so is novel to the cereals. From this analysis it was also possible to identify a number of maize homoeologues that have been retained since allotetraploidisation of the maize genome. The evolutionary evidence also suggested that changes in protein sequence between orthologues are minimal and evolutionary pressure has been towards gene conservation.

Although a barley genome gene sequence scaffold is now available (Mayer et al., 2011), the barley genome sequence was not available when the analysis described in this thesis was performed. Although further advanced than barley, sequencing of the maize genome was also a "work in progress" during this project. Significant recent advances have been made, and although some BACs are still not fully sequenced and assembled, in the last few months, anchoring of most BACs to the genome has been accomplished. Another recent addition to the list of sequenced genomes is the cereal model species *Brachypodium distachyon*. While this project was in progress, the *Brachypodium* sequencing efforts had just commenced and comprised only short reads that proved difficult to piece together. In recent times, however, annotated gene sequences have become available and will provide a useful resource for future research on the endo-(1,4)-β-glucanase gene family.

Sequence gathered for the phylogenetic analysis was used in design of primers for the transcript analyses of maize and barley genes. Comparative analysis of the transcript patterns provided further proof of evolutionary conservation of this gene family. In essence, six endo-(1,4)-β-glucanase genes in barley showed similar transcript patterns to their maize orthologues. Co-transcription of several genes provided evidence in barley that multiple endo-(1,4)-β-glucanases may be acting in concert during cell wall synthesis. Genes associated with wall synthesis included members from the GH9A sub-family as well as a gene that was from clade B1 (Figure 2-7), whose members are phylogenetically related to the GH9A subfamily of genes. Genes associated with secondary cell wall synthesis also

included genes belonging to clade B3 (Figure 2-7), identified as novel to cereals in the phylogenetic analysis. Furthermore co-transcription between several endo-(1,4)-β-glucanase genes and other genes known to be important in cell wall synthesis were elucidated.

Future work could include analysis of microarray data to determine if any patterns exist between transcription factors and those cell wall synthesis genes that are co-transcribed. Additional proof of co-transcription as well as tissue specificity could be performed with the use of *in situ* hybridization techniques. A transcript analysis provides information regarding gene activity, but evidence of mRNA translation into a functional protein would provide additional proof. To enable identification of endo-(1,4)-β-glucanase expression in cereals by western blot analysis, specific antibodies are required. Due to amino acid sequence similarities in this gene family the successful production of specific antibodies may be difficult. A valid alternative approach is zymography (Béguin, 1983). This assay is not specific, but if an iso-electric focusing (Ahlgren et al., 1967) approach was adopted using a 2-D gel, mass spectrometry could be employed to identify the protein(s) producing the hydrolytic activity.

The co-transcriptional patterns observed between endo-(1,4)-β-glucanases and other genes involved in cell wall synthesis, such as expansins, xyloglucan endotransglycosylases (*XET*s) and *CESA4* were used as the basis for planned *in vitro* co-expression experiments using barley cell wall preparations as the substrate. Proteins for these co-expression studies were to be obtained using heterologous expression of five different endo-(1,4)-β-glucanase constructs in both *E.coli* and *P. pastoris* expression systems. However, active and soluble expressed protein was not obtained.

For the prokaryote system a large amount of protein was produced, however it was chiefly insoluble and was contained within the inclusion bodies of the cells. Several strategies were identified that could be used in the future to improve yields of soluble protein. These included the design of several primers to produce cDNAs that have sequence variations at the 3' and 5' ends (Gräslund et al., 2008). Another option included refolding of the protein contained within the

inclusion bodies, particularly if it is already partially folded (Burgess et al., 2009).

For the *Pichia* system, very low quantities of protein were harvested and further work could include analysis of the *Pichia* cells to determine if protein was being produced but not excreted. It is also possible that protein was produced but being degraded or stored in the cell due to delays in the transcription process (Varenne et al., 1984). Codon usage that is specific to the host species may be a useful option.

Other systems used for heterologous expression include those employing baculovirus particles (Geisse et al., 1996; Kost et al., 2005; Brondyk et al., 2009). This system involves homologous recombination or site directed transposition of the DNA of the protein of interest into the modified viral DNA and amplification of the virus in one of several insect cell lines grown in suspension cell cultures. Plant heterologous expression systems have also been developed, particularly successful being use of the genus *Nicotiana*. A system developed by Icon Genetics uses *Agrobacterium tumefaciens* transformed with a viral vector (Marillonnet et al., 2005). *Agrobacterium*-mediated transformation of the *Nicotiana* spp allows for an expression system that produces large quantities of protein that are post-translationally modified *in planta*. Reported yields of 25% to 40% of total soluble protein are achieved (Marillonnet et al., 2005).

Finally, *Agrobacterium*-mediated transformation of barley was performed using two GH9A endo-(1,4)-β-glucanase cDNAs behind a 2 x 35s overexpression promoter. The T1 generation of the *HvCEL1* transformation showed a range of phenotypes, and transcript analysis indicated that some plants were overexpressing the transgene, as expected, while in others the transgene had been inactivated. Plants in which *cis*-inactivation of the transgene was occurring were not dissimilar to the control plants, whereas, *trans*-inactivation resulted in the production of knockdown plants that were severely dwarfed because both transgene and endogenous transcripts were being suppressed. Cellulose levels in these dwarfed plants were much reduced compared to the control plants. Loss of endo-(1,4)-β-glucanase activity resulted in reduction in elongation of epidermal cells and a lower number of these cells spanning the length of the internode

compared to the control plants. Also, there was an increase in the number of cells spanning the radial width of the stem internode. Conversely, plants that were overexpressing the cDNA were of a similar phenotype to the control plants and had similar or higher levels of cellulose in the stem. There was an increase in the number of epidermal cells along the length of the internode and a reduction in the number of cells across the radial width of the internode compared to the control plants. This result leads to speculation that loss of HvCEL1 activity is affecting cell growth but not cell division.

Further characterisation of both the overexpressing and low expressing lines, could include several different analyses. An examination of the methylation patterns of the *HvCEL1* transgene and endogenous gene as well as a search for small RNAs associated with the gene would provide proof, or otherwise, of gene inactivation and the cause, which is only speculative at this stage (Llave et al., 2002; Zilberman et al., 2003). The plants with inactivated transcript, although dwarfed, showed normal stem anatomy. However, because cellulose levels were so low, a more detailed investigation of the cell wall could prove interesting. Relative levels of other cell wall constituents should also be investigated and may aid in characterizing the role of the endo-(1,4)-β-glucanases in cell wall synthesis. A protein analysis, although problematic, would also be of benefit in characterising these plants. Takahashi et al. (2009)'s *in muro* glucanase activity assay could be performed, and may indicate increased or decreased levels of endo-(1,4)-β-glucanase activity.

Beyond the scope of this project but part of the wider context of cellulose and cellulose synthesis in plants is biofuels production from lignocellulosic biomass. Currently, biofuels production largely competes with the food supply to both animals and humans, because grain is the chief source of raw materials and has global consequences (Rosegrant, 2008). In 2007, channelling of foodstuffs, particularly maize, for biofuels production created relative food shortages, and accounted for 30% of the increase in the global grain prices. The use of waste materials such as cereal stubbles for biofuels production lessens the problems associated with the use of food related materials (Gressel, 2008). However, digestion of lignocellulosic biomass comes with its own inherent problems.

Current ethanol production methods from vegetative materials are expensive and account for up to 19% of the capital expenditure (Aden et al., 2002). They are also high energy consumption activities and often use hazardous or dangerous products such as acids and ammonia in pretreatment processes to make the cellulose material more available for digestion by hydrolase enzymes (Heinze and Koschella, 2005). Thus altering pretreatment processes would be of great value and could be achieved by modifying the cell wall.

So the challenge ahead is to produce crop plants with more easily accessed cellulose ie soluble or low crystallinity and altered non-cellulosic components, without compromising plant growth and/or yield. Changing expression levels of the endo-(1,4)-β-glucanases may provide an answer. For example, when Takahashi et al. (2009) overexpressed an endo-(1,4)-β-glucanase *KORRIGAN* (GH9A) gene in *Arabidopsis*, the resultant plants showed total cellulose levels similar to the wildtype, but these plants had reduced crystalline cellulose levels, ie, non-crystalline cellulose levels were increased. It is very likely that a more controlled approach to changing the expression levels in a number of cell wall synthesis genes may provide a way of making the lignocellulosic biomass more amenable to digestion yet still retain plant growth and yield.

Future transformation work could include that of barley with overexpression vectors using other endo-(1,4)-β-glucanase cDNAs, such as those that were co-transcribed with *HvCEL1*. Another technique that may prove useful is the employment of inducible promoters. This may reduce the chances of both transgene and endogenous gene inactivation and also have a more specific effect on secondary cell wall production. Discovery of common transcription factors among the cellulose and cell wall synthesis genes is currently underway and their overexpression may prove to be a more effective way of altering cellulose levels.

This project was based on the premise that increasing the amount of cellulose in the stems of barley and maize will increase the stem strength of these plants and reduce plant lodging. With more knowledge of how the endo-(1,4)-β-glucanases and other enzymes involved in secondary cell wall synthesis are controlled, changing expression patterns of specific endo-(1,4)-β-glucanases has the

potential to not only increase the amount of cellulose in the cell wall but also alter the crystalline structure to improve hydrolytic enzyme access for digestion.

BIBLIOGRAPHY

Aden, A., Ruth, M., Ibsen, K., J., J., Neeves, K., Sheehan, J. and Wallace, B. (2002) Lignocellulosic biomass to ethanol process design and economics utilizing co-current dilute acid prehydrolysis and enzymatic hydrolysis for corn stover. *Technical Report, National Renewable Energy Laboratory*, 1-154.

Ahlgren, E., Eriksson, K.E. and Vesterbe.O (1967) Characterization of Cellulases and Related Enzymes by Isoelectric Focusing Gel Filtration and Zone Electrophoresis I. Studies on Aspergillus Enzymes. *Acta Chem. Scand.* **21**, 937-944.

Akar, T., Avci, M. and Dusunceli, F. (2007) Barley: Post-Harvest Operations. *The Central Research Institute for Field Crops*. Mejía, D. P., FAO (Technical) and Parrucci, E. H. t. (eds.), FAO

Appenzeller, L., Doblin, M., Barreiro, R., Wang, H., Niu, X., Kollipara, K., Carrigan, L., Tomes, D., Chapman, M. and Dhugga, K.S. (2004) Cellulose synthesis in maize: isolation and expression analysis of the cellulose synthase (CesA) gene family. *Cellulose*. **11**, 287.

Arioli, T., Peng, L., Betzner, A.S., Burn, J., Wittke, W., Herth, W., Camilleri, C., ouml, Hofte, H., Plazinski, J., et al. (1998) Molecular analysis of cellulose biosynthesis in *Arabidopsis*. *Science*. **279**, 717-720.

Azuma, T., Sumida, Y., Kaneda, Y., Uchida, N. and Yasuda, T. (1996) Changes in cell wall polysaccharides in the internodes of submerged floating rice. *Plant Growth Regulation*. **19**, 183-187.

Baba, Y., Ishida, Y., Oda, M., Iiyama, K. and Akita, S. (2001) Decomposition of (1-3,1-4)-beta-glucan and expression of the (1-3,1-4)-beta-glucanase gene in rice stems during ripening. *Plant Production Science*. **4**, 230-234.

Baker, C.K., Fullwood, A.E. and Colls, J.J. (1990) Lodging of winter barley (*Hordeum vulgare* L.) in relation to its degree of exposure to sulphur dioxide. *New Phytologist*. **114**, 191-197.

Bauer, W.D., Talmadge, K.W., Keegstra, K. and Albersheim, P. (1973) The Structure of Plant Cell Walls: II. The Hemicellulose of the Walls of Suspension-cultured Sycamore Cells. *Plant Physiol.* **51**, 174-187.

Béguin, P. (1983) Detection of cellulase activity in polyacrylamide gels using Congo red-stained agar replicas. *Analytical Biochemistry*. **131**, 333-336.

Bendtsen, J.D., Nielsen, H., von Heijne, G. and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology*. **340**, 783-795.

Bernal, A.J., Yoo, C.-M., Mutwil, M., Jensen, J.K., Hou, G., Blaukopf, C., Sorensen, I., Blancaflor, E.B., Scheller, H.V. and Willats, W.G.T. (2008)

Functional Analysis of the Cellulose Synthase-Like Genes CSLD1, CSLD2, and CSLD4 in Tip-Growing Arabidopsis Cells. *Plant Physiology*. **148**, 1238-1253.

Berry, P.M., Sterling, M. and Mooney, S.J. (2006) Development of a model of lodging for barley. *Journal of Agronomy and Crop Science*. **192**, 151-158.

Bhandari, S., Fujino, T., Thammanagowda, S., Zhang, D., Xu, F. and Joshi, C. (2006) Xylem-specific and tension stress-responsive coexpression of KORRIGAN endoglucanase and three secondary wall-associated cellulose synthase genes in aspen trees. *Planta*. **224**, 828-837.

Bisgrove, S.R. (2008) The roles of microtubules in tropisms. *Plant Science*. **175**, 747-755.

Blanc, G. and Wolfe, K.H. (2004) Functional Divergence of Duplicated Genes Formed by Polyploidy during Arabidopsis Evolution. *Plant Cell*. **16**, 1679-1691.

Boudet, A.M. (2000) Lignins and lignification: Selected issues. *Plant Physiology and Biochemistry*. **38**, 81-96.

Boyd, D. and Beckwith, J. (1990) The role of charged amino acids in the localization of secreted and membrane proteins. *Cell*. **62**, 1031-1033.

Brandenberger, R., Khrebtukova, I., Thies, R.S., Miura, T., Jingli, C., Puri, R., Vasicek, T., Lebkowski, J. and Rao, M. (2004) MPSS profiling of human embryonic stem cells. *BMC Developmental Biology*. **4**, 10.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., Eletr, S., Albrecht, G., Vermaas, E., Williams, S. R., Moon, K., Burcham, T., Pallas, M., DuBridge, R. B., Kirchner, J., Fearon, K., Mao, J., Corcoran, K., (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature: Biotechnology*. **18**, 630.

Brierley, R.A., Davis, G.R. and Holtz, G.C. (1994) Production of Insulin-Like Growth Factor-1 in Methylotrophic Yeast Cells. Patent, U. S. (ed.)

Brondyk, W.H., Richard, R.B. and Murray, P.D. (2009) Chapter 11 Selecting an Appropriate Method for Expressing a Recombinant Protein. *Methods in Enzymology*, pp 131-147, Academic Press

Brown, R.M., Jr (2004) Cellulose structure and biosynthesis: What is in store for the 21st Century? *Journal of Polymer Science Part A: Polymer Chemistry*. **42**, 487-495.

Brummell, D.A., Catala, C., Lashbrook, C.C. and Bennett, A.B. (1997) A membrane-anchored E-type endo-1,4-beta-glucanase is localized on Golgi and plasma membranes of higher plants. *PNAS*. **94**, 4794-4799.

Brummell, D.A., Hall, B.D. and Bennett, A.B. (1999) Anti-sense suppression of endo-1,4-B-glucanase Cel2 mRNA accumulation increases the force required to break fruit abscission zones but does not affect fruit ripening. *Plant Molecular Biology*. **40**, 615-622.

Burgess, R.R., Richard, R.B. and Murray, P.D. (2009) Chapter 17 Refolding Solubilized Inclusion Body Proteins. *Methods in Enzymology*, pp 259-282, Academic Press

Burton, R.A. and Fincher, G.B. (2009) (1,3;1,4)-β-D-Glucans in Cell Walls of the Poaceae, Lower Plants, and Fungi: A Tale of Two Linkages. *Mol Plant*, ssp063.

Burton, R.A., Ma, G., Baumann, U., Harvey, A.J., Shirley, N.J., Taylor, J., Pettolino, F., Bacic, A., Beatty, M., Simmons, C.R., et al. (2010) A Customized Gene Expression Microarray Reveals That the Brittle Stem Phenotype fs2 of Barley Is Attributable to a Retroelement in the HvCesA4 Cellulose Synthase Gene. *Plant Physiology*. **153**, 1716-1728.

Burton, R.A., Shirley, N.J., King, B.J., Harvey, A.J. and Fincher, G.B. (2004) The CesA gene family of barley. Quantitative analysis of transcripts reveals two groups of co-expressed genes. *Plant Physiology*. **134**, 224-236.

Burton, R.A., Wilson, S.M., Hrmova, M., Harvey, A.J., Shirley, N.J., Medhurst, A., Stone, B.A., Newbigin, E.J., Bacic, A. and Fincher, G.B. (2006) Cellulose synthase-like CslF genes mediate the synthesis of cell wall (1,3;1,4)-β-D-glucans. *Science*. **311**, 1940-1942.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V. and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics *Nucleic Acids Research*. **37**, D233-D238.

Carpita, N., Tierney, M. and Campbell, M. (2001a) Molecular biology of the plant cell wall: searching for the genes that define structure, architecture and dynamics. *Plant Molecular Biology*. **47**, 1-5.

Carpita, N.C., Defernez, M., Findlay, K., Wells, B., Shoue, D.A., Catchpole, G., Wilson, R.H. and McCann, M.C. (2001b) Cell Wall Architecture of the Elongating Maize Coleoptile. *Plant Physiol.* **127**, 551-565.

Carpita, N.C. and Gibeaut, D.M. (1993) Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *The Plant Journal*. **3**, 1-30.

Carpita, N.C. and McCann, M.C. (2002) The functions of cell wall polysaccharides in composition and architecture revealed through mutations. *Plant and Soil*. **247**, 71.

Cassells, A.C. and Curry, R.F. (2001) Oxidative stress and physiological, epigenetic and genetic variability in plant tissue culture: Implications for micropropagators and genetic engineers. *Plant cell tissue and organ culture*. **64**, 145-157.

Chanliaud, E., Burrows, Kathleen, Jeronimidis, George, Gidley, Michael, (2002) Mechanical properties of primary plant cell wall analogues. *Planta*. **215**, 989.

Chanliaud, E., De Silva, Jacquie, Strongitharm, Barbara, Jeronimidis, George, Gidley, Michael J. (2004) Mechanical effects of plant cell wall enzymes on cellulose/xyloglucan composites. *The Plant Journal*. **38**, 27-37.

Chen, J. and Rattray, M. (2006) Analysis of tag-position bias in MPSS technology. *BMC Genomics*. **7**, 77.

Ching, A., Dhugga, K., Appenzeller, L., Meeley, R., Bourett, T., Howard, R. and 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.

Christianson, M.L. (2005) Codon usage patterns distort phylogenies from or of DNA sequences. *Am. J. Botany*. **92**, 1221-1233.

Clare, J.J., Romanos, M.A., Rayment, F.B., Rowedder, J.E., Smith, M.A., Payne, M.M., Sreekrishna, K. and Henwood, C.A. (1991) Production of Epidermal Growth Factor in Yeast: High-Level Secretion; Using *Pichia pastoris* Strains Containing Multiple Gene Copies *Gene*. **105**, 205-212.

Clontech SMART PCR cDNA Synthesis Kit User Manual

Cosgrove, D.J. (2005) Growth of the plant cell wall. *Nature Reviews Molecular Cell Biology*. **6**, 850.

Cosgrove, D.J. and Durachko, D.M. (1994) Autolysis and extension of isolated walls from growing cucumber hypocotyls. *Journal of Experimental Botany*. **45**, 1711-1719.

Coutinho, P.M., Deleury, E., Davies, G.J. and Henrissat, B. (2003) An Evolving Hierarchical Family Classification for Glycosyltransferases. *Journal of Molecular Biology*. **328**, 307-317.

Crepet, W.L. and Feldman, G.D. (1991) The Earliest Remains of Grasses in the Fossil Record. *American Journal of Botany*. **78**, 1010-1014.

Curtis, M.D. and Grossniklaus, U. (2003) A Gateway Cloning Vector Set for High-Throughput Functional Analysis of Genes in Planta. *Plant Physiology*. **133**, 462-469.

Das, I., Prabhakar and Indira, S. (2008) Role of stalk-anatomy and yield parameters in development of charcoal rot caused byMacrophomina phaseolina in winter sorghum. *Phytoparasitica*. **36**, 199-208.

Davies, G. and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases. *Structure*. **3**, 853-859.

Davies, G.J., Gloster, T.M. and Henrissat, B. (2005) Recent structural insights into the expanding world of carbohydrate-active enzymes. *Current Opinion in Structural Biology*. **15**, 637-645.

del Campillo, E., Abdel-Aziz, A., Crawford, D. and Patterson, S.E. (2004) Root cap specific expression of an endo-β-1,4-D-glucanase (cellulase): a new marker to study root development in *Arabidopsis*. *Plant Molecular Biology*. **56**, 309-323.

Delmer, D.P. (1999) Cellulose biosynthesis: Exciting times for a difficult field of study. *Annual Review of Plant Physiology and Plant Molecular Biology*. **50**, 245-276.

Dhugga, K.S., Barreiro, R., Whitten, B., Stecca, K., Hazebroek, J., Randhawa, G.S., Dolan, M., Kinney, A.J., Tomes, D., Nichols, S., et al. (2004) Guar seed β-mannan synthase is a member of the cellulose synthase super gene family. *Science*. **303**, 363-366.

Dickison, W.C. (2000) *Integrative Plant Anatomy*, 1 ed., Academic Press

Doblin, M.S., Kurek, Isaac, Jacob-Wilk, Deborah and Delmer Deborah P. (2002) Cellulose biosynthesis in plants: from genes to rosettes. *Plant cell Physiology*. **43**, 1407-1420.

Doblin, M.S., Pettolino, F.A., Wilson, S.M., Campbell, R., Burton, R.A., Fincher, G.B., Newbigin, E. and Bacic, A. (2009) A barley cellulose synthase-like CSLH gene mediates (1,3;1,4)-β-D-glucan synthesis in transgenic Arabidopsis. *Proceedings of the National Academy of Sciences*. **106**, 5996-6001.

Donaldson, L. and Xu, P. (2005) Microfibril orientation across secondary cell wall of *Radiata* pine tracheids. *Trees*. **19**, 644-653.

Druka, A., Muehlbauer, G., Druka, I., Caldo, R., Baumann, U., Rostoks, N., Schreiber, A., Wise, R., Close, T., Kleinhofs, A., et al. (2006) An atlas of gene expression from seed to seed through barley development. *Functional & Integrative Genomics*. **6**, 202-211.

Dunnoyer, P. and Voinnet, O. (2005) The complex interplay between plant viruses and host RNA-silencing pathways. *Current Opinion in Plant Biology*. **8**, 415-423.

Duvick, D.N., Cassman, K.G. (1999) Post green revolution trends in yield potential in temperate maize in the north-central United States. *Crop Science*. **39**, 1662.

Echezone, B.C. (2007) Corn-stalk lidging and borer damage as influenced by varying corn densities and planting geometry with soybean (*glycine max*. L. Merril). *International Agrophysics*. **21**, 133-143.

Elmayan, T., Balzergue, S., Beon, F., Bourdon, V., Daubremet, J., Guénet, Y., Mourrain, P., Palauqui, J.-C., Vernhettes, S., Vialle, T., et al. (1998) Arabidopsis Mutants Impaired in Cosuppression. *The Plant Cell*. **10**, 1747-1757.

Esau, K. (1977) *Anatomy of Seed Plants*, John Wiley and Sons inc.

Esechie, H.A., Rodriguez, V. and Al-Asmi, H. (2004) Comparison of local and exotic maize varieties for stalk lodging components in a desert climate. *European Journal of Agronomy*. **21**, 21.

Eyre-Walker, A. and Bulmer, M. (1993) Reduced synonymous substitution rate at the start of enterobacterial genes. *Nucleic Acids Research*. **21**, 4599-4603.

Farrokhi, N., Burton, R.A., Brownfield, L., Hrmova, M., Wilson, S.M., Bacic, A. and Fincher, G.B. (2006) Plant cell wall biosynthesis: genetic, biochemical and functional genomics approaches to the identification of key genes. *Plant Biotechnology Journal*. **4**, 145-167.

Ferrarese, L., Moretto, P., Trainotti, L., Rascio, N. and Casadoro, G. (1996) Cellulase involvement in the abscission of peach and pepper leaves is affected by salicylic acid. *J. Exp. Bot.* **47**, 251-257.

Ferrarese, L., Trainotti, L., Gattolin, S. and Casadoro, G. (1998) Secretion, purification and activity of two recombinant pepper endo-[beta]-1,4-glucanases expressed in the yeast Pichia pastoris. *FEBS Letters*. **422**, 23-26.

Fincher, G.B. (2009) Revolutionary Times in Our Understanding of Cell Wall Biosynthesis and Remodeling in the Grasses. *Plant Physiol.* **149**, 27-37.

Fincher, G.B. and Stone, B.A. (2004) Chemistry of non-starch polysaccharides. *Encyclopaedia of Grain Science*. Wrigley, C., Conke, H. and Walker, C. E. (eds.), pp 206-223

Flavell, R. (1994) Inactivation of Gene Expression in Plants as a Consequence of Specific Sequence Duplication. *PNAS*. **91**, 3490-3496.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.-l. and Postlethwait, J. (1999) Preservation of Duplicate Genes by Complementary, Degenerative Mutations. *Genetics*. **151**, 1531-1545.

Fry, S. (1993) Plant cell expansion: loosening the ties. *Current Biology*. **3**, 355-357.

Fukunaga, R. and Doudna, J.A. (2009) dsRNA with 5' overhangs contributes to endogenous and antiviral RNA silencing pathways in plants. *EMBO J*. **28**, 545-555.

Gale, M.D. and Devos, K.M. (1998) Comparative genetics in the grasses. *Proceedings of the National Academy of Sciences of the United States of America*. **95**, 1971-1974.

Gaut, B.S. and Doebley, J.F. (1997) DNA sequence evidence for the segmental allotetraploid origin of maize. *Proceedings of the National Academy of Sciences of the United States of America*. **94**, 6809-6814.

Gaut, B.S., Morton, B.R., McCaig, B.C. and Clegg, M.T. (1996) Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene Adh parallel rate differences at the plastid gene rbcL. *Proceedings of the National Academy of Sciences of the United States of America*. **93**, 10274-10279.

Gaut, B.S., Muse, S.V., Clark, W.D. and Clegg, M.T. (1992) Relative rates of nucleotide substitution at the rbcl locus of monocotyledonous plants. *Journal of Molecular Evolution*. **35**, 292-303.

Gebler, J., Gilkes, N.R., Claeyssens, M., Wilson, D.B., Beguin, P., Wakarchuk, W.W., Kilburn, D.G., Miller, R.C.J., Warren, R.A. and Withers, S.G. (1992) Stereoselective hydrolysis catalyzed by related β-1,4-glucanases and β-1,4-xylanases. *Journal of Biological Chemistry*. **267**, 12559-12561.

Geisse, S., Gram, H., Kleuser, B. and Kocher, H.P. (1996) Eukaryotic Expression Systems: A Comparison. *Protein Expression and Purification*. **8**, 271-282.

Gibeaut, D.M., Pauly, M., Bacic, A., and Fincher, G.B. (2005) Changes in cell wall polysaccharides in developing barley (*Hordeum vulgare*) coleoptiles. *Planta*. **221**, 729.

Gibeaut, D.M., Pauly, Markus, Bacic, Antony, Fincher, Geoffrey B. (2005) Changes in cell wall polysaccharides in developing barley (*Hordeum vulgare*) coleoptiles. *Planta*. **221**, 729.

Gilad, R., Rabinovich, L., Yaron, S., Bayer, E.A., Lamed, R., Gilbert, H.J. and Shoham, Y. (2003) CelI, a noncellulosomal Family 9 enzyme from *Clostridium thermocellum*, is a progressive endoglucanase that degrades crystalline cellulose. *Journal of Bacteriology*. **185**, 391-398.

Gilkes, N.R., Henrissat, B., Kilburn, D.G., Miller, R.C.J. and Warren, R.A.J. (1991) Domains in microbial β-1,4-glycanases: sequence conservation, function and enzyme famillies. *Microbiological Reviews*. **55**, 303-315.

Goldman, N. and Yang, Z. (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol*. **11**, 725-736.

Grabber, J.H. (2005) How do lignin composition, structure and cross-linking affect biodegradability? A review of cell wall model studies. *Crop Science*. **45**, 820-831.

Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R. (1981) Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucl. Acids Res.* **9**, 213-b-.

Gräslund, S., Nordlund, P., Weigelt, J., Hallberg, B.M., Bray, J., Gileadi, O., Knapp, S., Oppermann, U., Arrowsmith, C., Hui, R., et al. (2008) Protein production and purification. *Nature Methods*. **5**, 135(12).

Gressel, J. (2008) Transgenics are imperative for biofuel crops. *Plant Science*. **174**, 246-263.

Harche, M., Chadli, R. and Catesson, A.M. (1990) Diversity of cellulose microfibril arrangement in the cell walls of *Lygeum spartum* leaves. *Ann Bot*. **65**, 79-86.

Hartati, S., Sudarmonowati, E., Park, Y.W., Kaku, T., Kaida, R., Baba, K.i. and Hayashi, T. (2008) Overexpression of poplar cellulase accelerates growth and disturbs the closing movements of leaves in sengon. *Plant Physiol.*, pp.108.116970.

Hartung Frank, Blattner Frank R. and Puchta Holger (2002) Intron gain and loss in the evolution of the conserved eukaryotic recombination machinery. *Nucl. Acids Res.* **30**, 5175-5181.

Hatfield, R.D., Wilson, J.R. and Mertens, D.R. (1999) Composition of cell walls isolated from cell types of grain sorghum stems. *Journal of the Science of Food and Agriculture*. **79**, 891-899.

Hayashi, T. and Maclachlan, G. (1984) Pea Xyloglucan and Cellulose : I. Macromolecular Organization. *Plant Physiol.* **75**, 596-604.

Hayashi, T., Marsden, M.P.F. and Delmer, D.P. (1987) Pea Xyloglucan and Cellulose: VI. Xyloglucan-Cellulose Interactions in Vitro and in Vivo. *Plant Physiol.* **83**, 384-389.

Hayashi, T., Yoshida, K., Park, Y.W., KonishiT. and Baba, K. (2005) Cellulose metabolism in plants. *International Review of Cytology*. **247**, 1-33.

Hazen, S.P., Scott-Craig, J.S. and Walton, J.D. (2002) Cellulose Synthase-Like Genes of Rice. *Plant Physiol.* **128**, 336-340.

Hedden, P. (2003) The genes of the Green Revolution. *Trends in Genetics*. **19**, 5-9.

Heger, A. and Ponting, C.P. (2007) Evolutionary rate analyses of orthologs and paralogs from 12 Drosophila genomes. *Genome Res.* **17**, 1837-1849.

Heinze, T. and Koschella, A. (2005) Solvents applied in the field of cellulose chemistry: a mini review. *Polı´meros* **15**, 84-90.

Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemistry journal*. **280**, 309-316.

Henrissat, B. and Bairoch, A. (1993) New famillies in the classification of glycosyl hydrolases based on amino acid sequence. *Biochemistry Journal*. **293**, 781-788.

Heuchert, J.C., Marks, J.S. and Mitchell, C.A. (1983) Strengthening Tomato Shoots by Gyratory Shaking. *Journal of American Society for Horticultural Science*. **108**, 801-805.

Hill, J. (2007) How a Corn Plant Develops - Identifying Stages of Development University, I. S. (ed.)

His, I., Driouich, A., Nicol, F., Jauneau, A. and Hofte, H. (2001) Altered pectin composition in primary cell walls of *korrigan*, a dwarf mutant of *Arabidopsis* deficient in a membrane-bound endo1,4-B-glucanase. *Planta*. **212**, 348-358.

Hrmova, M., Farkas, V., Harvey, A.J., Lahnstein, J., Wischmann, B., Kaewthai, N., Ezcurra, I., Teeri, T.T. and Fincher, G.B. (2009) Substrate specificity and catalytic mechanism of a xyloglucan xyloglucosyl transferase HvXET6 from barley (*Hordeum vulgare* L.). *FEBS*. **276**, 437-456.

Hrmova, M., Farkas, V., Lahnstein, J. and Fincher, G.B. (2007) A Barley Xyloglucan Xyloglucosyl Transferase Covalently Links Xyloglucan, Cellulosic Substrates, and (1,3;1,4)-beta-D-Glucans. *J. Biol. Chem.* **282**, 12951-12962.

http://bio.lundberg.gu.se/edu/translat.html.

http://signal.salk.edu/cgi-bin/tdnaexpress.

http://www.cazy.org/CAZY/

http://www.cazy.org/fam/acc_GH.html.

http://www.gramene.org/

http://www.ncbi.nlm.nih.gov/.

http://www.tigr.org/.

Idiris, A., Tohda, H., Kumagai, H. and Takegawa, K. (2010) Engineering of protein secretion in yeast: strategies and impact on protein production. *Applied Microbiology and Biotechnology*. **86**, 403-417.

Ina, Y. (1996) Pattern of Synonymous and nonsynonymous substitutions: an indicator of mechanisms of molecular evolution. *Journal of Genetics*. **75**, 91-115.

Invitrogen Elongase® Enzyme Mix. *Cat. No. 10480-010*.

Invitrogen (2002a) BL21 Star (DE3) One Shot® BL21 Star (DE3)pLysS One Shot® Chemically Competent Cells Version D

Invitrogen (2002b) pET-DEST42 Gateway™ Vector A destination vector for high-level, inducible expression in E. coli, Catalog no. 12276-010 Version B.

Invitrogen (2004-2005) Gateway® BP Clonase™ II Enzyme Mix. .

Invitrogen (2006) pCR®8/GW/TOPO® TA Cloning® Kit.

Invitrogen (2008) pPICZ α A, B and C *Pichia* expression vectors for selection on Zeocin™ and purification of secreted, recombinant proteins Catalogue number V195-20 Version F. *User Manual*

Irimia, M. and Roy, S.W. (2008) Spliceosomal introns as tools for genomic and evolutionary analysis. *Nucl. Acids Res.* **36**, 1703-1712.

Jampatong, S., Darrah, L. L., Krause, G. F., Barry, B. D. (2000) Effect of one- and two-eared selection on stalk strength and other characters in maize. *Crop Science*. **40**, 605-611.

Jedel, P.E., Helm, J.H. (1991) Lodging effects on a semidwarf and two standard barley cultivars. *Agronomy Journal*. **83**, 151-168.

Jezowski, S., Surma, M., Krajewski, P., Adamski, T. (2003) Genotype-enviroment interaction of barley DH lines in terms of morphological and physical traits of the stem and the degree of lodging. *International Agrophysics*. **17**, 57-60.

Jones, J.D.G., Gilbert, D.E., Grady, K.L. and Jorgensen, R.A. (1987) T-DNA structure and gene expression in petunia plants transformed by Agrobacterium tumefaciens C58 derivatives. *Molecular and General Genetics MGG*. **207**, 478-485.

Joshi, C., Bhandari, S., Ranjan, P., Kalluri, U., Liang, X., Fujino, T. and Samuga, A. (2004) Genomics of cellulose biosynthesis in poplars. *New Phytologist*. **164**, 53-61.

Joshi, C.P., Thammannagowda, S., Fujino, T., Gou, J.-Q., Avci, U., Haigler, C.H., McDonnell, L.M., Mansfield, S.D., Mengesha, B., Carpita, N.C., et al. (2011) Perturbation of Wood Cellulose Synthesis Causes Pleiotropic Effects in Transgenic Aspen. *Molecular Plant*. **4**, 331-345.

Jung, H.G. and Casler, M.D. (2006) Maize stem tissues: Cell wall concentration and composition during development. *Crop Science*. **46**, 1793-1800.

Juy, M., Amit, A.G., Alzari, P.M., Poljak, R.J., Claeyssons, M., Beguin, P. and Aubert, J.P. (1992) Three dimensional structure of a thermostable bacterial cellulase. *Nature*. **357**, 89-91.

Kanazawa, A., O'Dell, M. and Hellens, R.P. (2007) Epigenetic Inactivation of Chalcone Synthase-A Transgene Transcription in Petunia Leads to a Reversion of the Post-Transcriptional Gene Silencing Phenotype. *Plant Cell Physiol.* **48**, 638-647.

Kawamura, O., Nijima, N., Niimi, M. and Akashi, R. (2002) Dry Matter and Lignin Distributions in Maize (*Zea mays* L.) Internodes. *Grassland Science*. **47**, 578-582.

Kay, B.K., Williamson, M.P. and Sudol, M. (2000) The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J.* **14**, 231-241.

Kellogg, E.A. (1998) Relationships of cereal crops and other grasses. *Proceedings of the National Academy of Sciences of the United States of America*. **95**, 2005-2010.

Kemmerer, E.C. and Tucker, M.L. (1994) Comparative study of cellulases associated with adventitious root initiation, apical buds, and leaf, flower, and pod abscission zones in soybean. *Plant Physiology* **104**, 557-562.

Khademi, S., Guarino, L.A., Watanabe, G.T. and Meyer, E.F. (2002) Structure of an endoglucanase from termite, *Nasutitermes takasagoensis*. *Acta Crystallographica Section D*. **58**, 653-659.

Kimura, M. and Ohta, T. (1974) On Some Principles Governing Molecular Evolution. *Proceedings of the National Academy of Sciences of the United States of America*. **71**, 2848-2852.

Koshland Jr, D.E. (1953) Stereochemistry and the mechanism of enzymatic reactions. *Biological Review of the Cambridge Philosophical Society*. **28**, 416-436.

Kost, T., Condreay, J. and DL, J. (2005) Baculovirus as versatile vectors for protein expression in insect and mammalian cells. . *Nature Biotechnology*. **23**, 567-575.

Krauss, V., Pecyna, M., Kurz, K. and Sass, H. (2005) Phylogenetic Mapping of Intron Positions: A Case Study of Translation Initiation Factor eIF2{gamma}. *Mol Biol Evol*. **22**, 74-84.

Lafarguette, F., Leplé, J.C., Déjardin, A., Laurans, F., Costa, G., Lesage-Descauses, M.C. and Pilate, G. (2004) Poplar genes encoding fasciclin-like arabinogalactan proteins are highly expressed in tension wood. *New Phytologist*. **164**, 107-121.

Lambert, D. (2007) *The Field Guide to Geology,* Checkmark Books, New York

Lane, D.R., Wiedemeier, A., Peng, L., Hofte, H., Vernhettes, S., Desprez, T., Hocart, C.H., Birch, R.J., Baskin, T.I., Burn, J.E., et al. (2001) Temperature-sensitive alleles of RSW2 link the Korrigan endo-1,4-β-glucanase to cellulose synthesis and cytokinesis in *Arabidopsis*. *Plant Physiology*. **126**, 278-288.

Lang, P.M., Mallet, J.B., Berry, W.A.J., (1986) The effect of tillage system on lodging in maize resulting from a severe rainstorm. *South African Journal of Plant Soil*. **3**, 165-170.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*. **23**, 2947-2948.

Lawton-Rauh, A. (2003) Evolutionary dynamics of duplicated genes in plants. *Molecular Phylogenetics and Evolution*. **29**, 396-409.

Lee, Y. and Kende, H. (2001) Expression of beta -Expansins Is Correlated with Internodal Elongation in Deepwater Rice. *Plant Physiol.* **127**, 645-654.

Lee, Y. and Kende, H. (2002) Expression of alpha -Expansin and Expansin-Like Genes in Deepwater Rice. *Plant Physiol.* **130**, 1396-1405.

Lerouge, P., Cabanes-Macheteau, M., Rayon, C., Fischette-Lainé, A.-C., Gomord, V. and Faye, L. (1998) N-Glycoprotein biosynthesis in plants: recent developments and future trends. *Plant Molecular Biology*. **38**, 31-48.

Lerouxel, O., Choo, T.S., Seveno, M., Usadel, B., Faye, L., Lerouge, P. and Pauly, M. (2002) Rapid structural phenotyping of plant cell wall mutants by enzymatic oligosaccharide fingerprinting. *Plant Physiology*. **130**, 1754-1763.

Levy, I., Shani, Z. and Shoseyov, O. (2002) Modification of polysaccharides and plant cell wall by endo-1,4-β-glucanase and cellulose-binding domains. *Biomolecular Engineering*. **19**, 17-30.

Lewis, D., Bacic, A., Chandler, P.M. and Newbegin, E.J. (2009) Aberrant cell expansion in the *elongation* mutants. *Plant and Cell Physiology*. **50**, 554-571.

Li, M., Xiong, G., Li, R., Cui, J., Tang, D., Zhang, B., Pauly, M., Cheng, Z. and 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, Y., Qian, Q., Zhou, Y., Yan, M., Sun, L., Zhang, M., Fu, Z., Wang, Y., Han, B., Pang, X., et al. (2003) BRITTLE CULM1, Which Encodes a COBRA-Like Protein, Affects the Mechanical Properties of Rice Plants. *Plant Cell*. **15**, 2020-2031.

Libertini, E., Li, Y. and McQueen-Mason, S. (2004) Phylogenetic Analysis of the Plant Endo-β-1,4-Glucanase Gene Family. *Journal of Molecular Evolution*. **58**, 506.

Liebhardt, W.C., Stangel, P.J. and Murdock, T.J. (1968) A Mechanism for Premature Parenchyma Breakdown in Corn (*Zea Mays* Lt .). *Agronomy Journal*. **60**, 496-499.

Liepman, A.H., Wilkerson, C.G. and Keegstra, K. (2005) Expression of cellulose synthase-like (Csl) genes in insect cells reveals that CslA family members encode mannan synthases. *Proceedings of the National Academy of Science*. **102**, 2221-2226.

Linder, H.P. (1987) The Evolutionary History of the Poales/Restionales: A Hypothesis. *Kew Bulletin*. **42**, 297-318.

Linder, M. and Teeri, T.T. (1997) The roles and function of cellulose-binding domains. *Journal of Biotechnology*. **57**, 15-28.

Livak, K.J. and Schmittgen, T.D. (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $22^{-\Delta\Delta CT}$ Method. *Methods*. **25**, 402-408.

Llave, C., Xie, Z., Kasschau, K.D. and Carrington, J.C. (2002) Cleavage of Scarecrow-like mRNA Targets Directed by a Class of Arabidopsis miRNA. *Science*. **297**, 2053-2056.

Lynch, M. and Conery, J.S. (2000) The Evolutionary Fate and Consequences of Duplicate Genes. *Science*. **290**, 1151-1155.

Lynch, M. and Force, A. (2000) The Probability of Duplicate Gene Preservation by Subfunctionalization. *Genetics*. **154**, 459-473.

Marillonnet, S., Thoeringer, C., Kandzia, R., Klimyuk, V. and Gleba, Y. (2005) Systemic Agrobacterium tumefaciens-mediated transfection of viral replicons for efficient transient expression in plants. *Nat Biotech*. **23**, 718-723.

Martin, S.A., Darrah, L. L. and Hibbard, B. E. (2004) Divergent selection for rind penetrometer resistance and its effects on european corn borer damage and stalk traits in corn. *Crop Science*. **44**, 711-717.

Master, E.R., Rudsander, U.J., Zhou, W., Henriksson, H., Divne, C., Denman, S., Wilson, D.B. and Teeri, T.T. (2004) Recombinant Expression and Enzymatic Characterization of PttCel9A, a KOR Homologue from *Populus tremula* x *tremuloides*. *Biochemistry*. **43**, 10080-10089.

Mattanovich, D., Gasser, B., Hohenblum, H. and Sauer, M. (2004) Stress in recombinant protein producing yeasts. *Journal of Biotechnology*. **113**, 121-135.

Mattinen, M.-L., Linder, M., Teleman, A. and Annila, A. (1997) Interaction between cellohexoase and cellulose binding domains from *Trichoderma reesei* cellulases. *FEBS letters*. **407**, 289-290.

Mayer, K.F.X., Martis, M., Hedley, P.E., Å imkovÃ¡, H., Liu, H., Morris, J.A., Steuernagel, B., Taudien, S., Roessner, S., Gundlach, H., et al. (2011) Unlocking the Barley Genome by Chromosomal and Comparative Genomics. *The Plant Cell Online*. **4**, 1249-63.

McQueen-Mason, S. and Cosgrove, D.J. (1994) Disruption of Hydrogen Bonding Between Plant Cell Wall Polymers by Proteins that Induce Wall Extension. *Proceedings of the National Academy of Sciences*. **91**, 6574-6578.

McQueen-Mason, S., Durachko, D.M. and Cosgrove, D.J. (1992) Two Endogenous Proteins That Induce Cell Wall Extension in Plants. *The Plant Cell*. **4**, 1425-1433.

Messing, J., Bharti, A.K., Karlowski, W.M., Gundlach, H., Kim, H.R., Yu, Y., Wei, F., Fuks, G., Soderlund, C.A., Mayer, K.F.X., et al. (2004) Sequence composition and genome organization of maize. *Proceedings of the National Academy of Sciences of the United States of America*. **101**, 14349-14354.

Meyer, P., Heidmann, I. and Niedenhof, I. (1993) Differences in DNA-methylation are associated with a paramutation phenomenon in transgenic petunia. *The Plant Journal*. **4**, 89-100.

Meyers, B.C., Tej, S.S., Vu, T.H., Haudenschild, C.D., Agrawal, V., Edberg, S.B., Ghazal, H. and Decola, S. (2004) The Use of MPSS for Whole-Genome Transcriptional Analysis in Arabidopsis. *Genome Research*. **14**, 1641-1653.

Molhoj, M., Pagant, S. and Hofte, H. (2002) Towards understanding the role of membrane-bound endo-β-1,4-glucanases in cellulose biosynthesis. *Plant Cell Physiology*. **43**, 1399-1406.

Molhoj, M., Ulvskov, P. and Degan, D. (2001) Characterisation of a functional soluble form of a *Brassica napis* membrane-anchored endo-1,4-β-glucanase heterologously expressed in *Pichia pastoris*. *Plant Physiology*. **127**, 674-684.

Moller, I., Sørensen, I., Bernal, A.J., Blaukopf, C., Lee, K., Øbro, J., Pettolino, F., Roberts, A., Mikkelsen, J.D., Knox, J.P., et al. (2007) High-throughput mapping of cell-wall polymers within and between plants using novel microarrays. *The Plant Journal*. **50**, 1118-1128.

Moore, G., Devos, K.M., Wang, Z. and Gale, M.D. (1995) Cereal Genome Evolution: Grasses, line up and form a circle. *Current Biology*. **5**, 737-739.

Morel, J.-B., Godon, C., Mourrain, P., Béclin, C., Boutet, S.p., Feuerbach, F., Proux, F. and Vaucheret, H. (2002) Fertile Hypomorphic ARGONAUTE (ago1) Mutants Impaired in Post-Transcriptional Gene Silencing and Virus Resistance. *The Plant Cell*. **14**, 629-639.

Morel, J.-B., Mourrain, P., Béclin, C. and Vaucheret, H. (2000) DNA methylation and chromatin structure affect transcriptional and post-transcriptional transgene silencing in Arabidopsis. *Current Biology*. **10**, 1591-1594.

Morrison, T.A., Kessler, J.R. and Buxton, D.R. (1994) Maize Internode Elongation Patterns. *Crop Science*. **34**, 1055-1060.

Mourrain, P., van Blokland, R., Kooter, J. and Vaucheret, H. (2007) A single transgene locus triggers both transcriptional and post-transcriptional silencing through double-stranded RNA production. *Planta*. **225**, 365-379.

Muse, S.V. and Gaut, B.S. (1994) A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol Biol Evol*. **11**, 715-724.

Nakano, M., Nobuta, K., Vemaraju, K., Tej, S.S., Skogen, J.W. and Meyers, B.C. (2006) Plant MPSS databases: signature-based transcriptional resources for analyses of mRNA and small RNA. *Nucleic Acids Research*. **34**, D731-735.

Napoli, C., Lemieux, C. and Jorgensen, R. (1990) Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *Plant Cell*. **2**, 279-289.

Needs, P.W. and Selvendran, R.R. (1993) An improved methylation procedure for the analysis of complex polysaccharides including resistant starch and a critique of the factors which lead to undermethylation. *Phytochemical Analysis*. **4**, 210-216.

Nei, M. and Gojobori, T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*. **3**, 418-426.

Nelson, N. (1944) A photometric adaptation of the Somogyi method for the determination of glucose. *Journal of Biological Chemistry*. **153**, 375-380.

Nicol, F., His, I., Jauneau, A., Vernhettes, S., Canut, H. and Hofte, H. (1998) A plasma membrane-bound putative endo-1,4-β-D-glucanase is required for normal wall assembly and cell elongation in *Arabidopsis*. *The EMBO Journal*. **17**, 5563-5576.

Nishitani, K. and Tominaga, R. (1992) Endo-xyloglucan transferase, a novel class of glycosyltransferase that catalyzes transfer of a segment of

xyloglucan molecule to another xyloglucan molecule. *Journal of Biological Chemistry*. **267**, 21058-21064.

Novagen Competent cells: What a difference a strain makes, Merck Biosciences

Novagen (2006) pET System Manual

Nunan, K.J., Davies, C., Robinson, S.P. and Fincher, G.B. (2001) Expression patterns of cell wall-modifying enzymes during grape berry development. *Planta*. **V214**, 257.

Page, R.D.M. (1996) Tree View: An application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12**, 357-358.

Paredez, A.R., Persson, S., Ehrhardt, D.W. and Somerville, C.R. (2008) Genetic Evidence That Cellulose Synthase Activity Influences Microtubule Cortical Array Organization. *Plant Physiol.* **147**, 1723-1734.

Park, Y.W., Tominaga, R., Sugiyama, J., Furuta, Y., Tanimoto, E., Samejima, M., Sakai, F. and Hayashi, T. (2003) Enhancement of growth by expression of poplar cellulase in *Arabidopsis thaliana*. *The Plant Journal*. **33**, 1099-1106.

Passarge, E., Horsthemke, B. and Farber, R.A. (1999) Incorrect use of the term synteny. *Nature Genetics*. **23**, 387.

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., et al. (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature*. **457**, 551-556.

Paterson, A.H., Bowers, J.E. and Chapman, B.A. (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proceedings of the National Academy of Sciences of the United States of America*. **101**, 9903-9908.

Peng, J., Richards, D.E., Hartley, N.M., Murphy, G.P., Devos, K.M., Flintham, J.E., Beales, J., Fish, L.J., Worland, A.J., Pelica, F., et al. (1999) `Green revolution' genes encode mutant gibberellin response modulators. *Nature*. **400**, 256-261.

Peng, L., Kawagoe, Y., Hogan, P. and Delmer, D. (2002) Sitosterol-β-glucoside as primer for cellulose synthesis in plants. *Science*. **295**, 147-150.

Penning, B.W., Hunter, C.T., III, Tayengwa, R., Eveland, A.L., Dugard, C.K., Olek, A.T., Vermerris, W., Koch, K.E., McCarty, D.R., Davis, M.F., et al. (2009) Genetic Resources for Maize Cell Wall Biology. *Plant Physiol.* **151**, 1703-1728.

PLEXdb (2001-2006) Plant Genome Database, PLEXdb

Prest, T.J., Cantrell, R.P. and Axtell, J.D. (1983) Heritability of Lodging Resistance and Its Association with Other Agronomic Traits in a Diverse Sorghum Population. *Crop Sci*. **23**, 217-221.

Qiagen (2008) QIAsymphony® Ni-NTA Handbook.

Ray, P.M. (1962) Cell Wall Synthesis and Cell Elongation in Oat Coleoptile Tissue. *American Journal of Botany*. **49**, 928-939.

Reinartz, J., Bruyns, E., Lin, Jing-Zhong, Burcham, T., Brenner, S., Bowen, B., Kramer, M. and Woychik, R. (2002) Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. *Briefings in Functional Genomics and Proteomics*. **1**, 95-104.

Richmond, T.A. and Somerville, C.R. (2001) Integrative approaches to determining Csl function. *Plant Molecular Biology*. **47**, 131-143.

Robert, S., Mouille, G. and Höfte, H. (2004) The mechanism and regulation of cellulose synthesis in primary walls: lessons from cellulose-deficient Arabidopsis mutants. *Cellulose*. **11**, 351-364.

Rogozin, I.B., Sverdlov, A.V., Babenko, V.N. and Koonin, E.V. (2005) Analysis of evolution of exon-intron structure of eukaryotic genes. *Brief Bioinform*. **6**, 118-134.

Rosegrant, M.W. (2008) Biofuels and Grain Prices: Impacts and Policy Responses. *Testimony for the U.S. Senate Committee on Homeland Security and Governmental Affairs*, International Food Policy Research Institute

Roulin, S., Buchala, A.J. and Fincher, G.B. (2002) Induction of (1-3,1-4)-β-D-glucan hydrolases in leaves of dark-incubated barley seedlings. *Planta*. **V215**, 51.

Sahdev, S., Khattar, S. and Saini, K. (2008) Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Molecular and Cellular Biochemistry*. **307**, 249-264.

Salnikov, V.V., Grimson, M. J., Delmer, D. P. and Haigler, C. H. (2001) Sucrose synthase localizes to cellulose synthesis sites in tracheary elements. *Phytochemistry*. **57**, 823.

Salse, J., Bolot, S., Throude, M., Jouffe, V., Piegu, B., Quraishi, U.M., Calcagno, T., Cooke, R., Delseny, M. and Feuillet, C. (2008) Identification and Characterization of Shared Duplications between Rice and Wheat Provide New Insight into Grass Genome Evolution. *Plant Cell*. **20**, 11-24.

Sankoff, D., Zheng, C. and Zhu, Q. (2007) Polyploids, genome halving and phylogeny. *Bioinformatics*. **23**, i433-439.

Sato, S., Kato, T., Kakegawa, K., Ishii, T., Liu, Y.-G., Awano, T., Takabe, K., Nishiyama, Y., Kuga, S., Sato, S., et al. (2001) Role of the putative membrane-bound endo-1,4-b-glucanase Korrigan in cell elongation and cellulose synthesis in *Arabidopsis thaliana*. *Plant Cell Physiology*. **42**, 251-263.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F.S., Pasternak, S., Liang, C.Z., Zhang, J.W., Fulton, L., Graves, T.A., et al. (2009) The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science*. **326**, 1112-1115.

Schreiber, A.W., Shirley, N.J., Burton, R.A. and Fincher, G.B. (2008) Combining transcriptional datasets using the generalised singular value decomposition. *BMC Bioinformatics*. **9**, 335.

Schrick, K., Fujioka, S., Takatsuto, S., Stierhof, Y.-D., Stransky, H., Yoshida, S. and Jurgens, G. (2004) A link between sterol biosynthesis, the cell wall, and cellulose in *Arabidopsis*. *The Plant Journal*. **38**, 227-243.

Schulte, D., Close, T.J., Graner, A., Langridge, P., Matsumoto, T., Muehlbauer, G., Sato, K., Schulman, A.H., Waugh, R., Wise, R.P., et al. (2009) The International Barley Sequencing Consortium--At the Threshold of Efficient Access to the Barley Genome. *Plant Physiology*. **149**, 142-147.

Sexton, R., del Campillo, E., Duncan, D. and Lewis, L.N. (1990) The purification of an anther cellulase (β(1,4)-glucan hydrolase) from *Lathyrus odoratus* L. and its relationship to the similar enzyme found in abscission zones. *Plant Science*. **67**, 169-176.

Shah, T., de Villiers, E., Nene, V., Hass, B., Taracha, E., Gardner, M.J., Sansom, C., Pelle, R. and Bishop, R. (2006) Using the transcriptome to annotate the genome revisited: Application of massively parallel signature sequencing (MPSS). *Gene*. **366**, 104.

Shani, Z., Dekel, M., Roiz, L., Horowitz, M., Kolosovski, N., Lapidot, S., Alkan, S., Koltai, H., Tsabary, G., Goren, R., et al. (2006) Expression of endo-1,4-B-glucanase (cel1) in Arabidopsis thaliana is associated with plant growth, xylem development and cell wall thickening. *Plant Cell Reports*. **V25**, 1067.

Shani, Z., Dekel, M., Tsabary, G., Goren, R. and Shoseyov, O. (2004) Growth enhancement of transgenic poplar plants by overexpression of Arabidopsis thaliana endo-1,4-B-glucanase (cel1). *Molecular Breeding*. **14**, 321-330.

Shani, Z., Dekel, M., Tsabary, G. and Shoseyov, O. (1997) Cloning and characterization of elongation specific endo-1,4-β-glucanase (cel1) from Arabidopsis thaliana. *Plant Molecular Biology*. **34**, 837-842.

Shivprasad, S., Pogue, G.P., Lewandowski, D.J., Hidalgo, J., Donson, J., Grill, L.K. and Dawson, W.O. (1999) Heterologous Sequences Greatly Affect

Foreign Gene Expression in Tobacco Mosaic Virus-Based Vectors. *Virology*. **255**, 312-323.

Smith, G.E., Summers, M.D. and Fraser, M.J. (1983) Production of human beta interferon in insect cells infected with a baculovirus expression vector. *Molecular Cell biology*. **3**, 2156-2165.

Solingen, P., Meijer, D., Kleij, W., Barnett, C., Bolle, R., Power, S. and Jones, B. (2001) Cloning and expression of an endocellulase gene from a novel streptomycete isolated from an East African soda lake. *Extremophiles*. **5**, 333-341.

Somerville, C., Bauer, S., Brininstool, G., Facette, M., Hamann, T., Milne, J., Osborne, E., Paredez, A., Persson, S., Raab, T., et al. (2004) Toward a systems approach to understanding plant cell walls. *Science*. **306**, 2206-2211.

Stolovitzky, G.A., Kundaje, A., Held, G.A., Duggar, K.H., Haudenschild, C.D., Zhou, D., Vasicek, T.J., Smith, K.D., Aderem, A. and Roach, J.C. (2005) Statistical analysis of MPSS measurements: Application to the study of LPS-activated macrophage gene expression. *PNAS*. **102**, 1402-1407.

Studier, F.W. (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*. **41**, 207-234.

Suyama, M., Torrents, D. and Bork, P. (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609-612.

Swigonova, Z., Lai, J., Ma, J., Ramakrishna, W., Llaca, V., Bennetzen, J.L. and Messing, J. (2004) Close Split of Sorghum and Maize Genome Progenitors. *Genome Research*. **14**, 1916-1923.

Swigoňová, Z., Lai, J., Ma, J., Ramakrishna, W., Llaca, V., Bennetzen, J.L. and Messing, J. (2004) Close Split of Sorghum and Maize Genome Progenitors. *Genome Research*. **14**, 1916-1923.

Szyjanowicz, P.M.J., McKinnon, I., Taylor, N.G., Gardiner, J., Jarvis, M.C. and Turner, S.R. (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.

Takahashi, J., Rudsander, U.J., Hedenstrom, M., Banasiak, A., Harholt, J., Amelot, N., Immerzeel, P., Ryden, P., Endo, S., Ibatullin, F.M., et al. (2009) KORRIGAN1 and its Aspen Homolog PttCel9A1 Decrease Cellulose Crystallinity in Arabidopsis Stems. *Plant Cell Physiol.* **50**, 1099-1115.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Mol Biol Evol*. **24**, 1596-1599.

Taniguchi, M., Miura, K., Iwao, H. and Yamanaka, S. (2001) Quantitative Assessment of DNA Microarrays--Comparison with Northern Blot Analyses. *Genomics*. **71**, 34-39.

Taylor, N.G., Howells, R.M., Huttly, A.K., Vickers, K. and Turner, S.R. (2003) Interactions among three distinct CesA proteins essential for cellulose synthesis. *Proceedings of the National Academy of Science*. **100**, 1450-1455.

Thompson, J.E. and Fry, S.C. (2001) Restructuring of wall-bound xyloglucan by transglycosylation in living plant cells. *The Plant Journal*. **26**, 23-34.

Tollenaar, M. (1989) Genetic improvement in grain yield of commercial maize hybrids grown in Ontario from 1959 to 1988. *Crop science*. **29**, 1365-1371.

Trainotti, L., Spolaore, S., Pavanello, A., Baldan, B. and Casadoro, G. (1999) A novel E-type endo-β-1,4-glucanase with a putative cellulose-binding domain is highly expressed in ripening strawberry fruits. *Plant Molecular Biology*. **40**, 323.

Urbanowicz, B.R., Bennett, A.B., del Campillo, E., Catala, C., Hayashi, T., Henrissat, B., Hofte, H., McQueen-Mason, S.J., Patterson, S.E., Shoseyov, O., et al. (2007a) Structural Organization and a Standardized Nomenclature for Plant Endo-1,4-beta-Glucanases (Cellulases) of Glycosyl Hydrolase Family 9. *Plant Physiology*. **144**, 1693-1696.

Urbanowicz, B.R., Catala, C., Irwin, D., Wilson, D.B., Ripoll, D.R. and Rose, J.K.C. (2007b) A Tomato Endo-beta-1,4-glucanase, SlCel9C1, Represents a Distinct Subclass with a New Family of Carbohydrate Binding Modules (CBM49). *Journal of Biological Chemistry*. **282**, 12066-12074.

Valdivia, E.I.G.d. and Isaksson, L.A. (2005) Abortive translation caused by peptidyl-tRNA drop-off at NGG codons in the early coding region of mRNA. *FEBS Journal*. **272**, 5306-5316.

Van Sandt, V.S.T., Suslov, D., Verbelen, J.-P. and Vissenberg, K. (2007) Xyloglucan Endotransglucosylase Activity Loosens a Plant Cell Wall. *Annals of Botany*. **100**, 1467-1473.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. and Speleman, F. (2002) Accurate normalisation of real-time quantitative RT-PCR data by geometric averaging if multiple control genes. *Genome Biology*. **3**.

Varenne, S., Buc, J., Lloubes, R. and Lazdunski, C. (1984) Translation is a non-uniform process : Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *Journal of Molecular Biology*. **180**, 549-576.

Vaucheret, H., Béclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J.-B., Mourrain, P., Palauqui, J.-C. and Vernhettes, S. (1998) Transgene-induced gene silencing in plants. *The Plant Journal*. **16**, 651 -659.

Vaucheret, H., Beclin, C. and Fagard, M. (2001) Post-transcriptional gene silencing in plants. *J Cell Sci*. **114**, 3083-3091.

Vera, A., González-Montalbán, N., Arís, A. and Villaverde, A. (2007) The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. *Biotechnology and Bioengineering*. **96**, 1101-1106.

Vlasenko, E.Y., Ryan, A.I., Shoemaker, C.F. and Shoemaker, S.P. (1998) The use of capillary viscometry, reducing end-group analysis, and size exclusion chromatography combined with multi-angle laser light scattering to characterize endo-1,4-[beta]--glucanases on carboxymethylcellulose: a comparative evaluation of the three methods. *Enzyme and Microbial Technology*. **23**, 350-359.

Vogel, J.P., Garvin, D.F., Mockler, T.C., Schmutz, J., Rokhsar, D. and Bevan, M.W. (2010) Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature*. **463**, 763-768.

Wei, F., Coe, E., Nelson, W., Bharti, A.K., Engler, F., Butler, E., HyeRan, K., Goicoechea, J.L., Mingsheng, C., Seunghee, L., et al. (2007) Physical and Genetic Structure of the Maize Genome Reflects Its Complex Evolutionary History. *PLoS Genetics*. **3**, e123-1263.

White, A.R. and Brown, R.M. (1981) Enzymatic hydrolysis of cellulose: visual characterization of the process. *Proceedings of the National Academy of Science*. **78**, 1047-1051.

Whitney, S.E.C., Gidley, M.J. and McQueen-Mason, S.J. (2000) Probing expansin action using cellulose/hemicellulose composites. *The Plant Journal*. **22**, 327-334.

Wierzbicki, A., Haag, J. and Pikaard, C. (2008) Noncoding Transcription by RNA Polymerase Pol IV/PolV Mediates Transcriptional Silencing of Overlapping and Adjacent Genes. *Cell*. **135**, 635-648.

Williams, S. (2006) http://en.wikipedia.org/wiki/Glycoside_hydrolase

Williamson, M.P. (1994) The structure and function of proline-rich regions in proteins. *Biochemistry journal*. **297**, 249-260.

Wilson, I.B.H. (2002) Glycosylation of proteins in plants and invertebrates. *Current Opinion in Structural Biology*. **12**, 569-577.

Wise, R.P., Caldo, R.A., Hong, L., Shen, L., Cannon, E.K. and Dickerson, E.J. (2007) BarleyBase/PLEXdb: A Unified Expression Profiling Database for Plants and Plant Pathogens. *Methods in Molecular Biology*. Edwards, D. (ed.), pp 347-363, Humana Press, Totowa, NJ

Withers, S.G., Dombroski, D., Berven, L.A., Kilburn, D.G., Miller, R.C.J., Warren, R.A.J. and Gilkes, N.R. (1986) Direct NMR determination of the stereochemical course of hydrolyses catalysed by glucanase components of the cellulase complex. *Biochemical and Biophysical Research Communications*. **139**, 487-494.

Wolters-Arts, A.M.C., van Amstel, T. and Derksen, J. (1993) Tracing cellulose microfibril orientation in inner primary cell walls. *Protoplasma*. **175**, 102-111.

Xu, J.-H. and Messing, J. (2008a) Diverged Copies of the Seed Regulatory Opaque-2 Gene by a Segmental Duplication in the Progenitor Genome of Rice, Sorghum, and Maize. *Mol Plant*. **1**, 760-769.

Xu, J.-H. and Messing, J. (2008b) Organization of the prolamin gene family provides insight into the evolution of the maize genome and gene duplications in grass species. *Proceedings of the National Academy of Sciences*. **105**, 14330-14335.

Yahraus, T., Chandra, S., Legendre, L. and Low, P.S. (1995) Evidence for a Mechanically Induced Oxidative Burst. *Plant Physiol.* **109**, 1259-1266.

Yang, Z., Gu, S., Wang, X., Li, W., Tang, Z. and Xu, C. (2008) Molecular Evolution of the CPP-like Gene Family in Plants: Insights from Comparative Genomics of Arabidopsis and Rice. *Journal of Molecular Evolution*. **67**, 266-277.

Yang, Z. and Nielsen, R. (2000) Estimating Synonymous and Nonsynonymous Substitution Rates Under Realistic Evolutionary Models. *Mol Biol Evol*. **17**, 32-43.

Yokoyama, R. and Nishitani, K. (2004) Genomic basis for cell-wall diversity in plants. A comparative approach to gene families in rice and *Arabidopsis*. *Plant Cell Physiology*. **45**, 1111-1121.

Yoshida, K., Imaizumi, N., Kaneko, S., Kawagoe, Y., Tagiri, A., Tanaka, H., Nishitani, K. and Komae, K. (2006) Carbohydrate-binding module of a rice endo-β-1,4-glycanase, OsCel9A, expressed in auxin-induced lateral root primordia, is post-translationally truncated. *Plant and Cell Physiology*. **47**, 1555.

Yoshida, K. and Komae, K. (2006) A rice family 9 glycoside hydrolase isozyme with broad substrate specificity for hemicelluloses in type II cell walls. *Plant and Cell Physiology*. **47**, 1541.

You, J., Cohen, R.E. and Pickart, C.M. (1999) Construct for High-Level Expression and Low Misincorporation of Lysine for Arginine During Expression of ET-Encoded Eukaryotic Proteins in *Escherichia coli*. *Biotechniques*. **127**, 950-954.

Zheng, C., Zhu, Q. and Sankoff, D. (2006) Genome Halving with an Outgroup. *Evolutionary Bioinformatics* **2**, 295-302.

Zheng, Z., Xing, Y., He, X.-J., Li, W., Hu, Y., Yadav, S.K., Oh, J. and Zhu, J.-K. (2010) An SGS3-like protein functions in RNA-directed DNA methylation and transcriptional gene silencing in Arabidopsis. *The Plant Journal*. **62**, 92-99.

Zhou, H.-L., He, S.-J., Cao, Y.-R., Chen, T., Du, B.-X., Chu, C.-C., Zhang, J.-S. and Chen, S.-Y. (2006) OsGLU1, a putative membrane-bound endo-1,4-β-D-glucanase from rice, affects plant internode elongation. *Plant Molecular Biology*. **60**, 137.

Zilberman, D., Cao, X. and Jacobsen, S.E. (2003) ARGONAUTE4 Control of Locus-Specific siRNA Accumulation and DNA and Histone Methylation. *Science*. **299**, 716-719.

Zilberman, D., Cao, X., Johansen, L.K., Xie, Z., Carrington, J.C. and Jacobsen, S.E. (2004) Role of *Arabidopsis* ARGONAUT4 in RNA-directed DNA methylation triggered by inverted repeats. *Current Biology*. **14**, 1214-1220.

Zuber, M.S., Kang, M.S. (1978) Corn lodging slowed by sturdier stalks. *Crops Soils*. **30**, 13-15.

Zuo, J., Niu, Q.-W., Nishizawa, N., Wu, Y., Kost, B. and Chua, N.-H. (2000) Korrigan, an *Arabidopsis* endo-1,4-β-glucanase, localizes to the cell plate by polarized targeting and is essential for cytokinesis. *Plant Cell*. **12**, 1137-1152.

Appendix A

PRIMERS

Table 1: 3'SMART RACE Primers. F and R indicates forward and reverse primers respectively. For the forward gene specific primers, the first listed primer is the outside primer and the other primers, if there is more than one, are nested primers.

Gene/EST/contig	Primer2
BM369748	QF-CGGGAAAGATTATGCTGATTGGGGAAGGC F-GTCGATGAGCTCCTCTGGGCAGCAAGTTGGC
TC144321	F2-GCCCTACAAGCACAGCAAACAGTTCATCGGGTG F-GGCCGCCCGAGCTCCAACCCCAACGTGC
BM469939	QF-GGCGGTGCGGTGGGCGACGGACTACCTGAT F2-GGACATGGACACCCCTCGCACCGTCTACAAG
TC136010	QF-CATACTGTCGTGGAGCCTGCTTGAGTACGG
BM377061	F3-GCGTCAATGGCGAAGCTCGTTTCTGGACAC QF-CAAGTGGGGCACGGACTACTTCATCAAGGC
TC130995	F-GCTCCAACATGAACAAGAACCCCGGCGGC F2-GTGACCAACGCCGTCTTCCTCCTGACCGTC
BM442502	QF-TGTCGTGGAGCGTGGTGGAGTTCGGCG F-CCCTGTCCGACGGCAACGCCTCCAATGTGGACC
TC153576	F4-CCCACCTCCTCCTCCACCACGCCCAG QF-CGGGGCGGGCGAGCTACCTGGACTACGTG F-CACTACCTGTGCGCGTGCATGGGCAAGAAC F2-CACTACCTGTGCGCGTGCATGGGCAAGAAC
BI956782	F-GACCTGGACCGGGGCGCCATGATGC
BI956376	F-GCATGTCCTACATGGTCGGCTACGGC F2-GGCAGCTCCATCCCGTCGGTG
BI779097	F-GAACGCCGACAACTTCATCTGCACCCTCCTCC F2-CGTCGCTGTCGTTCCTCCTGCTCGCCTAC
TC153055	F-GCTGCTGTGGGCGGCGGCGTGGCTGC
3' RACE CDS Primer A	AAGCAGTGGTATCAACGCAGAGTAC(T)30V N (N = A, C, G, or T; V = A, G, or C)
Universal Primer A Mix (UPM)	R long-CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT R short- CTAATACGACTCACTATAGGGC
Nested Universal Primer A (NUP)	R-AAGCAGTGGTATCAACGCAGAGT

Table 2: primers used for production of probes for hybridisation of membranes produced from a Morex barley BAC library

Gene	Forward primer	Reverse primer	Size
HvCel3	CGCTCCAGTTCTTCCAGGTCCAG	CACAAATGACGGCCTCAGTGTC	859
HvCel5	GGACGACCACTACTGCTGGAACC	GGACTCCATGTAGTCGGCATAAAGC	660
HvCel11	CAAAGGGTCAGCTGGAGGGC	GTGGACCAGCCCGTGCCG	670
HvCel14	GGGGCACCGACTACTTCATCAAGGCG	CCATCCTTCCTGTGTATCAAGAAACCTC	1300

Table 3: primers below were used for barley QPCR and BAC identification. Primers HvCEL1 to HvCEL14 were used for QPCR and identification of BAC clones that contained endo-β-1,4,D-glucanase genes. Cells highlighted in yellow were used for BAC clone identification only. Size refers to amplified product size.

Gene or EST	Forward Primer	Reverse primer	Size
HvCel1	CAAATGGAGGGAGCTTGGAAAC	TGATCCGTCATACCGATACAGAAGTAG	169
HvCel2	CATGTTCATAGGAGCGGATTGC	GCACTTTGCACACATCCACACA	118
HvCel3	TGGTTCCAAGTGACGGTAATGCG	CACCAGCAAACCAGATAGGAGTTGACC	174
HvCel4	GCCGTACCGCAAGAGTATTG	CAGCCCATACACCCTCCTGA	151
HvCel5	CGGTGTTGAGCAGATCCAT	CAAGAAACATCTACAGTAGTACAGGC	194
HvCel6	CCAGTGGAGATGACGGTGTAA	CCACAGGTACACGAGGGAAAC	138
HvCel7	GCATTATCGGTCCCAAGATTGTTCGT	GCGAGCTGCAACAACAACCTGAGT	219
HvCel8	CTACTCGCTGCGGGCAAT	ATCGGCATCCGTTACACGTAGAAAT	175
HvCel10	CCACAAGTACTGCGGCGGAGTG	GCTTGTCAAATGCGTCCACCC	196
HvCel11	GCTCGTCTGAACTGAAACCAATG	CACTTTATAGCGCAACAATACCATTC	202
HvCel12	CTCTCGGCCAAGTGTAGATAAAGGTG	GTTCCTTTCTTTCACTAAATAAATGTCTCTC	141
HvCel14	CGAGGTGTAAATGGATTAGATAGGGG	CCTTTTCTCGTATGTACAGGCAGCG	211
TC130995	GCTGATTGAGGAGAAGGAGCTGAG	GGTGTCGGCGTCCTCGTCC	274
TC153576	GGGCGGGCGAGCTACCTGG	GGCGGCGGCGTTCTTGC	231
779097	CTCCAACCACCCCCAGATCG	GGCTGGCGTAGTTGGCGTAG	207
469939	CGGTGTACGTGCAGGTGGG	GCTTGGAGTAGGCGGGGTC	197
TC153055	CTACTCCCAGCGCCTCCTCGACCG	GAGCACCGCCTCGTTCTTCTTGATG	220
442502	CGGTGTACGTGCAGGTGGG	GCTTGGAGTAGGCGGGGTC	197

Table 4: primers used for maize QPCR of the endo-β-(1,4,)-glucanase genes. Size refers to product size.

Gene or EST	Forward Primer	Reverse Primer	Size
ZmCel1	GGGAGAGTGATAGATGACGAACGC	CAGGCAAGGGAGTGGCTCTGG	136
ZmCel2	CGCAGCACTGATGAACGGG	TACGAGTACAAGCAGTAGGACCAATC	185
ZmCel3	CCTATTCTTTTGTTCATTCTTTTCTACGG	CCTATTCTTTTGTTCATTCTTTTCTACGG	149
ZmCel7	CCGTGTGGAAGCCTTGAGAG	GCGTTTCTGTTAGACACCCTCG	184
ZmCel10	CCCTAGCGTTAATTGTTGCCG	GAGCCGTGGAGCACAGTTTG	208
ZmCel11	CTGGAGCGATGCGTGTACGAG	GTACACACACAACAGAAGAAACGTGG	208
ZmCel12	CATGGAAGCCATGACCACATACG	CCATCCGTTCATTAACTGTAATCCAC	149
ZmCel14	CAAAGCATCAACACAGGGAAGTGG	GACCAAACAAGTTCAGTGCAAACCC	163
ZmCel18	CGCAGCCTCTGAATCTTTCTTGG	GACCAAACAAGTTCAGTGCAAACCC	182
ZmCel19	GTTTGGTGAGTTCGTTTAGCTTTG	CCCTCTTGGTTCGCATTCC	180
ZmCel20	CGTCCTGTCCTGAGTAGTGTCCCTG	GACAATGAGGATGACCGAGGTTG	124
ZmCel31	GACCCGAGCTTCGTCCTGTG	CCTGTTGGTCACCGTCACTGC	143

Appendix B

Sequences used to probe Morex barley BAC library.

>HvCel3_probe

GGCGGGGGGACTCTGCGCTGACCGATGGGAAGGAAGCCAAGCTGGAT
CTCTCCAAGGGGCTGTACGATGCCGGGGATCATATGAAGTTCACCTTC
CCCATGGCCTTCACGGCGACGGTGCTCGCGTGGTCGGTGCTCGAGTAC
GGGGATCAGATGACCGCCGCCAAGCAGCTGGACCCAGCGCTTGATGC
GCTCCGGTGGGTCACCGATTTCCTCATCGCCGCACATCCTTCGGACAA
CGTGTTCTACATTCAGGTTGGTGATCCTGATCTAGATCACGACTGCTG
GCAAAGGCCAGAAACAATGACTGAGAAAAGGCCGCTCACACAGATT
AACCCAAAGAACCCTGGGTCAGATATCGCTGCTGAGGCAGCAGCAGC
CCTGGCGGCAGCATCGATGGTTTTCAAATCCAGTGACACTACATATTC
TGATTCACTTCTTCAACATGCCCAGAAGTTATTTACTTTTGCTGATACT
TATAAAGGCATCGCCAGCGACACCTATCCAAAGCTTCAGAAGTACTA
TAATTCTACTGGTTATGTCGATGAGCTCCTCTGGGCAGCAAGTTGGCT
CTATCATGCCACAGGAGATCAGACATATCTCAGTTATGTGACTGTACA
GAACGGGAAAGATTATGCTGATTGGGGAAGGCCCACATGGTTCAGCT
GGGATGACAAGCTCCCAGGGACACAGATCCTTTTGTCCAGAGTTAATT
TCTTTGGATCCAAAGAGATATCTAATGCTGAAAACAGTGGACTCAAA
ATGTACAGAGACACTGCTGAGGCCGTCATTTGT

>HvCel5_probe

GGCAGCACGCCGGACGACCACTACTGCTGGAACCGGCCGGAGGACAT
GGCGTACCCGCGGCCCGTGCAGCAGGCGAGCGCGGCCCCGGACCTGG
GCGGCGAGGTGGCGGCGGCGCTGGCGTCGGCGTCCATCGTGTTTCGC
GACAACGCCGCCTACTCCAAGAAGCTGGTGAGAGGCGCGGCGACGGT
GTACAAGTTCGCGCGCGCCAACGGCAAGCGCACCCCCTACTCCCGGG
GCAATCCGGCCATCGAGTACTACTACAACTCCACCAGCTACTGGGAC
GAGTTCATGTGGAGCGCCTCCTGGATGTACTACGCCACCGGCAACAA

CAGCTACATCAGCTTCGCCACCGACCCCAGGCTCCCCAAGAACGCCA
AGGCCTACTTCAACATCCTCGACTTCTCCGTCTTCAGCTGGGACAACA
AGCTCCCCGGCGCCACCCTGCTGCTGTCGCGGCTGCGCATGTTCCTCA
ACCCCGGCTACCCCTACGAGGAGTCGCTCATGGGATACCACAACGTC
ACCAGCATGAACATGTGCATGTATTTCCCCAAGTTCAACGCGTTCAAC
TTCACCAAGGGCGGGCTCGCCCAGTTCAACCACGGCCAGGGCCAGGC
GCTGCAGTACGCCGTCGCCAACTCCTTCCTCGCCGCGCTTTATGCCGA
CTACATGGAGTCC

>HvCel11_probe

CAAAGGGTCAGCTGGAGGGCCAACTCCGGCCTGCTCGACGGGAAGGC
CAACGGGGTGGATCTCACGGGCGGCTACTACGACGCCGGCGACAACG
TCAAGTTCGGGTTGCCGATGGCGTTCACGGTGACCATGATGGCGTGG
AGCGTCCTCGAGTACGGCGGGCAGATGGAGGCGGCCGGCGAGCTCGG
CAACGCCCTCGAGGCCGTCAAGTGGGGCACCGACTACTTCGTCAAGG
CGCACCCGGAGGCCAAGGTGCTCTATGGAGAGGTGGGAGACGGTGAC
ACCGACCACGGCTGCTGGCAGCGGCCGGAGGACATGTCGACGAGCCG
TCAGGCGTACCGCCTCGACCCCCAGCGCCCGGGGTCCGACCTCGCCG
GGGAGACCGCCGCCGCAATGGCCGCGGCGTCGCTCGTGTTCCGCCGC
CCCAACCCGGGTTACGCCAACCAACTCCTGCAACACTCCAAGCAGCT
GTTTGATTTCGCCGACAAGTACCGGGGAAGGTACGACGCCAGCATCC
CCGTGGCGAGGAATTTCTACCAATCAGTCAGCGGATACAGCGACGAG
CTGCTGTGGGCGGCGGCGTGGCTGTTCCAGGCGACCGACGACAGGCG
CTACCTGGACTACGTTGCCAACAACGCCGACGCGATGGGCGGCACGG
GCTGGTCCACC

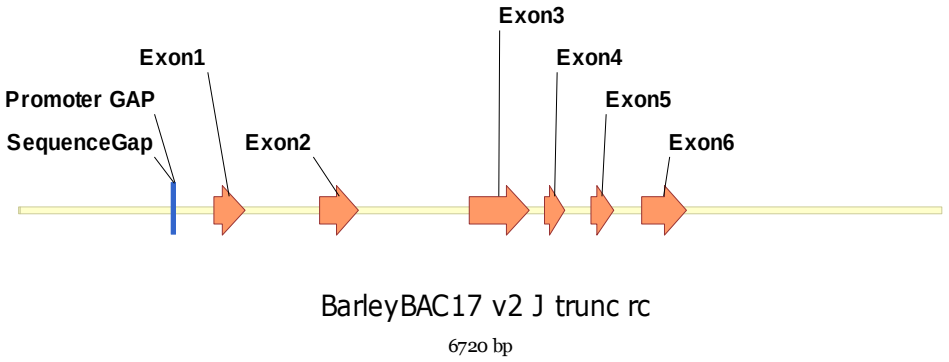
>**HvCel14_probe**

TGGGCCGAGGTGGGAGACGGCGACACTGATCACTACTGCTGGCAGCG
GCCGGAGGACATGACTTACGTCCCGGCAGGCGTACAAGGTCGACCGC
GACCACCCGGGCTCCGACGTCGCCGGCGAGACGGCCGCCGCCATGGC
CGCCGCCTCCGTGGTCTTCCCCAAATCCAACCCGCACTACGCCCACCT
CCTCCTCCACCACGCCCAGAAATTGTTCGACTTCGCGGACGGGTACAG

GGGCTAATACGACAGCAGCATATCGGAGGTGAAGAGCTACTACGCGT
CGGTTTTTTTTTTTTCAAGGACGAGCTGCTGTGGGCCGCCCTGTGGCTC
CACCTCGCCACTGGCAATCCCACCTACCTCGATTACGTCGTCGATAAC
GCCCTCTTACTTCTTCCCGCACCGGCTGGGCCGTCACCTTATTCAGCTG
GGACGTCAAGTACGCCGGCGTCCAGATCCTCGCCGCAAGGCTGCTGC
TGAGGGGAGAGCACACGGCGGAGCAGAAGAGGACGCTGGAGCAGTA
CCAGGCCAAGGCGGAGCACTACGGGTGCGCGTGCCTGGGCCGGAACG
CGGCCGGCGGCGAGGACGCCAACGTGGAGCGGAGCCCGGGAGGGAT
GCTCTACATCCGGCAGTGGAACAACATGCAGTACGTGACCAACGCGG
CGTTCCTGCTCTCCACCTACGCCGACTACCTGGCGGGGGCCGGCGTGG
GGACGGTGGCCTGCGCCGGCGGCGAGACGGCGGGCGCCGGCGAGGT
GGCCGCGCTGGCCAGGGCGCAGGTGGACTACGTGCTGGGCAGCAACC
CGAGGGGGGTGAGCTACCTCGTGGGGTACGGCGCCAAGTACCCCGCC
AGGGTGCACCACCGCGCGGCATCCATCGTGCCCTACAAGCACAGCAA
ACAGTTCATCGGGTGCACGCAGGGGTTCGACCACTGGTTCGGCCGCC
CGAGCTCCAACCCCAACGTGCCCGTCGGCGCCATCGTCGGCGGGCCC
GACCGGAGGGACAGGTTCAGGGACAACCGGGACAANTACATGCAGA
CGGAGGCGTGCACGTACAACACCGCGCCCGTGGTCGGGATGTTCGCC
AAGCTCAACCGGATGGCCAGGGAGGAGAGGGTGGCGGCCGAGGTGT
AAATGGATTAGATAGGGTTATAGGTCCAGGGGGCTGTACAATTTTCCC
CATTCTTTTGTTCATTTTTCTCTCCTTCTTTGG

Appendix C

HvCEL3_BAC17



HvCEL3_BAC17_sequence

caactaaatccttccgggccccgccgccagcaacccggacctcccacccggttcttcccgaacgtgcacccagcc
gtttgttcctcagtccctgccaccccaccccaccaacccgcgccgcgcggcgagcgagcttagttaaccgcagctc
cagcagcaggccccgacgcgcgcgagcagccggaggaaggaggtttctgagcgggcggcgaatgcgatagga
ggaagggaggggagggcgagatggggtccaagaccaaggggtgctgcgggtggctgatcgtggcgctggtgct
ggcgctggtcgccacggcggcggtggtggccatcatgaagcggaagcccggaggaggcggaggggccggga
gccggccgcccaagccgctccccgtccccggcccgcccggggacgtcgactccaagtacggcgacgcgctcg
gggtcgcgctccagttcttccaggtccagaagggtgagtgagtctccgtcccccgcattgccgccgccgcgatccc
gtgccgaagcattgctccgggcatcgaggcttggtcggtgccgtcgggctctgctttgcttagattcattctcgtatcg
tgtgccatctcctctccccggtggtcggtggagaggccaatcctctcttcgtctctgctgtctttcggtggactacgctg
gacggctggaggatttcgtgctaattcgagcgtgcatttcctcatggttttgacctctatttggaaaactgctcccacttc
ttctgttctgtgactgctcgaagttgctggaaatgcgggctggctcgcagatttcactcaaatctctcatctagctgtga
attttagctgcttctgatgccagagctgctcctttagctattggagactgctgctttggcaatcccatgctccaatgccat
gcttgggggtgtcggctgagcgaaaccgacaatctctgctgctgcttcagctctgatccgtccaagaaagccttgtt
aacttctcttctgttctgcagcggggaagttggtggacaaccagatcccgtggcgggggactctgcgctgaccgat
gggaaggaagccaagctggatctctccaaggggctgtacgatgccggggatcatatgaagttcaccttccccatgg
ccttcacggcgacggtgctcgcgtggtcggtgctcgagtacggggatcagatgaccgccgccaagcagctggac
ccagcgcttgatgcgctccggtgggtcaccgatttcctcatcgccgcacatccttcggacaacgtgttctacattcag
gttagaaatgttgtccacttgtttgcctctaggtaaagagtttgaggcatattttgggaattactaggatgaatctttggttt
atgggctttaattgtgggtatattttcattggatctatctttggttaatggcagttcagtagaggtgagagtttggtaggta
gggaaacaagagtccaggtgagatgtcatgatcaggacacctaatgtttaggagccacttagttttctgcttcatatgtt
caaaatgacacatttgttgtgatccttggcactgttggacaagggcaaacctaatgttccttccttcccgtccttccattt
ccttttgtcttcatagctgcattccttcattatgtggtaccacaagacaaatgtttttaaaactgaatctgagttgtactttat
actgttttataaaccttaaaatcatataaagtttaactgttcatagtaggggatctcaagtgtttcagtcaactgagtttagt
cagagatactgaacagggtgtaaactgcatttcatactgaagaataatgttttctagcctatagcctgttggctaaggta
aacaattcagtttaagcggcagttcagtgaaatcagatctaaacattaataaatcaggcaactgaatcactttattcctgt
ttgtccatctaaaaaaataatatcatttgggtagccaatttcagcttcaggcgttgataaattcttgtaaattcagaacagc
aagagatgcatgattgtcttttagtttaaccgctggaaccagcatttgaaaagttgttaccttgtaggttggtgatcctgat
ctagatcacgactgctggcaaaggccagaaacaatgactgagaaaaggccgctcacacagattaacccaaagaac

cctgggtcagatatcgctgctgaggcagcagcagccctggcggcagcatcgatggttttcaaatccagtgacacta
catattctgattcacttcttcaacatgcccagaagttatttacttttgctgatacttataaaggcatcgccagcgacaccta
tccaaagcttcagaagtactataattctactggttatgtcgatgagctcctctgggcagcaagttggctctatcatgcca
caggagatcagacatatctcagttatgtgactgtacagaacgggaaagattatgctgattggggaaggcccacatgg
ttcagctgggatgacaagctcccagggacacaggtttgcttgcatgaactaagccgtgtgaattattacctcagcaga
gttactttcaaattctctcttcttacaaggataagtaagtaccttaagtgctcttgtctattctgtgcagatccttttgtccag
agttaatttctttggatccaaagagatatctaatgctgaaaacagtggactcaaaatgtacagagacactgctgaggcc
gtcatttgtgggcttctaccagattctccatcagccactggtagcaggactagcggtatgttgaaatttctatttgttgatt
tggttttatgagtccatgtagggatgttttatgcattatgcattatgcattatgggtgcccgtcatgttcgctatcggcata
gagaatgtcataatatgtttattgtaatatgtattacacttgtgcatagtttcaccatatcatcatattcccaatgcaggagg
gctggtatggataagtgggtggaattctctgcagcatggcataaatgcagcattccttgctgttgtttacagcgactac
atgctgacatcacaaacagccgcagtggagtgcagtggaaagtattttagcccgacagatattcgcaatttcgctata
tcacaggtgagctaaggaaaaacttgcaaactaccattctgtcaatccatttctttatctgtagtatttcagtaaaccatct
tgctgttgacgataattgtgcgtttcgccatgacccatgggcaaattaggggcgcgatgtctgtctttcatcatgtgtgc
tgaagggtgtatgtatctgatttattatgccatgtcgttttgtcaggcaaattacgtactgggtgacaacacaatgaagct
tagctacctggtgggatatgggagcaactatccacagcaggtacaccataggggggcatccatccccgcggatgc
gaaaacgggctgcaaggggttcgagtacctccactcatccagcccgaacccgaatgtcgccatgggggccgttgtt
ggcgggccatttcagaatgacactttcgtggacttgcgcgacaatgcggtgcagacagagtccagcacttacaaca
gcggcaccttcgtcggcctactggctggcctggtcaccacctcgtccgtcgcaaagacattcacctagagctctttct
caagcaacgacaatggtggcttttatcaacccgggaagcatatacaattagaggaatgtttttttttttttggttccaagt
gacggtaatgcgcaatccagtgagacctgtttcagatttgttcattggccgtcacttaggtgggtgattagcagtagtc
attagtgccacattactcttgcttgtataatacatatatggtaatttttaccatgtggtcaactcctatctggtttgctggtgt
acctgttgtagctatgggtattgtatttgtagggcagatatcatttgttcattcagaggtggtttttggtacgtagccctgat
gattgttctgtaatgccttaccaatggaacaagcatatattcacgtgcgtgtgtaattttcttacactatgttccatgcaca
cagggaacacctgtcaacttctgcagggaaaaaatatggttcatcaggatctcatgtacagcgtgatgactagtattgt
gtgattagtgttggaaaagccacatggtccacatgggagcacgg

>HvCel3_ORF BAC17

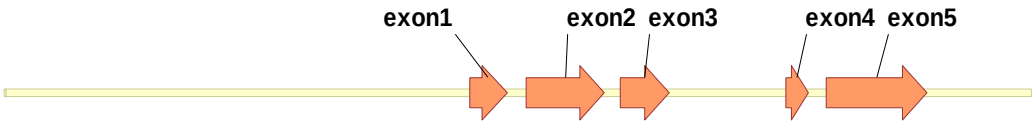
atggggtccaagaccaaggggtgctgcgggtggctgatcgtggcgctggtgctggcgctggtcgccacggcggc
ggtggtggccatcatgaagcggaagcccggaggaggcggaggggggccggagccggccgcccaagccgctcc
ccgtccccggcccgcccggggacgtcgactccaagtacggcgacgcgctcggggtcgcgctccagttcttccag
gtccagaaggcggggaagttggtggacaaccagatcccgtggcgggggggactctgcgctgaccgatgggaagg
aagccaagctggatctctccaaggggctgtacgatgccggggatcatatgaagttcaccttccccatggccttcacg
gcgacggtgctcgcgtggtcggtgctcgagtacggggatcagatgaccgccgccaagcagctggacccagcgct
tgatgcgctccggtgggtcaccgatttcctcatcgccgcacatccttcggacaacgtgttctacattcaggttggtgat
cctgatctagatcacgactgctggcaaaggccagaaacaatgactgagaaaaggccgctcacacagattaaccca
aagaaccctgggtcagatatcgctgctgaggcagcagcagccctggcggcagcatcgatggttttcaaatccagtg
acactacatattctgattcacttcttcaacatgcccagaagttatttacttttgctgatacttataaaggcatcgccagcga
cacctatccaaagcttcagaagtactataattctactggttatgtcgatgagctcctctgggcagcaagttggctctatc
atgccacaggagatcagacatatctcagttatgtgactgtacagaacgggaaagattatgctgattggggaaggccc
acatggttcagctgggatgacaagctcccagggacacagatcctttgtccagagttaatttctttggatccaaagaga
tatctaatgctgaaaacagtggactcaaaatgtacagagacactgctgaggccgtcatttgtgggcttctaccagattc
tccatcagccactggtagcaggactagcggagggctggtatggataagtgggtggaattctctgcagcatggcata
aatgcagcattccttgctgttgtttacagcgactacatgctgacatcacaaacagccgcagtggagtgcagtggaaa
gtattttagcccgacagatattcgcaatttcgctatatcacaggcaaattacgtactgggtgacaacacaatgaagctta

gctacctggtgggatatgggagcaactatccacagcaggtacaccataggggggcatccatccccgcggatgcga
aaacgggctgcaaggggttcgagtacctccactcatccagcccgaacccgaatgtcgccatgggggccgttgttg
gcgggccatttcagaatgacactttcgtggacttgcgcgacaatgcggtgcagacagagtccagcacttacaacag
cggcaccttcgtcggcctactggctggcctggtcaccacctcgtccgtcgcaaagacattcacctag

>HvCel3_amino acid _BAC17

mgsktkgccgwlivalvlalvataavvaimkrkpgggggrsrppkplpvpgppgdvdskygdalgvalqff
qvqkagklvdnqipwrgdsaltdgkeakldlskglydagdhmkftfpmaftatvlawsvleygdqmtaakql
dpaldalrwvtdfliaahpsdnvfyiqvgdpdldhdcwqrpetmtekrpltqinpknpgsdiaaeaaaalaaas
mvfkssdttysdsllqhaqklftfadtykgiasdtypklqkyynstgyvdellwaaswlyhatgdqtylsyvtvq
ngkdyadwgrptwfswddklpgtqillsrvnffgskeisnaensglkmyrdtaeavicgllpdspsatgsrtsgg
lvwisgwnslqhginaaflavvysdymltsqtaavecsgkyfsptdirnfaisqanyvlgdntmklsylvgygs
nypqqvhhrgasipadaktgckgfeylhssspnpnvamgavvggpfqndtfvdlrdnavqtesstynsgtfvg
llaglvttssvaktft*

HVCEL4_BAC242



BAC242 HvCel4 rc intronexon

5360 bp

HVCEL4_BAC242_sequence

cctcaaaagaaggaaaagtgcaacattttcctggactttgatcacaacaaaaccctgtcaaaatccccgcagaaaga
cacattccaaaaccctcttatccattcggtctggctaaacgaagaaacgaaaccacattatacgcactcctcgacatttt
cctctggagtactatacgagcatgcttgaatgccatgccacattttctctgatggagcaacggccaaaacgaaccaaa
ccaacggcacagacggctgcaactgcagagcgagtactagcagtagttctgtggagtaatgcagcagcacattca
agccttcagctccaggccaaccaagccaccggcgccgcaacgttcgccacgtccacgtgcagccgccgcattgg
cctctgccgggcatcctaggcctccttcccctcccgtcactccggcctgcgcccgtcacatgccgtatttaacctccc
gcccgttcggctccgacaccacctcactcacctctcctccgccggccactattaaaccacggccctgctgactccag
tagacgcacacacactcgctcctcgctctgtcactcagccgccgagagcaccgatccaacggtgctgcattgccatt
gcattagcttaatttcttccggtcagccgcagcaccggtcttgccgttggcggtcagggcggcaatggggctgtcgt
ctgcgcttgtgctggcgctgctgctcttgccggcggtggcgtcggccgggcatcacgactacggcgacgccctcc
acaagagcattctcttcttcgagggtcagcgatccggcaggcttccgcccgaccagcgccttcgctggcgccgcga
ctccggcctacacgacggcgccgccgccggcgtacgactcgcatctcatcttacaatgttctgcttcgtttatactgta
ttttcttttgtgcaaaagtggtgatggttccatttccatggcaatgcgctgcaggtggacctgacgggagggtactacg
acgccggcgacaacgtgaagttcgggttcccgatggcgttcacggcgacgctcatgtcgtgggggctgatcgactt
cggccgcagcttcgggccgcacaaggaggaggcgaggaaggcggtgcggtgggcgacggactacttcatgaa
ggccacggccaagcccaacacggtgtacgtgcaggtgggcgacgccttcaaggaccactcctgctgggagcgg
cccgaggacatggacaccccgcgcaccgtctacaaggtcgacccctcccacccgggctccgacgtcgccgccg
agaccgccgccgccctcgccgccggctccatcgtcttccgcgacgccgaccccgtctactcccagcgcctcctcg
accgcgccatggccgtaagtcgcaaactgtctccactctgacgactcaaatgtgccgtgcgtgtcagctcagctctg
acatgaaaaaaatgagcaggtgttcaagttcgcggacaggtaccggggcgcctacagcagcagcctgcacgcgg

cggtgtgcccctgctactgcgacttcgacgggtaccaggacgagctgctgtgggcggcggcgtggctgcacaag
gcgtcgcggaggagggagtaccggcagtacatcaagaagaacgaggtggtgctcggcgccagcgactccatca
acgagttcggctgggacaacaagcacgccggcatcaacgtcctcatctccaaggtgatatcccttccgcatgcagct
ttggcctttgtatgcagtacagaactacagatgctgataatgataaaagagtacacataagctagcaagccactcgaa
tctaatgattgactgtcttgcctgctgctttacgccttttaagctttacaaatgtggattcagttagtttcttatgatgggcac
atacgtcttgccaacggatcagccattctgaaattaccctgactggtagatgctactgattcctgaatctggatgatata
ccccggttcagattagattaatcaatgaccagtcagtagtaaacaaatggagtgtgtatagtgcacggtggtggcgta
ccagcaaataggagacttcacgaggaggatacttgtctcttctatcttggcctctcctacctaggcccctgtgctccgc
atgctgccggcgcggcaagcggtagcggcacttttcagaggctctcacgggtctccactctcgactcgtcatgcca
gtcatgtccttcctgccattcgaatcaagttacaaccagctaataaaactggatcatagtactgcacttgctgtcagtact
aatatgttgttgtgttcttgattgatcgaccaggaggtcctaatggggaaggacgagtacttccagtccttccgcgtca
acgccaacaacttcatgtgcaccctcctccccggcatctccgaccacccccagatccagtactccccaggtttgtttc
gccacaaacacagcaagaaaacatgaggttcatcttgtgaacaactgggaggagtgatcactgatcaatcaagaaa
cggaatgcaggcgggctgctgttcaaggtcgggggcagcaacatgcagcacgtgacgtcgctctccttcctcatcc
tggcctactccaactacctgagccacgcgggcgcgcacgtctcctgcggcggcggcggggcggacggcgccgc
ctacgaatctccggcagggtggccaaccggcaggtggactacatcctcggggacaacccgctgcgcatgtcctac
atggtcggatgtcgagaccgcggtttcccgctccggcatccaccaccgcggcagctccatcccgtctgtggcagcg
cacccggcgaggatcgggtgcaaggccggggcggcgtactatgcgaccgcggcgcccaacccgaacctgctg
gtgggcgccgtcgtcggcgggccgaccgacgccaccgacgctttccccgacgcgcgcgccgtattccagcagtc
cgagcccaccacgtacatcaacgcgcctctcatgggcctcctcgcctacttctcagcccaccccaaccccgccgag
tgggccgacgactaacactcgccatctagacacagggaaataaaagatccaaaaagatctctctctgactgacttgg
ccgtaccgcaagagtattggagtagtgcatcattagccattcatttgtttggtttaataataactgtatctaattattatttatt
tagttatttatctatagccatttcctctcctgcgaaaaagaaaagatcaggagggtgtatgggctgtatctgtagtgttgtt
atacttgtaagatgtttacgtgctcatattggcactacgtaaaatgtttatcaataaaagagcttctagtttttctgacgtct
atgaaatgtttaccggtggctacagaaagatagaaaaagatgctgacctcaaataccagcccatgagatcacatggc
aaatgttgtgttataattgtaagttgtaagattttagcgtgctaagaatattgtactcactctgtctccgaatttacttctctta
gat

>HvCel4_ORF

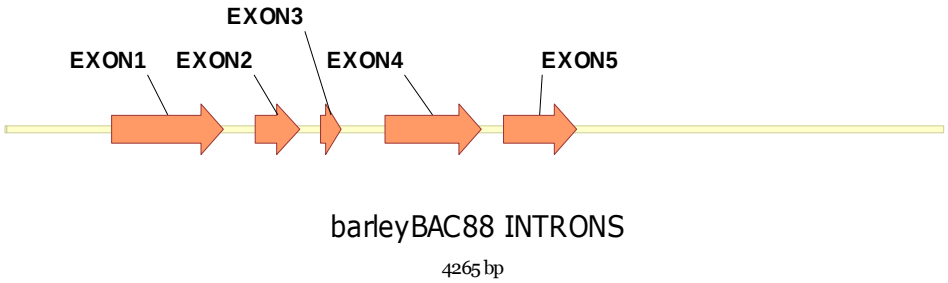
atggggctgtcgtctgcgcttgtgctggcgctgctgctcttgccggcggtggcgtcggccgggcatcacgactacg
gcgacgccctccacaagagcattctcttcttcgagggtcagcgatccggcaggcttccgcccgaccagcgccttcg
ctggcgccgcgactccggcctacacgacggcgccgccgccggcgtggacctgacgggagggtactacgacgc
cggcgacaacgtgaagttcgggttcccgatggcgttcacggcgacgctcatgtcgtgggggctgatcgacttcggc
cgcagcttcgggccgcacaaggaggaggcgaggaaggcggtgcggtgggcgacggactacttcatgaaggcc
acggccaagcccaacacggtgtacgtgcaggtgggcgacgccttcaaggaccactcctgctgggagcggcccg
aggacatggacaccccgcgcaccgtctacaaggtcgacccctcccacccgggctccgacgtcgccgccgagac
cgccgccgccctcgccgccggctccatcgtcttccgcgacgccgaccccgtctactcccagcgcctcctcgaccg
cgccatggccgtgttcaagttcgcggacaggtaccggggcgcctacagcagcagcctgcacgcggcggtgtgcc
cctgctactgcgacttcgacgggtaccaggacgagctgctgtgggcggcggcgtggctgcacaaggcgtcgcgg
aggagggagtaccggcagtacatcaagaagaacgaggtggtgctcggcgccagcgactccatcaacgagttcgg
ctgggacaacaagcacgccggcatcaacgtcctcatctccaaggaggtcctaatggggaaggacgagtacttcca
gtccttccgcgtcaacgccaacaacttcatgtgcaccctcctccccggcatctccgaccacccccagatccagtact
ccccaggcgggctgctgttcaaggtcgggggcagcaacatgcagcacgtgacgtcgctctccttcctcatcctggc
ctactccaactacctgagccacgcgggcgcgcacgtctcctgcggcggcggcggggcggacggcgccgcctac
gaatctccggcagggtggccaaccggcaggtggactacatcctcggggacaacccgctgcgcatgtcctacatgg
tcggatgtcgagaccgcggtttcccgctccggcatccaccaccgcggcagctccatcccgtctgtggcagcgcac
ccggcgaggatcgggtgcaaggccggggcggcgtactatgcgaccgcggcgcccaacccgaacctgctggtg

ggcgccgtcgtcggcgggccgaccgacgccaccgacgctttccccgacgcgcgcgccgtattccagcagtccga
gcccaccacgtacatcaacgcgcctctcatgggcctcctcgcctacttctcagcccaccccaaccccgccgagtgg
gccgacgactaa

>HvCel4_amino acid BAC242 _HvCel4

mglssalvlallllpavasaghhdygdalhksilffegqrsgrlppdqrlrwrrdsglhdgaaagvdltggyydag
dnvkfgfpmaftatlmswglidfgrsfgphkeearkavrwatdyfmkatakpntvyvqvgdafkdhscwer
pedmdtprtvykvdpshpgsdvaaetaaalaagsivfrdadpvysqrlldramavfkfadryrgaydsslhaav
cpcycdfdgyqdellwaaawlhkasrrreyrqyikknevvlgasdsinefgwdnkhaginvliskevlmgkd
eyfqsfrvnannfmctllpgisdhpqiqyspggllfkvggsnmqhvtslsflilaysnylshagahvscggggad
gaayespagwptgrwttssgttrcacptwsdvetavsrsgihhrgssipsvaahparigckagaayyataapnp
nllvgavvggptdatdafpdaravfqqsepttyinaplmgllayfsahpnpaewadd*

HVCEL7_BAC88



HVCEL7_BAC88_sequence

ttatccgaaaactcgagttctcctttcaccataaaatcaaaaccacgacaccagtaatggaagcatagccctttgtttat
caaagcctcagctggcacctgccattgccatgtgaacgagctcgcatcttgcagaaaaaaactgccccatctcattct
cgagggagagaaaaaaaatgaacaggcccgtctttaaaagcagcgctccagtgccaactataatggcagagcag
gcagagaggaggagaggtggccggtagccggtatcggtagttaaccgagcagtagtgctctgctcttctctcctcct
gcacgcggcacacttgttcgctcgctcgctcatggctgccgtcctccggtgcctcctgctggccctggccgtcctgg
ttgtccgcaatgccgccgccctgccgcaccacggccccgccaagcacgactacagggacgctctcaccaagtcc
atcctcttcttcgagggccagaggtccggcaggctcccgccgtcccagcgcgtgtcctggcgccggagctccggc
ctatccgacggctcctccgtcaaggcgagcacctccggccgttctttcttgctgcatttcccctgcattgcattgtgttc
aatggctgacattgcctcttttccccaggttgatctgaccggagggtactacgacgccggggacaacgtcaagttcg
ggttcccgctggccttcagctcgaccatgctggcgtggagcgtgctggagttcggcgggatgatgaagggcgagc
tgcagcacgcccgggacgccgtccgctggggctccgactacctcctcaaggccaccgcccaccccgacaccgtc
tacgtccaggtccctccctcccccgcccccgtcttcctcgttccggcactgttgattgattgatgcatggctcgggc
gctccttgccgcaaagtacaaagccatggaccacgcaccactcaactcactgtgagcgtgaccgactgtcgctagg
ttggggacgcgggcaaggaccacgcgtgctgggagcggccggaggacatggacacgccgaggacggtgtaca
aggtcgaccccagcacccccggctccgacgtcgccgccgagaccgccgccgcgctcgccgcggcatccctcgt
cttccgcaagtccgacccggcctactccagccggctcgtcgccagggccaagagggtacaagcaccctacacact
ccccatccaccattgcctcctcccctgccgcatctttctcgccgtgacgctgaccagctgctctgttccgcgtaggtgt

tcgagttcgctgacaagcacaggggcgtgtacagcgccaagctctcgtcatacgtctgcccctactactgttcctact
ccggataccaggtagggtcccagctgctgccaccattattgtctttttcttgctctgctttccctcccattggctggttcc
acagacctgtgagcgagagagcatgggcgccatgcatgcgctctgccctatgctcggcgtcgcctgcttttcttcctc
tcgtctcgtgccgtttgtttgtttaatgggaccgatgcatgactggaaacaggacgagctgctatggggcgccgcatg
gctacaccgggcgaccaagagccccgtctacctcagctacatcaagctgaacggccagctgctcggcgccgacg
agcaggacaacaccttcggatgggacaacaagcacgccggcgccaggatcctgctctccaagtccttcctggtgc
agaaggtgggcgcgctgcaggagtacaaggcacacgcggacagcttcatctgctccatggtgcccggcaccccc
accgaccagacgacctacacccggggcggcctcctcttcaagctcagcgacagtaacatgcaatacgtcacctcc
agctccttcctcctgctcacctacgccaagtacctcgtctccgccaagaagaccgtctcctgcggcggcgtcgccgt
caccccgcagcgcctccgcgccattgccaggcgtcaggtccgtcggcttcttgcttctccttctctaccatcgtcgtc
gtcgtcgtcaacggaagagcaggtttgatcaccatgccgtttgtttcttccttgctgcaggttgactacctgctggggag
caacccgatgggcatgtcgtacatggtggggtacggcgccaagtacccgaagaagctgcaccaccgggcctcct
cgctgccctccgtcgcggcgcacccggggaagatcgggtgctcgcaggggttcaccggcctctactccggcgcg
gccaacccgaacgtgcacgtcggcgccgtcgtgggcgggccgaaccagaacgaccagttccccgaccagcgca
gcgactacgagcactcggagccggccacctacatcaacgcgcccctcgtcggggcgctcgcctacctcgcccact
cctccggccagctctaaaaaaaaaatccagggacgcgccgcattatcggtcccaagattgttcgttgccttgattgttc
cggcgatctaattgattgttgcgatccggagaccccagtgcaagcaacactagtgggtcttgatttatccatcctactta
ctactgctctagcctctaggtaggtgttaagtacgagtgctaagtacgtccaactgtgattcatgtacaacatgtactca
ggttgttgttgcagctcgcaactcgttatgaaatcaagtctttgtgtcgagtgattctacgtttctactcggtccgtccttta
ctcggttactga

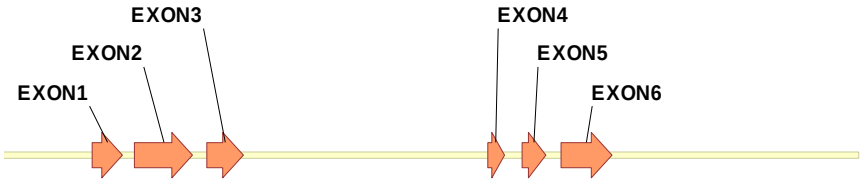
>HvCel7_ORF

atggctgccgtcctccggtgcctcctgctggccctggccgtcctggttgtccgcaatgccgccgccctgccgcacc
acggccccgccaagcacgactacagggacgctctcaccaagtccatcctcttcttcgagggccagaggtccggca
ggctcccgccgtcccagcgcgtgtcctggcgccggagctccggcctatccgacggctcctccgtcaaggcgagc
acctccggccgttctttcttgctgcatttcccctgcattgcattgtgttcaatggctgacattgcctcttttccccaggttga
tctgaccggagggtactacgacgccggggacaacgtcaagttcgggttcccgctggccttcagctcgaccatgctg
gcgtggagcgtgctggagttcggcgggatgatgaagggcgagctgcagcacgcccgggacgccgtccgctggg
gctccgactacctcctcaaggccaccgcccaccccgacaccgtctacgtccaggttggggacgcgggcaaggac
cacgcgtgctgggagcggccggaggacatggacacgccgaggacggtgtacaaggtcgaccccagcacccccc
ggctccgacgtcgccgccgagaccgccgccgcgctcgccgcggcatccctcgtcttccgcaagtccgacccggc
ctactccagccggctcgtcgccagggccaagagggtgttcgagttcgctgacaagcacaggggcgtgtacagcg
ccaagctctcgtcatacgtctgcccctactactgttcctactccggataccaggacgagctgctatggggcgccgcat
ggctacaccgggcgaccaagagccccgtctacctcagctacatcaagctgaacggccagctgctcggcgccgac
gagcaggacaacaccttcggatgggacaacaagcacgccggcgccaggatcctgctctccaagtccttcctggtg
cagaaggtgggcgcgctgcaggagtacaaggcacacgcggacagcttcatctgctccatggtgcccggcacccc
caccgaccagacgacctacacccggggcggcctcctcttcaagctcagcgacagtaacatgcaatacgtcacctc
cagctccttcctcctgctcacctacgccaagtacctcgtctccgccaagaagaccgtctcctgcggcggcgtcgccg
tcaccccgcagcgcctccgcgccattgccaggcgtcaggttgactacctgctggggagcaacccgatgggcatgt
cgtacatggtggggtacggcgccaagtacccgaagaagctgcaccaccgggcctcctcgctgccctccgtcgcg
gcgcacccggggaagatcgggtgctcgcaggggttcaccggcctctactccggcgcggccaacccgaacgtgc
acgtcggcgccgtcgtgggcgggccgaaccagaacgaccagttccccgaccagcgcagcgactacgagcactc
ggagccggccacctacatcaacgcgcccctcgtcggggcgctcgcctacctcgcccactcctccggccagctcta
a

>HvCel7 amino acid BAC88 _cel7

maavlrclllalavlvvrnaaalphhgpakhdyrdaltksilffegqrsgrlppsqrvswrrssglsdgssvkasts
grsfllhfpcialcsmadiasfpqvdltggyydagdnvkfgfplafsstmlawsvlefggmmkgelqhardavr
wgsdyllkatahpdtvyvqvgdagkdhacwerpedmdtprtvykvdpstpgsdvaaetaaalaaaslvfrks
dpayssrlvarakrvfefadkhrgvysaklssyvcpyycsysgyqdellwgaawlhratkspvylsyiklngqll
gadeqdntfgwdnkhagarillsksflvqkvgalqeykahadsficsmvpgtptdqttytrggllfklsdsnmq
yvtsssfllltyakylvsakktvscggvavtpqrlraiarrqvdyllgsnpmgmsymvgygakypkklhhrass
lpsvaahpgkigcsqgftglysgaanpnvhvgavvggpnqndqfpdqrsdyehsepatyinaplvgalaylah
ssgql

HVCEL8, 779097



BarleyBAC93 EXONS rc

6030 bp

HVCEL8_ 779097_sequence

aaattgcactatagtgataaaactgaaaaagtagttttttgcattctttatgtccaaaatgtccatagtgataaaaccgga
aaaagtgacaagccactttcttcatctgtgagaggtttgcaatttgctctttttttttcattctttgtgtccaaaatgttgcaa
ggttgtaaaccaatgattgcacgtgaatacaggtactgtattagggacgaagccatggagattccaagaaaagttata
tcgatcggtcccaagacaggcatttctgagggtttggtacaccacttttgtgctcccaatcccaaatggcagtagccct
tggttggcaccggaagcgaccacggtccacggcagcgtccacaaggaccagttccacggacgccgcggccctc
gtcctcgccccgggcgccgcaacgctgccacgcgcacgtgtcgcgccgcgactggctccgcccccagcatccca
ttccctctctggctacaccctcacatggagcctcgccctcacatgccattaaacctggccacccccctctcccccccct
attaaatcctttgccgttcgccgctccagtgcagcagcagtccagcgcccgaaccgccgcgcacgcacagcgcca
accgactaaactgccgcccatcaacaccggtgaagcctagctttagctacctagcgaggcgattgcaatggcgtcg
ccggcgaggaaggcaatggtggcggccgcggcaatgctgctgctgctgctcctgtcgacggcggccgtcgcgca
gcacaactacggcgacgccctccacaagagcatcctcttcttcgagggccagcgctccggccgcctcccgcccga
ccagcgcgtccgctggcgccgcgactccgcgctcaacgacggcgccaccgccggggtgagtgacaccactaga
accgcctcctcttctttcgcgctccattgcattgattgttaacacggcggtgtgcgtgcgtaggtggacctgacgggc
gggtactacgacgcgggggacaacgtcaagttcgggttcccgatggcgttcacgacgacgctcatgtcgtggggg
ctgatcgacttcggcaagaacttcgggtcgcacggcggggaggcgcgggcggcggtgcggtgggcgacggact
acctgatgaaggcgacggcgacgcccaacacagtgtacgtgcaggtgggggacgccttccgcgaccactcgtgc
tgggagcggcccgaggacatggacacccctcgcaccgtctacaaggtcgacccggcgcacccgggctccgacg
tcgccgccgagaccgccgccgccctcgccgccggctccatcgtcttccgcgaggccgaccccgcctactccaag
cgcctcctcgaccgcgccatctccgtaagtgcattaaagccgcctcgtcatcacgtacctgccacgtcgattagacat
aagtatgtatgttcatttttcttcttcttggttgcgatgagaaggtgttcgagttcgcggacaagtaccgtggtgcgtaca
gcagcagcctgtactcggccgtgtgcccctgctactgcgactacgacggctaccaggacgagctgctgtggggcg
cggcgtggctgcacaaggcgtcgcgccagcgcgtgtaccgcgactacatcaagaagaacgaggtggcgctcgg
cgccagcgactccatcaacgagtttggctgggataacaagcacgccggcatcaacatcctcatctccaaggtacta

cagtagtaccatgcacatttcattcctgctaggctactcctagattacggacatacatgcatgttcctgttgactaggagt
agtagggtatgttactgtacttggtgtaacacagaccacacagtaggtgcacttgattcttgcacgttgctgactagcc
atgacggccactccaccactgcattgctgtgctgacctatggtatagagtgtgaagtagtatagccaggtactctagg
atgatgcgtcgacgcagggatgcctgaccctgttttggccgcttctcgtgtgtcattgcgtgattctgccctgctctgct
ctgctcggtttgtgtacacacaaacacaagagcaacgacgacactagtcctctggtgcccagcggcgcgcacatgc
gccggcttcccgccttttttcacgagtgctagtttatcttttctttcggtagccgcagcgggcaatgcgcagtcttggc
atgccatcgtcaagtttgagacgtgccatcctgacaatcagtgtgaatatgcatctcttactaattagccgtagcagata
tactaaactagaggatttctctgtgcttggagagcagggcactctgctcgcgtagtaaagcaagagcatgcatgggc
tggaaggccggccatgaggggaacctttgcacgcgcgctctgctctttgaccaagacagacatgtggggctgggt
tgtgctagggcccaaaagcaattttaaaagatcagacgaaaggtcggctcgatatcatgatactagtttcctagaggc
tatatactcaccagatattcctttgaacgtgcacgccatgttgggtgatgacctgcgttctcttgtgtcttgtctggattaa
ctgtcagaggcgaggaataattgtcccaaggctgctaggataacaaaacttagcagatcatgtgatgtacatcacagt
gggtgattaggcaatgagatcgaacatatgtgcgggctgagcatatgcatttgttctctccattggggtttatgatgcat
caatacttgttcggcgcatcagcatgtgagctgactcgattcttctggcccgactacacagacattcagcactttctgg
catccttggctgtcacgtggcactccacatggccgtatgatcagcgatggctactcacaccgtggcagagtgtactc
atcacatattcacatgcctggctggctggctagatagattggcgcagcatgtgtgcattattgggtgcctcattggaac
gtatcctagattcccactgcacgttgctggaccttggttggggccagattctatgtggataggttaaccaaaacctact
agtatatttttatcagcgctgccagtgatgataggctggttagtggatcaaagcaggcaggcgagcatgtttgtcccg
gtcgtcagtacacactggccctggccctggcccggcccggtctgcccatggggtcgtgtgccgtgctgtaaaacag
gcgaacagtggggatcgggcagtgcgctgtgggcgacttttcagtaccgtgctcacgggcctcgatcgtagcggg
atgtgccaaagccactgtcatttttcacttatcattagtagcactactgatgaaatgttggattagtgattcggctggtcg
ctgatggtgggatctgctgctgtggtgtttggctgtgtaggaggtgctgatgggcaaggactccttcttcaagtcgttc
caggtgaacgccgacaacttcatctgcaccctcctcccggggatctccaaccacccccagatcgagtactccccag
gtcaacctcccattgccaaccccatggctagtgtaatttccacacaatttagcttgatgtatgtgtcatgtagttcagcct
gtggagaagaacctttctgaaccgtctgaatatcgacgatgcaggtggcctgctgttcaaggtgggcaacagcaac
atgcagcacgtgacgtcgctgtcgttcctcctgctcgcctacgccaactacgccagccacgccaacgtgcgggtgc
cgtgcggcggctcctccgcctcgccggtcgtgctccgccgcgtcgccaagcgccaggtccgccaccaaacccgc
cgccgtgttttcctgccggaaactcgccgcgccgcgacgacccactgacctcctcctcgttttcctctgctgctgctg
ctgctgccaggtggactacatcctgggcgacaacccgctgcgcatgtcgtacatggtggggtacggggagcggtt
cccgcggcgcatccaccaccgcgggagctcgctgccgtcggtgtcggcgcacccggcgcggatcggatgcaag
gccggcgcggcatactacggcagcagcgcgcccaaccccaacctgctcgtcggcgccgtggtgggcgggccca
gcaacaccagcgacaacttccccgacgctcgggccgtcttccagcagtccgagcccaccacctacatcaacgccc
cgctgctgggcctcctcgcctacttctccgcccaccccaacctgccccagtccgacggcaccgactgaagatatat
gtggcccgctactcgctgcgggcaatgcggtggagtggccctctttagcggtcgtgggccgaggaaggaaagcct
ggttggtagtccttgttttgtacgggctgcgatgggctgatccgtcccgctgccttgctgcccctgtattattatttctata
tatttctacgtgtaacggatgccgatccctttcctgtaaaaaatggcaggaaaaggatcggtcgtgccatgccaagaca
agcaatctaatatatctatctaccactatatcatcatcgtcagttcgtcacgcttgttgcgatgcatttacgtccaccttcc
cctacttcttgtagttcagaggtcaatgtcatcgcgagaccaaacatgtcgccggaaagctgtgggcactgcccttca
cgcaatgatgcaagctccaaggtcaactccatatctgcaatgcaatgatgcagcgttgcttagcgagcacaactccct
ttttttggggtacgaaacg

>HvCel8_ORF

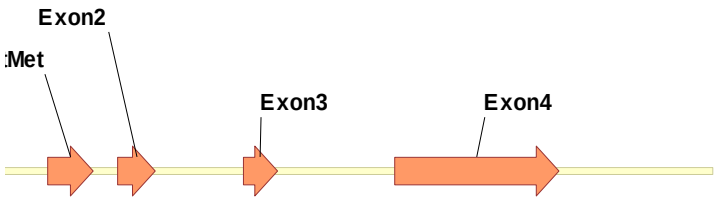
atggcgtcgccggcgaggaaggcaatggtggcggccgcggcaatgctgctgctgctcctgtcgacggcggc
cgtcgcgcagcacaactacggcgacgccctccacaagagcatcctcttcttcgagggccagcgctccggccgcct
cccgcccgaccagcgcgtccgctggcgccgcgactccgcgctcaacgacggcgccaccgccggggtggacct
gacgggcgggtactacgacgcgggggacaacgtcaagttcgggttcccgatggcgttcacgacgacgctcatgtc
gtgggggctgatcgacttcggcaagaacttcgggtcgcacggcggggaggcgcgggcggcggtgcggtgggc

gacggactacctgatgaaggcgacggcgacgcccaacacagtgtacgtgcaggtgggggacgccttccgcgac
cactcgtgctgggagcggcccgaggacatggacacccctcgcaccgtctacaaggtcgacccggcgcacccgg
gctccgacgtcgccgccgagaccgccgccgccctcgccgccggctccatcgtcttccgcgaggccgaccccgcc
tactccaagcgcctcctcgaccgcgccatctccgtgttcgagttcgcggacaagtaccgtggtgcgtacagcagca
gcctgtactcggccgtgtgcccctgctactgcgactacgacggctaccaggacgagctgctgtggggcgcggcgt
ggctgcacaaggcgtcgcgccagcgcgtgtaccgcgactacatcaagaagaacgaggtggcgctcggcgccag
cgactccatcaacgagtttggctgggataacaagcacgccggcatcaacatcctcatctccaaggaggtgctgatg
ggcaaggactccttcttcaagtcgttccaggtgaacgccgacaacttcatctgcaccctcctcccggggatctccaac
cacccccagatcgagtactccccaggtggcctgctgttcaaggtgggcaacagcaacatgcagcacgtgacgtcg
ctgtcgttcctcctgctcgcctacgccaactacgccagccacgccaacgtgcgggtgccgtgcggcggctcctccg
cctcgccggtcgtgctccgccgcgtcgccaagcgccaggtggactacatcctgggcgacaacccgctgcgcatgt
cgtacatggtggggtacggggagcggttcccgcggcgcatccaccaccgcgggagctcgctgccgtcggtgtcg
gcgcacccggcgcggatcggatgcaaggccggcgcggcatactacggcagcagcgcgcccaaccccaacctg
ctcgtcggcgccgtggtgggcgggcccagcaacaccagcgacaacttccccgacgctcgggccgtcttccagca
gtccgagcccaccacctacatcaacgccccgctgctgggcctcctcgcctacttctccgcccaccccaacctgccc
cagtccgacggcaccgactga

>HvCel8-779097 _BAC93

masparkamvaaaamlllllstaavaqhnygdalhksilffegqrsgrlppdqrvrwrrdsalndgatagvdlt
ggyydagdnvkfgfpmaftttlmswglidfgknfgshggearaavrwatdylmkatatpntvyvqvgdafrd
hscwerpedmdtprtvykvdpahpgsdvaaetaaalaagsivfreadpayskrlldraisvfefadkyrgayss
slysavcpcycdydgyqdellwgaawlhkasrqrvyrdyikknevalgasdsinefgwdnkhaginiliskev
lmgkdsffksfqvnadnfictllpgisnhpqieyspggllfkvgnsnmqhvtslsflllayanyashanvrvpcg
gssaspvvlrrvakrqvdyilgdnplrmsymvgygerfprrihhrgsslpsvsahparigckagaayygssapn
pnllvgavvggpsntsdnfpdaravfqqsepttyinapllgllayfsahpnlpqsdgtd*

HVCEL10, 153576



HVCEL10_153576_sequence

tcccccgcccatccccatttccgcacctcattcctcttcctccaacgaccgcccgaaccaccgggccacctgtccgtc
cgtccgccgcgtgcgtgcgtgcgtgcatatatcgaacaacgatcgaccggccggccgccgccggagcggagca
gcaagtccgtcgggtacgctcctcgaggaccggcgtatacaactgttggagaaagagagagtgagagagaggga
gagctatagcgatgataatgaggaggagcagcggcgcggcggcttgggttcttggcgtggccgtggcggcggcc
gcgacgttcttcctggcggcggaggcggccggcgcggcgcccgtgccgcacgactacgagcaggcgctgcgg
aagagcctgctctacttcgaggcgcagcggtcggggcggctgccgtacggccagcgcgtcgcctggcgcgacc
actccggcctcaccgacggcctcgagcaaggggtacgtccccgtcgtcgaattcttccgtcccttcttcttccttgcc

ggccggccgatcatcgtatcgttcggtcgatcaggatcagggtgctgacttaatttggatggatcttgcttgctgtgttt
gtgtttggttccattaggtggatttggtgggcggatactacgatgccggcgaccacgtcaagttcgggctgcccatgg
ccttcacggtgaccatgctgtcgtggagcttgcttgagtacggcggcgacgtggcggccgccggcgagctcgcgc
acgccctcgagtccatcaagtggggcaccgactacttcatcaaggcgcacaccaagcccgacgagctctgggcc
gaggtaatttttgtttttacaattcccagttacccttacccccatcgtaatcaccgccattcgctggctaatcatccaatca
acaacttcaaacaaacataaacaaaagctaaaaaacttgtagtacaaacatgcatgcacgcaccattggccgtgggc
tggcaaattgagttgagaaaacaaaaacaaaaacaaaaacaaaagctaaaaaacttgtagtatagattttaacttgaat
ttttttgaaagctgggggagaggaacatttgttaggtgggggctagaggacaagcaaagccaaggcgcggttcagt
cgagggatgcgcaaatctcacacctgcgcctgcacctgcacctgccactgcatgcatgattgtctcttagcctccaca
tggattagccaggaacttattaacaaatcacgacgcatatgcatgtctgtgtatctgtccgtctgcttccacggctttga
cgaaatgtgaacgaaatgattgactgatggtgctttgctttgcaggttggcgacggcgacaccgaccactactgctg
gcagcggccggaggacatgaccacgtcccgccaggcctacaaggtcgaccgcgagcgccccggctccgacgt
cgccggcgagaccgccgccgccctggccgccgcctccatggctttccgcgagaccaacccgcactacgcccac
ctcctcctccaccacgcccagcaggtttgtaattgtcaccttcactactactactactattacagtgcacgtagctaatttt
agtacactattccagagattagtactactacgagtggtaaattctttggcttgtgtggaatgtgtcgggtacgtgtagcg
gagctgctgggctgtagattctgaggaccagccaggaggcacgagtcaaagagggcagaggatttttctccctggt
tcacattggaaccgtgttcttctcgtcgctgctcgcgacggcacgtgccaacttgtccacatgacgtcatcgtcatgcc
gctcgcaaaatcgtcttctccacaaattgacaaattatataaaataaccgcgtgaaggttctgtcggtccggggaggc
cgacaggtgggtccacgaaagagctagcggcatcatgcagcacactgtcagcgtactactacgtggctgcgtgat
gcgacaggctacttaacgtaacaaccacatttccgttcacgttagggcctgcagaaacccacgcgattacaacgatg
cgatgcgatgcgatgcgatgatgggcgattattattgccgccgttggattatcgggggaatattagttggttggtttgct
tttttaatggcgtgtatgtacgtacttaagtgagcattgtgctgacatatgtctgtttttctgttggtggtacagctgttcga
gttcgctgacaagtacagaggcaagtacgacagcagcatcgcggaggtgaagagctactacgcgtccgtgagcg
gctaccacgacgagctgctctgggcggcgctctggctgcaccgcgccacggggcgggcgagctacctggactac
gtggtggacaacgcgcacgagttcggcggcacgggatgggccatcaccgagttcagctgggacgtcaagtacgc
cggcgtccagatcctcgccacccggctgctgctgcgtggcgagcacgaggagcgccaccgggccacgctccag
gggtaccgcgccaaggcggagcactacctgtgcgcgtgcatgggcaagaacgccgccgccgaggacaacgtc
gagcgcagccccggcgggatgctctacgtgcgccagtggaacaacatgcagtacgtcaccagcgccgccttcctc
ctctccgtctactcgggctacctcacggaggccggcgagggcgcggcggcgacgtgcggcgaggtggcggtcg
acgccggcgaggtgttcgcgcacgccagggcgcaggtggactacgtgctggggagcaacccgcgcgggatca
gctacctggtggggtacggcgccaagttcccggcgagggtgcaccaccgcgcggcgtccatcgtgccgtacaag
gacaggaaggagttcatcgggtgcgcgcaggggttcgacgactggttcggccgcaagagcgccaaccccaacgt
cgtcgtcggcgccatcgtcggcgggcccgaccgccacgacaggttcagggacgacagggtcaactacatgcag
acggaggcgtgcacctacaacaccgcccccatggtcggcatgttcgccatgctcaacagggtggcgcgccagca
ggaccccacgccggcgacgacacggccggagggcaccacaagtactgcggcggagtgaggtgtaaattatatat
atatactctaggcagtagtaatacgtgcagggggatctgtacaattttgcccctcctattctttgttcattcttttctggta
gacgcgaaggctcttctcttcttcgcagggattatttgtatacgggaaaaatagggtggacgcatttgacaagcaaat
gcgaattccctccccgtaaatataagataagaggaaaaatatatatgtatatgaaattagtctactggccggatttccaa
tatatgagaatgtttttccagcgaatgattgctaatattctggggaaaaaaacattgccgcctttctaatcatgcatcaca
gcgagctcgtgcaatgattaagcatctggtgaggaagattaaacaaattcaacgcgtcaagaaaagcatggcaatct
aa

>HvCel10_ORF

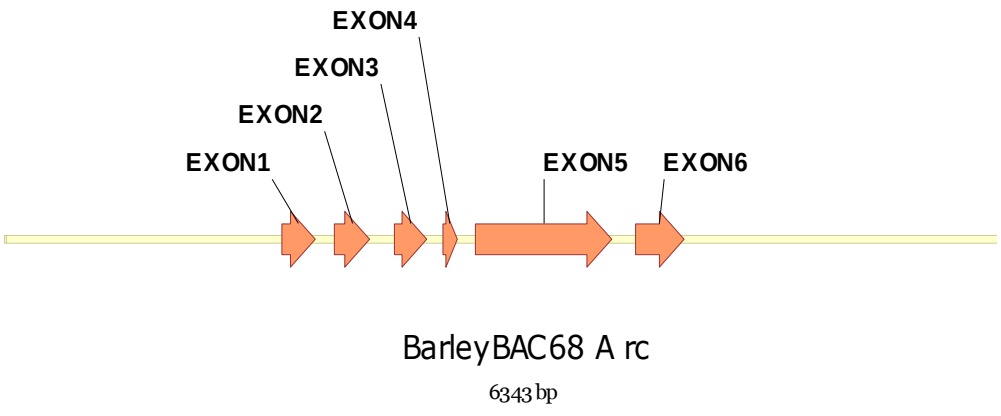
atgataatgaggaggagcagcggcgcggcggcttgggttcttggcgtggccgtggcggcggccgcgacgttcttc
ctggcggcggaggcggccggcgcggcgcccgtgccgcacgactacgagcaggcgctgcggaagagcctgctc

tacttcgaggcgcagcggtcggggcggctgccgtacggccagcgcgtcgcctggcgcgaccactccggcctcac
cgacggcctcgagcaaggggtggatttggtgggcggatactacgatgccggcgaccacgtcaagttcgggctgc
ccatggccttcacggtgaccatgctgtcgtggagcttgcttgagtacggcggcgacgtggcggccgccggcgagc
tcgcgcacgccctcgagtccatcaagtggggcaccgactacttcatcaaggcgcacaccaagcccgacgagctct
gggccgaggttggcgacggcgacaccgaccactactgctggcagcggccggaggacatgaccacgtcccgcca
ggcctacaaggtcgaccgcgagcgccccggctccgacgtcgccggcgagaccgccgccgccctggccgccgc
ctccatggctttccgcgagaccaacccgcactacgcccacctcctcctccaccacgcccagcagctgttcgagttcg
ctgacaagtacagaggcaagtacgacagcagcatcgcggaggtgaagagctactacgcgtccgtgagcggctac
cacgacgagctgctctgggcggcgctctggctgcaccgcgccacggggcgggcgagctacctggactacgtggt
ggacaacgcgcacgagttcggcggcacgggatgggccatcaccgagttcagctgggacgtcaagtacgccggc
gtccagatcctcgccacccggctgctgctgcgtggcgagcacgaggagcgccaccgggccacgctccaggggt
accgcgccaaggcggagcactacctgtgcgcgtgcatgggcaagaacgccgccgccgaggacaacgtcgagc
gcagccccggcgggatgctctacgtgcgccagtggaacaacatgcagtacgtcaccagcgccgccttcctcctctc
cgtctactcgggctacctcacggaggccggcgagggcgcggcggcgacgtgcggcgaggtggcggtcgacgc
cggcgaggtgttcgcgcacgccagggcgcaggtggactacgtgctggggagcaacccgcgcgggatcagctac
ctggtggggtacggcgccaagttcccggcgagggtgcaccaccgcgcggcgtccatcgtgccgtacaaggaca
ggaaggagttcatcgggtgcgcgcaggggttcgacgactggttcggccgcaagagcgccaaccccaacgtcgtc
gtcggcgccatcgtcggcgggcccgaccgccacgacaggttcagggacgacagggtcaactacatgcagacgg
aggcgtgcacctacaacaccgcccccatggtcggcatgttcgccatgctcaacagggtggcgcgccagcaggac
cccacgccggcgacgacacggccggagggcaccacaagtactgcggcggagtga

>Cel10 amino acid_BAC11

mimrrssgaaawvlgvavaaaatfflaaeaagaapvphdyeqalrksllyfeaqrsgrlpygqrvawrdhsglt
dgleqgvdlvggyydagdhvkfglpmaftvtmlswslleyggdvaaagelahalesikwgtdyfikahtkpd
elwaevgdgdtdhycwqrpedmttsrqaykvdrerpgsdvagetaaalaaasmafretnphyahlllhhaqql
fefadkyrgkydssiaevksyyasvsgyhdellwaalwlhratgrasyldyvvdnahefggtgwaitefswdv
kyagvqilatrlllrgeheerhratlqgyrakaehylcacmgknaaaednverspggmlyvrqwnnmqyvtsa
afllsvysgylteagegaaatcgevavdagevfaharaqvdyvlgsnprgisylvgygakfparvhhraasivpy
kdrkefigcaqgfddwfgrksanpnvvvgaivggpdrhdrfrddrvnymqteactyntapmvgmfamlnr
varqqdptpattrpegttstaae*

HVCEL11_BAC68



HVCEL11_BAC68_sequence

tttaaaatgtagaaaaaaaatcagtaacaaagcccacatacataacccagtagggctcagcccgaactgaaccagg
gtcaatccatttatgtaacactcaacaataaaactttagatgacctacaactttgtacacttttccaaaatcttgtaccgga
tatctgtttgccctgatattcttacttaatcagttttaatacctcatcgatcaaattaattgacgtaacttctatgcagtgtaaa
atagatattatatagaggctgtgtcaattaattcggattgagggaggactcgttttatgtcatgaacctcggatacactgt
tttaaaaccgaagaaattaatatgattttgagatgtgcacgatctgtctcatttgctttccaatatttagtctttagtgttttaa
ttttaatttgcgcaatcacgatttagatagccaacagattttgtactcgtgtctgaataaccagacctccgccacctttga
gggccgtctccgcgtcctcttccttccgacgggatacacgctataaatacatgctcccccggctcgagagcatccac
ccaaagcactacaccctgtggcattccctctttcgctagcctctagccatgggtgtcgtcgttgtcgtggtgcttgccg
ccatggtgcaggtggcgtactgcgggggcggtggcaccaccccgaactaccgcgaggcgctggccaagagcat
cctcttcctggaagcccagcggtccggcgtgctcccggccagccaaagggtcagctggaggtccaactccggcct
gctcgacgggaaggccaacggggtatgcaactgcaactgcccatttggaaactcatctcaaatcagggtgcagcc
aattaatctgtgcatgctatgaccagtatacattaacttgtggccacatgcattttaatttctttcaggtggatctcacggg
cggctactacgacgccggcgacaacgtcaagttcgggttgccgatggcgttcacggtgaccatgatggcgtggag
cgtcctcgagtacggcggggcagatggaggcggccggcgagctcggcaacgccctcgaggccgtcaagtggg
gcaccgactacttcgtcaaggcgcacccggaggccaaggtgctctatggagaggtagctcttcgcactactgcttc
gcttacttttgagttttgaaacgcactcgagtgcatgcacgtcgtcgtcgggtttacgtggtcatacgggtagtcggca
gctgaccaagtgtacgcatcttggaatggattcttgtttgtgcaggtgggagacggtgacaccgaccacggctgctg
gcagcggccggaggacatgtcgacgagccgtcaggcgtaccgcctcgacccccagcgcccggggtccgacctc
gccggggagaccgccgccgcaatggccgcggcgtcgctcgtgttccgccgctccaacccgggttacgccaacca
actcctgcaacactccaagcaggtttgccgattctatctataaactaacaatatgtattgatgaaatgatccaatagtca
atcctgacgtttcttggctcatttggtggatggctcagctgtttgatttcgccgacaagtaccggggaaggtacgacgc
cagcatccccgtggcgaggaatttctaccaatcagtcagcggatacagcgtaagctatcagatcagatcagatgga
cactccgcccgccattttcaagttcaacatccatcttgatctggttaactaacccgtacgtgtggtgtggtatatatctttc
aggacgagctgctgtgggcggcggcgtggctgttccaggcgaccgacgacaggcgctacctggactacgttgcc
aacaacgccgacgcgatgggcggcacgggctggtccaccggcatgttcagctgggacatcaagtaccccggcgt
gcaggtcctcgccgccaagatcctgctgcacggcaaagccggcgcgcacgccgccgtgctgcagcggtaccgg
cagaaggcggacttcttcgtctgcgcgtccctcggcaagcagggctccggcaacgtaaagcggacgcccggcgg
cctcctgcaccaccagaagtggaacaacctccagttcgtgaccagcggctcgtttctcctggccgcctactccgaca
gcctcgccgccgccgggcaggccggcgtgcggtgcccggccggggcggcgtcgccggccgagatcatggcgt
tcgccaagtcgcaggtggactacatcctggggagcaacccgaggggcacgagctacatggtcggctacggcttca
tttacccgctggaggcgcaccaccgcggcgcgtccatcgtgtccgtcaagtccaacccggcgttcgtgagctgccg
cggcgggtacgcgacctggtacccacgcaagggcagcaacccgaacctgctcgacggcgccatcgtcggcgggg
ccggacgagtacgacaacttcgccgacgagaggaacaactacgagcagaccgaggccaccacatacaacaacg
cgccgttcatgggcgtgctcgctcggctcgccgccggccaaggtggtggcggcaggtttgatcggtcagggctcg
acaaccagacctcactagtgccatctctttctgctggtaattaataatctacccaacatcatgcatggcccgagtcttgt
cattttaccatccaacccgagaggagatgactagctcttgattcttcacctcgtgcgcgttgcagccgccgatcaagct
gagcacaagtcgccgatcgtgatcgagcagaacgcgacggcgtcatggacggagaagggccggacgtaccgc
cggtacgcggtgacggtgaccaacaggtcgctgaacaagacggtgcacgagctgtacctcggcatcgccagtctc
tacggcccggtgacggggctcgacaagacgcggtatggccacgtgctccctggcacggccccgtcggtgcccgc
cggcgggagcgtcacgttcgagtacgcccacgccgcgccgccggcgaacgtctgggtcaccggctacaagctc
gtctgaactgaaaccaatggtagtcaagtttgtcaggttaacaagagggaagctggagcgatggggctagtgacaa
tctctcagccatcctccatacgctgttataataagctaagctaagttcgtaggagtaactagtagatgtgtgttatcgtga
ttgaaaatgtaagaggaatggtattgttgcgctataaagtgtataaatgctcgtgctttgcacggataaaatattatagta
ctccatgtgcctttatgcaaaacttttttaaatttaattcctaccaacttttttttggaccatggaatttggacgcacctaacaat
gttttttttgaaccggacctaacaatgttcgacgtgacttatttcagctctccgatgaaagaatagttgtgtacggaatag
tctttcgatcaattttatttttttgttgttcagaaaaacatgaatttatttttcattttggga

>HvCel11_ORF

atgggtgtcgtcgttgtcgtggtgcttgccgccatggtgcaggtggcgtactgcgggggcggtggcaccaccccga
actaccgcgaggcgctggccaagagcatcctcttcctggaagcccagcggtccggcgtgctcccggccagccaa
agggtcagctggaggtccaactccggcctgctcgacgggaaggccaacggggtggatctcacgggcggctacta
cgacgccggcgacaacgtcaagttcgggttgccgatggcgttcacggtgaccatgatggcgtggagcgtcctcga
gtacggcgggcagatggaggcggccggcgagctcggcaacgccctcgaggccgtcaagtggggcaccgacta
cttcgtcaaggcgcacccggaggccaaggtgctctatggagaggtgggagacggtgacaccgaccacggctgct
ggcagcggccggaggacatgtcgacgagccgtcaggcgtaccgcctcgacccccagcgcccggggtccgacct
cgccggggagaccgccgccgcaatggccgcggcgtcgctcgtgttccgccgctccaacccgggttacgccaacc
aactcctgcaacactccaagcagctgtttgatttcgccgacaagtaccggggaaggtacgacgccagcatccccgt
ggcgaggaatttctaccaatcagtcagcggatacagcgacgagctgctgtgggcggcggcgtggctgttccaggc
gaccgacgacaggcgctacctggactacgttgccaacaacgccgacgcgatgggcggcacgggctggtccacc
ggcatgttcagctgggacatcaagtaccccggcgtgcaggtcctcgccgccaagatcctgctgcacggcaaagcc
ggcgcgcacgccgccgtgctgcagcggtaccggcagaaggcggacttcttcgtctgcgcgtccctcggcaagca
gggctccggcaacgtaaagcggacgcccggcggcctcctgcaccaccagaagtggaacaacctccagttcgtga
ccagcggctcgtttctcctggccgcctactccgacagcctcgccgccgccgggcaggccggcgtgcggtgcccg
gccggggcggcgtcgccggccgagatcatggcgttcgccaagtcgcaggtggactacatcctggggagcaacc
cgaggggcacgagctacatggtcggctacggcttcatttacccgctggaggcgcaccaccgcggcgcgtccatcg
tgtccgtcaagtccaacccggcgttcgtgagctgccgcggcgggtacgcgacctggtacccacgcaagggcagc
aacccgaacctgctcgacggcgccatcgtcggcgggccggacgagtacgacaacttcgccgacgagaggaaca
actacgagcagaccgaggccaccacatacaacaacgcgccgttcatgggcgtgctcgctcggctcgccgccggc
caaggtggtggcggcaggtttgatcggtcagggctcgacaaccagacctcactacacaagtcgccgatcgtgatc
gagcagaacgcgacggcgtcatggacggagaagggccggacgtaccgccggtacgcggtgacggtgaccaac
aggtcgctgaacaagacggtgcacgagctgtacctcggcatcgccagtctctacggcccggtgacggggctcgac
aagacgcggtatggccacgtgctccctggcacggccccgtcggtgcccgccggcgggagcgtcacgttcgagta
cgcccacgccgcgccgccggcgaacgtctgggtcaccggctacaagctcgtctga

>HvCel11 amino acid__BAC68

mgvvvvvvlaamvqvaycggggttpnyrealaksilfleaqrsgvlpasqrvswrsnsglldgkangvdltgg
yydagdnvkfglpmaftvtmmawsvleyggqmeaagelgnaleavkwgtdyfvkahpeakvlygevgd
gdtdhgcwqrpedmstsrqayrldpqrpgsdlagetaaamaaaslvfrrsnpgyanqllqhskqlfdfadkyrg
rydasipvarnfyqsvsgysdellwaaawlfqatddrryldyvannadamggtgwstgmfswdikypgvqv
laakillhgkagahaavlqryrqkadffvcaslgkqgsgnvkrtpggllhhqkwnnlqfvtsgsfllaaysdslaa
agqagvrcpagaaspaeimafaksqvdyilgsnprgtsymvgygfiypleahhrgasivsvksnpafvscrgg
yatwyprkgsnpnlldgaivggpdeydnfadernnyeqteattynnapfmgvlarlaagqggggrfdrsgldn
qtslhkspivieqnataswtekgrtyrryavtvtnrslnktvhelylgiaslygpvtgldktryghvlpgtapsvpa
ggsvtfeyahaappanvwvtgyklv*

HVCEL12_PARTIAL_BAC74

HVCEL12_PARTIAL_BAC74_sequence

tgcttaccggtaaccacatacacctacctatacccttgctgaggcctccccctcgcatggtggcggatcgggtccac
cccgcccgtgtgtggaagtaaatcgcttttaatgttctggtttaggggaagctgaaaaggccctcccctcggaaacag
gggcggtaaagaaaaattcttctgaccccaaaggcttgtatgcgggttttgctttgggaaggcgggggggagacgac

tggccttgggttgggcgggggcgcttgcgcaccatgagccggggccgcatccggcgggcagctggtagccccgt
ggtcgccgcggtggtgctcctgcaggcggccgtcctggcactgtccctcggctgcgatggcgccgcggtgacca
ccgacgcggaggcggagccggactaccgggaggcgctgtccaagagcctgctctacttcgaggcgcagcgctc
cgggcggctgccgtacaaccagcgcgtgaggtggcgggggcattcggggatcaccgacgggctgcagcaggg
cgtggacctggtgggcgggtactacgacgccggcgaccacgtcaagttcggcctgcccatggccttcaccgtcac
catgctctcctggggcgccatcgagttcggcgacgacatcgccgccgccggcgagtggcaacacacgctggagg
ccatcaagtggggcaccgactacttcatcaaggcgcacaccgagcccttcgtctactgggccgaggtaccaacgc
acgcacgaacgtcgcccattttccttttctttctgtttttgtttttaccgccacgacggccggcgtcccgatcgcccgcgcg
accgcttcgaggcgacacgggaaaagcggatgcccgaggcgactgccattaatttactgctgcgatgcgaatgat
ggcgggcggatgtctgacgagacgggccccgcaggtgggggacggcgacacggaccactactgctggcagc
ggcccggaggacaatgacgacgtcgcggaaggcgtaccgggtggacaaggagaaccccgggctccgacctcg
tcgccgagacccgcccccggcgctcgccgccgcctccatcgtcttccgcgcctccaacccgcattactccgacatc
ctcctccaccacgcgaaacaggtaagtacgtcctctacgtgcttataataactcatcaccagcacgggacaaggtag
ctagctgtatgctatgttctgactgcgaccggtgcagctgttcgagttcggggacaggtacaggggcaactacgaca
gcagcatcgcggaggtgcggagcttctacgcgtcggtgagcgggtaccgggacgagctgctgtgggcggccttc
tggctgcaccgggcaacggggaaggaggagttcctccgctacgccgtcgacaaggccgactgcttcggcggcgt
cggatgggccatgaccgagttcagctgggacgtcaagtatgccggcctccaggtcctcgcctccaaggtacatact
tgttccactattgtcacgaactccctccttccgcagattaacataaaaacgtttaaatcactagtttaataagtagtaattta
taaacgttttttatattattagattactagtttagtaagctacgtagaagaatttgctttcagaaataagtacgtagtacgaa
atttgcacatcaacatcttcaattcctccggcactggatgacccctc

>HvCel12_partialORF (Lower case sequence indicates that obtained by BAC sequence. Uppercase sequence was obtained from the TIGR database)

Atgagccggggccgcatccggcgggcagctggtagccccgtggtcgccgcggtggtgctcctgcaggcggcc
gtcctggcactgtccctcggctgcgatggcgccgcggtgaccaccgacgcggaggcggagccggactaccggg
aggcgctgtccaagagcctgctctacttcgaggcgcagcgctccgggcggctgccgtacaaccagcgcgtgagg
tggcgggggcattcggggatcaccgacgggctgcagcagggcgtggacctggtgggcgggtactacgacgccg
gcgaccacgtcaagttcggcctgcccatggccttcaccgtcaccatgctctcctggggcgccatcgagttcggcga
cgacatcgccgccgccggcgagtggcaacacacgctggaggccatcaagtggggcaccgactacttcatcaagg
cgcacaccgagcccttcgtctactgggccgaggtgggggacggcgacacggaccactactgctggcagcggc
ccggaggacaatgacgacgtcgcggaaggcgtaccgggtggacaaggagaaccccgggctccgacctcgtcg
ccgagacccgcccccggcgctcgccgccgcctccatcgtcttccgcgcctccaacccgcattactccgacatcctc
ctccaccacgcgaaacagctgttcgagttcggggacaggtacaggggcaactacgacagcagcatcgcggaggt
gcggagcttctacgcgtcggtgagcgggtaccgggacgagctgctgtgggcggccttctggctgcaccgggcaa
cggggaaggaggagttcctccgctacgccgtcgacaaggccgactgcttcggcggcgtcggatgggccatgacc
gagttcagctgggacgtcaagtatgccggcctccaggtcctcgcctccaagNNNNNNNNNCACTAC
CTGTGCGCGTGNCCTACGGCTGAACAACGGCAGCAACGTGGACCGTA
GCCCCGGCGGCATGCTGTACGTCCGGCAGTGGAACAACCTCCAGTAC
GCCTCCAGCGCCGCCTTCCTGCTCACCGTCTACTCGCGGTACCTTGCC
GGCGCCGGCGCCGGCGCCCGCCTCCGCTGCCCCGACGCGCCGGCGGT
CCCGCCGTCGGAGCTCCTGGCGCTCGCGCGCTCGCAGGCCGAGTACA
TCCTGGGCCGCAACCCGCTGCGGCTGAGCTACATGGTGGGCTACGGC
CCCCGCTACCCCGCGCACGTGCACCACCGCGCCGCCTCCATCGTGTCG
CACAAGGCCAACAACCGTTTCATCGGCTGCATGCAGGGGTTCGACGA
CTGGTTCGCCCGCAAGCGGCCCAACCCCAACGTGCTCACCGGCGCCA
TCGTCGGCGGGCCCAACTGCCGCGACGAGTTCAGGGACGACCGCGGG
AATTACGTGCAGACCGAGGCGTGCACCTACAACACGGCGCCCATGGT

CGCCGTCTTCGCCAGGCTGCACAACCTGTCCACGGCAGCGGCGGAGG
AGCGGTGCCGCCCAGGGACGGCGCCGGGGCTCTCGGCCAAGTGTAGA
TAAAGGTGGTATTAGCTTGGGTGGGAAATTCTTTTTTCTTTCTCGTTTA
CGGCCTTACTTCCATATAAGATTTCTTTTTGAAGTGTTAATATGAGAG
ACATTTATTTAGTGAAAGAAAGGAACAAAGCATAGCAGTGAATGAAC
AATAAGGTGGAGACCGATCNAGGAAGGTGAAGATTAGATACTTGTAT
TCTCAAGTTATAACCAGGTGCCTCAGAAAGCAAATTTAAAGAG

>HvCEL12 amino acid

msrgrirraagspvvaavvllqaavlalslgcdgaavttdaeaepdyrealsksllyfeaqrsgrlpynqrvrwrgh
sgitdglqqgvdlvggyydagdhvkfglpmaftvtmlswgaiefgddiaaagewqhtleaikwgtdyfikaht
epfvywaevggrrhgplllaaarrtmettsrkayrvdkenpglrprrrdpppalaaasivfrasnphysdillhha
kqlfefgdryrgnydssiaevrsfyasvsgyrdellwaafwlhratgkeeflryavdkadcfggvgwamtefsw
dvkyaglqvlaskxTTCARXLRLNNGSNVDRSPGGMLYVRQWNNLQYASSAA
FLLTVYSRYLAGAGAGARLRCPDAPAVPPSELLALARSQAEYILGRNPLR
LSYMVGYGPRYPAHVHHRAASIVSHKANNRFIGCMQGFDDWFARKRPN
PNVLTGAIVGGPNCRDEFRDDRGNYVQTEACTYNTAPMVAVFARLHNLS
TAAAEERCRPGTAPGLSAKCR

HVCEL14_BAC73

HVCEL14_BAC73_sequence

gggcaagagcggggagcggcaaggcgcacaactacgagggaggcgctgcggaagagcgtggtctacttcgag
gcgcagcggtcggggcggctgccgcacgggcagcgcgtgccgtggcgcgaccactcgggcctcaccgacggc
ctggagcaaggggtggacctggtgggcggctactacgacgccggcgaccacgtcaagttcggcctgcccatggc
cttcaccgtcaccatgctgtcctggagcatgatcgagtacggcgccgacgtcgcggcggccggcgagctcgcgca
cgcgctcgaggccatcaagtggggcaccgactacttcatcaaggcgcacacgcagcccgacgagctgtgggccg
aggtacgtacctacatgccgacgaacgatctcgtgtcacctcgcctcggctagctgtgttcttccgttcagacacgat
cgatcatgaccatgacatgcatatgcaaatgagaccgaccagttccaagccaagctagctgttgcgcacgcaagaa
agagagagggagagaccaattcaatcatgcacacctagctaccttgttgcactttgatcgattacttttttttcatgaatt
atgttatagataccctatatctttttttctgtagaaagatcgattatatattgtatatataggagtattttctggttggggaaga
aaagaaaaaaggggttctttctggtgaataacaacaagcgtgttaggtaggcggccagtttgatttgatgaacgtacgt
acgaacgaacgaacgaacggcagcactaaactatgcgtgtttttttctttcttccacacatggcgcgcataggtttcaa
aaaagctttcgtcggtccatgcatccgaaccgtctgatcagtttccttttttattttttgagctggaaataatctccttttactt
gcaacacaacaacaacaggtgggagacggcgacacggatcactactgctggcagcggccggaggacatgacg
acgtcgcggcaggcgtacaaggtcgaccgcgaccacccgggctccgacgtcgccggcgagacggccgccgcc
atggccgccgcctccatggtcttccgcaaatccaacccgcactacgcccacctcctcctccaccacgcccagcagg
tacatacatacatacatacatacatactcgacacacacacacacacacgtatacattgacgtacgcggtgcatg
caaggtcctatgtggaaaaatggaacggaaaaaaatatcggtgcaagtttggagcttcgtgtagcaccagtctgtcc
gcgtactggattttggaacatagcgcgcgtattggtagctttggtcaagggggtgccggtcttttcatgggcggtgtt
cttttcttgtgccgcacatgccaagttgccaacacctcaacacgatcgacgtcgatccatcgtaggggctggctgccg
tgtggggccgccggaacgcaacctctatctggtggacgggggtcatgggaaactttccatttgctgcctgcctccgg
ccatgcatgtatatatgattaggtggccagttcaatcaatcacatgcttaactgggaacattattaggaagcggtcgtc
gtgtcatcatcatgtgctaattgaaccgctccgtattcacacccactgacctggccacccacctgctgcatgctcttca

agcgtcaactataaaatgcgtcaatgaatgatacgatacgtatgcaaccttaattatcattagggctttcttaaataattta
tgaacaaaccaaactaacagcgttacaatcttagttgttgttggctgatcatgtttactttggctttttgtgaacgtgcgtg
cagttgttcgagttcgcggacgggtacaggggccaatacgacagcagcatagccggaggtgaagagctactacg
cgtcggtgagcgggtacaaggacgagctgctgtgggccgccctgtggctccaccgcgccaccggcaaggccag
ctacctcgactacgtcgtcgacaacgcccacgacttcggcggcaccggctgggccatcaccgagttcagctggga
cgtcaagtacgccggcgtccagatcctcgccgcaagggtacgtaagctatcatgcatgtgaaggatcgatcactag
acggccggccggtcactgacgaggagaccaaatgattaattaattgcagctgctgctgaggggagagcacacggc
ggagcagaagaggacgctggagcagtaccaggccaaggcggagcactacgtgtgcgcgtgcctgggccggaa
cgcggccggcggcgaggacgccaacgtggagcggagcccgggagggatgctctacatccggcagtggaacaa
catgcagtacgtgaccaacgcggcgttcctgctctccacctacgccgactacctggcgggggccggcgtggggac
ggtggcctgcgccggcggcgagacggcgggcgccggcgaggtggccgcgctggccagggcgcaggtggact
acgtgctgggcagcaacccgaggggggtgagctacctcgtggggtacggcgccaagtaccccgccagggtgca
ccaccgcgcggcatccatcgtgccctacaagcacagcaagcagttcatcgggtgcacgcaggggttcgaccactg
gttcggccgccggagctccaaccccaacgtgctcgtcggcgccatcgtcggcgggcccgaccggagggacagg
ttcagggacaaccgggacaactacatgcagacggaggcgtgcacgtacaacaccgcgcccatggtcgggatgtt
cgccaagctcaaccggatggccagggaggagaaggtggcggccgaggtgtaaatggattagataggggttatag
gtccaggggctgtacaattttccccattcttttgttcatttttctctccttctttggaggtttcttgatacacaggaaggatg
gagatgggtgtagaaatacgaggaaaaaacaaagggtggaggcattcacatgtgacatgcaaatgtaaattcgctg
cctgtac

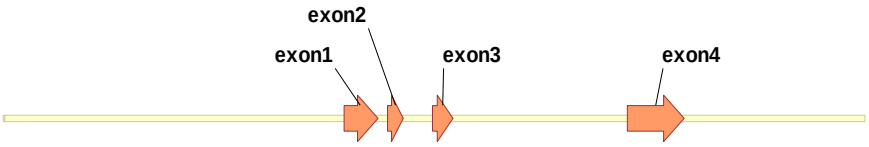
>HvCel14_ORF

atgagatcggggaggggcaccgcagccgcggcatgtctccgggccgcgctgctgctaccggcgctgctgacgat
gctgatcgcggcggcgtcgtcttcggcggcggaggggggcaagagcgggagcggcaaggcgcacaactacga
ggaggcgctgcggaagagcctggtctacttcgaggcgcagcggtcggggcggctgccgcacgggcagcgcgt
gccgtggcgcgaccactcgggcctcaccgacggcctggagcaaggggtggacctggtgggcggctactacgac
gccggcgaccacgtcaagttcggcctgcccatggccttcaccgtcaccatgctgtcctggagcatgatcgagtacg
gcgccgacgtcgcggcggccggcgagctcgcgcacgcgctcgaggccatcaagtggggcaccgactacttcat
caaggcgcacacgcagcccgacgagctgtgggccgaggtgggagacggcgacacggatcactactgctggca
gcggccggaggacatgacgacgtcgcggcaggcgtacaaggtcgaccgcgaccacccgggctccgacgtcgc
cggcgagacggccgccgccatggccgccgcctccatggtcttccgcaaatccaacccgcactacgcccacctcct
cctccaccacgcccagcagttgttcgagttcgcggacgggtacaggggccaatacgacagcagcatagccgagg
tgaagagctactacgcgtcggtgagcgggtacaaggacgagctgctgtgggccgccctgtggctccaccgcgcc
accggcaaggccagctacctcgactacgtcgtcgacaacgcccacgacttcggcggcaccggctgggccatcac
cgagttcagctgggacgtcaagtacgccggcgtccagatcctcgccgcaaggctgctgctgaggggagagcaca
cggcggagcagaagaggacgctggagcagtaccaggccaaggcggagcactacgtgtgcgcgtgcctgggcc
ggaacgcggccggcggcgaggacgccaacgtggagcggagcccgggagggatgctctacatccggcagtgg
aacaacatgcagtacgtgaccaacgcggcgttcctgctctccacctacgccgactacctggcgggggccggcgtg
gggacggtggcctgcgccggcggcgagacggcgggcgccggcgaggtggccgcgctggccagggcgcagg
tggactacgtgctgggcagcaacccgaggggggtgagctacctcgtggggtacggcgccaagtaccccgccag
ggtgcaccaccgcgcggcatccatcgtgccctacaagcacagcaagcagttcatcgggtgcacgcaggggttcg
accactggttcggccgccggagctccaaccccaacgtgctcgtcggcgccatcgtcggcgggcccgaccggag
ggacaggttcagggacaaccgggacaactacatgcagacggaggcgtgcacgtacaacaccgcgcccatggtc
gggatgttcgccaagctcaaccggatggccagggaggagaaggtggcggccgaggtgtaa

>HvCel14 amino acid BAC73_Cel14

rksvvyfeaqrsgrlphgqrvpwrdhsgltdgleqgvdlvggyydagdhvkfglpmaftvtmlswsmieyg
advaaagelahaleaikwgtdyfikahtqpdelwaevgdgdtdhycwqrpedmttsrqaykvdrdhpgsdv
agetaaamaaasmvfrksnphyahlllhhaqqlfefadgyrgqydssiaevksyyasvsgykdellwaalwlh
ratgkasyldyvvdnahdfggtgwaitefswdvkyagvqilaarlllrgehtaeqkrtleqyqakaehyvcaclg
rnaaggedanverspggmlyirqwnnmqyvtnaafllstyadylagagvgtvacaggetagagevaalaraq
vdyvlgsnprgvsylvgygakyparvhhraasivpykhskqfigctqgfdhwfgrrssnpnvlvgaivggpdr
rdrfrdnrdnymqteactyntapmvgmfaklnrmareekvaaev

TC130995



BAC60 Cel130699 BACseq

10735 bp

TC130995_BAC60_sequence

tcaactgcagcagcagagcagtactaggactaggagtagctggtactgacttgtcccgggacgtcgctgtcctccc
cctgccaagccaccgcacctccaccgccaccgccatgcatgccattgccaagccccaaacccaaagcttcttcctc
ctttcctttcctttgtttcaacacagcccaccaacccaacatcgccatccattcctaccactttgtttgtatatctgcttccc
cgccccgcccatatataatcgcaccagcatcggccatgtacagcgaaggagatattgtcgaggaagaggaagagg
aagaggaagaggtccgccgctaaagctggaagtcaattactccctcgtcggttatctatatcagtttccatttccagctt
aagctaccaccgagcgacgaggttcacttttccgaggaggcgccggtccacttccccttcttggaaaccctttattcc
ggcgaccgcgggcgagtccggccaccggtggccattttcgtctacggcgggccgggggtcgatcgatcgcgcc
gtcgacgcttggttattcgcagtttctttcttggccggcgcgcgcgcgggccgggcgggcggtcggtgagatgagc
gatgtcacgggcaggttcgcgctggcggcggccgtggtggcggtgtcgctggtgatggcgtcggcgtcggcggc
gcggcacgactacggggacgcgctgtccaagagcctgctctacttcgagtcgcagcgctcggggcggctgccgt
acaaccagcgggtgcggtggcgcggccactcggggctcaccgacgggctggagcaggggggtggacctggtgg
gagggtactacgacgccggcgaccacgtcaagttcggcctccccatggccttcaccgtcaccatgctctcctgggc
cgtcctcgagtacggcgaccagatcgccgccgccggcgagggcgcccacgcgctccagtccatcaagtggggc
accgactacttcatcaaggcccacaccgcccccaacgagctctggacccaggtcagtcagcactcatcgccatgac
cggccgcgcgcgcgcgcagaatgtagtgtgtaaacacgtacgtacacacactgtatcttttctgacccgatcgatcg
acgtgtggcgcaggtcggcgacggcgactcggaccactactgctggcagcggccggaggacatgacgacgtcg
cggcgggcgtacaaggtggacgccgagcaccccggctcggaggtggccgccgagacggccgccgccatggc
cgccgcgtccatcgtgttccggcgagccggcgacgcgcactacgcgcacctcctcctccaccacgcgcagcagg
tacgctgtcacacgtattactactacatacactacttgccgcacgaaatcaatcaatcaacataaagctagcagaa
aattcagactagattagcagcaccagcagcttaatttccctttaagctagcttttgtcagctcgatcccttacacgaaac
tcacgattaaaatacgacagtcgatccactctagccgcagctgggtcgtctggtacgtgacagcgagcagctcacc
gtccgtcctacacacacaacaacacacacgcttcggcaatacgtattgtatatactcttctacatacatacggagtaca
tgcattaatttgaatttgtacgtacgtatgtatttatacgtgcagctgttcgagttcgcggacaagtacaggggccggta
cgacgacagcgtgcgggtggtgaagagcttctacccgtcggcgagcgggtacaaggacgagctgctgtgggcg
gcgttgtggctgcaccgcgccaccggcggccgcgactacctcgactacgtcgtctccaacgccgacgacttcggc
ggcaccggctggggcgtctccgagttcagctgggacatcaagtaccccggcctacagatcctcgcctccaaggtc

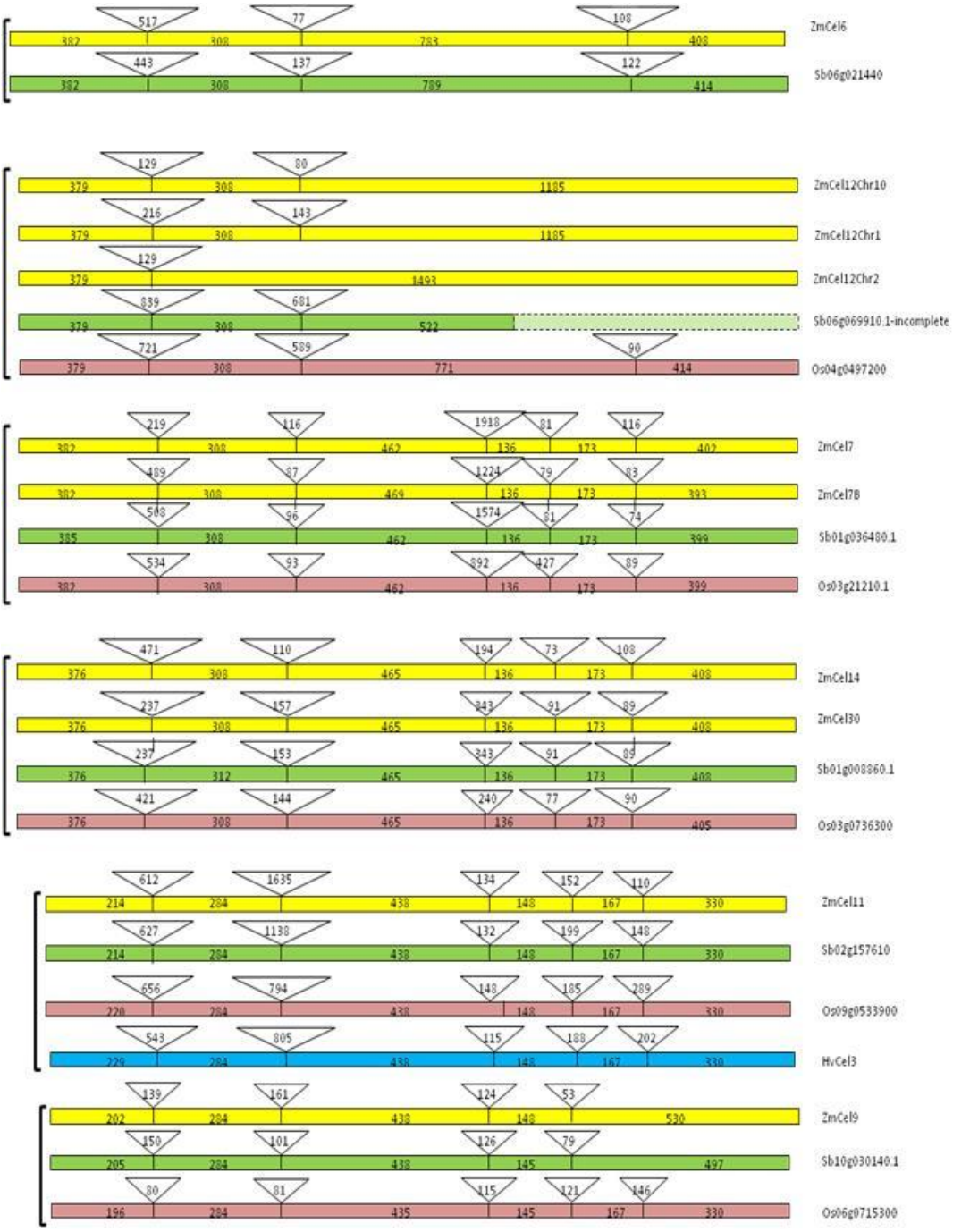
agtactaccgatcagtgatcagtctccattgctgacgcacgcacgcaagcaagcaacccatgcatgcatgcaccaa
cctatccacctttcctttctttggaccggtcaaacttgcttacctttttcatcgctgcatgcatgataagaatgatttcctctc
gccaaaaataggatcattccaccgccgatcgcggcagctagctgtaaattagtcaaagccatacgtcaattactactc
tactagttagtccagtccaaaatattaaaattagcagtccaaaatatagttagttggcaatttttaaagaaaattactactg
tggtaatcagtagcagccgtatagtactacgtaacaaaataacaggggttaatactgcatggtatgaatgcggttaatt
cgtcccatatacgtactctactgtagtaactgacagacgtgccaccggtttctgcatcgttagcatgaggagcaagcc
gcgtgtttggtgcaccctagcttttgcttagccacgcaccctctctctctagcttgcaagcatggctacctagcatgcttt
cttggggtgcttccgctggggttgatggatgcgccatgcccaacaactccgcccggaccggccatctctccacgctt
cttctagtagaatgcatatacatatcccaaaagcctttcgaccgaattgcaccgcgtgttctcgcgtagacgttgattttg
ttgttagagggcccggcgcccatgtgcagactccagttaagtatatataagcatatgagatggttgatgatgattctca
acaaaaaagctaatcatggttggggagcagagcagagcagacgatgctcagcccaccaagagggccgggtgaa
aagagatcagagatgcccaaatttgttttatatacagtacttaatacaaacatgcaactagctaggagagatcttttgat
ctaccctacccggcctacaccacctgcgcctttgtcctgtgagcaaccatatttccaacacgttcctttcacatgtgctc
ctcctggtgtctgccatgccatgccaatttattaaaaaaaccttccggaaacagaccggtttcctaacctaacctggatt
gtccgcgaggtgacggccgcactgtcactcaggtctcgttgggagttggactgccttcctaatcataactcctttgac
cacaccttctgtttctttctcccatgcatatgcttagaacgcaaggcagcatgcggctagctagctcctcttgccatgag
ggtgagggtgaaggctcatcgcatgcctctttctttggtcatccacaagctgtttttgtccctggcatgggttgctgctgt
ttgggatgtgatgtctgaagcaagcgattcaactaatgattgatggatactctggctggacatcttacttaattacttgcc
caagtggagtggaggagtcgaatggtcatttcttccgtcgtttaattaatgtatatcggctaatcaagttggtgagagga
gggaaaaaaatagtagaagcacatgagacaaaaaaaatattcaaccgacagtagctgcacggttttctttcttcttctt
ggattaagtagagtgcatgtctgtgtgggagaagcaaagtggttggatcacaggttgagcacctactagtaattgag
cccacaaggatcgtgtggcccatacgtacggcccatttttctttgatgttcgggttggtttataacctggaaaattcatta
tgtactgcgctgaaattaagttccatgataatatatgccattcgaccatggcatctgtcagttgatgtcgaaattctgcac
tcaaatattaccagacaaatggaggacaggttgattacgatcactgtgcttaagtcaaccctaataattaagtttggcac
atagtttgactactactgtagaaggtgctcacaaacgtcctgatcgacatgcagcctctaagcctatagaggtccacat
gtttcagaaagggtcaactgttttatttacaaaaggtcaaagggttcaaagccaggtcttatatttacaaatcagctgtca
attaaaaaccgaaactagccgtgcatcatcggctgtttagttagcaataattctgttttgatccgtgcgattaacaatcct
ctctcttggccattgatcgatcagctgctgattgaggagaaggagctgagctcggagtaccggacggtgctggaga
agtaccggtccaagggagagttctacgtgtgctccaacatgaacaagaaccccggcggcgccgcccgcaacgcg
ccgcgcaccccggcggggctcatcttcatccgtccatggaacaacctccagtacgtgaccaacgccgtcttcctcct
gaccgtccactcggaccacctggcggcgctcggggagcccctccgctgcaccgtggacgaggacgccgacacc
tcggacacgggcggcgagacggcggcggacgtcgtggtgggggcggaggaggtgctggcgttcgccaaggc
gcaggcggactacatcctgggcagcaacccgatggagacgagctacctggtggggtacggggagcagtacccg
cggcgggtgcaccaccgggcggcgtcgacggcgtcgctgacggacgagaaggggttcatcgggtgcacgcag
gggttcgactcgtggtacagcgcgggggagaggaacccgcacgacctcgtcggcgccgtcgtcggcgggcccg
acggcgaggaccggttcaacgaccagcggggcgcgtacatgcagaccgaggcatgcacgtacaacaccgcgc
ccatggtcggcgtcttctccaggctcatggagctccacgagaagcagcggctccgggagcaacgagacctgtgaa
gcggcccatgcatcgtcgtcggtgaatatgttcagtagtaggaatgttaattgaataactgagaaaaaagaagagtg
ggaaaaggctttttttttactaggatggatggatgggtgggtgggtggatttgtacattctttttgttgattaattcgtata
ggttttgctttgtttcattcatgtaggaatggaatttcggaagaaaaaaattgagggaataaaacatcaccagattgatg
atggtgatggttgaccttgtgttgaccaagagggaggaggaaaggcgggaaaggttttttttttacctcactgacacc
gggtcccgcagtccagtttccccccgcctcaccgcgaaatttgaatttccatgttgattccacaccgcccaccgagaa
gcagtgcccgatgtatgttcggcgaaatgccatcgccggagctagcgctgccgtggtcgcgccccccgtgcgcgt
cactctgggggaaccgtcggccttgacgccgcgcttctcatctcggcgcggacgcaggcgtgcgtgccgggggtc
gatttcggtcgccgcgcgcgcagggtgctcgacgaaatggctcccttggctgcctgcgctgcatctccttcggggc
agtgcagtgcagtgcgtgccgacatgcccactgacaccgaggtcaa

>130995_ORF
atgagcgatgtcacgggcaggttcgcgctggcggcggccgtggtggcggtgtcgctggtgatggcgtcggcgtc
ggcggcgcggcacgactacggggacgcgctgtccaagagcctgctctacttcgagtcgcagcgctcggggcgg
ctgccgtacaaccagcgggtgcggtggcgcggccactcggggctcaccgacgggctggagcaggggtggac
ctggtgggagggtactacgacgccggcgaccacgtcaagttcggcctccccatggccttcaccgtcaccatgctct
cctgggccgtcctcgagtacggcgaccagatcgccgccgccggcgagggcgcccacgcgctccagtccatcaa
gtggggcaccgactacttcatcaaggcccacaccgcccccaacgagctctggacccaggtcggcgacggcgact
cggaccactactgctggcagcggccggaggacatgacgacgtcgcggcgggcgtacaaggtggacgccgagc
accccggctcggaggtggccgccgagacggccgccgccatggccgccgcgtccatcgtgttccggcgagccgg
cgacgcgcactacgcgcacctcctcctccaccacgcgcagcagctgttcgagttcgcggacaagtacaggggcc
ggtacgacgacagcgtgcgggtggtgaagagcttctacccgtcggcgagcgggtacaaggacgagctgctgtgg
gcggcgttgtggctgcaccgcgccaccggcggccgcgactacctcgactacgtcgtctccaacgccgacgacttc
ggcggcaccggctggggcgtctccgagttcagctgggacatcaagtaccccggcctacagatcctcgcctccaag
ctgctgattgaggagaaggagctgagctcggagtaccggacggtgctggagaagtaccggtccaagggagagtt
ctacgtgtgctccaacatgaacaagaaccccggcggcgccgcccgcaacgcgccgcgcaccccggcggggctc
atcttcatccgtccatggaacaacctccagtacgtgaccaacgccgtcttcctcctgaccgtccactcggaccacctg
gcggcgctcggggagcccctccgctgcaccgtggacgaggacgccgacacctcggacacgggcggcgagac
ggcggcggacgtcgtggtgggggcggaggaggtgctggcgttcgccaaggcgcaggcggactacatcctggg
cagcaacccgatggagacgagctacctggtggggtacggggagcagtacccgcggcgggtgcaccaccgggc
ggcgtcgacggcgtcgctgacggacgagaaggggttcatcgggtgcacgcaggggttcgactcgtggtacagcg
cgggggagaggaacccgcacgacctcgtcggcgccgtcgtcggcgggcccgacggcgaggaccggttcaacg
accagcggggcgcgtacatgcagaccgaggcatgcacgtacaacaccgcgcccatggtcggcgtcttctccagg
ctcatggagctccacgagaagcagcggctccgggagcaacgagacctgtgag

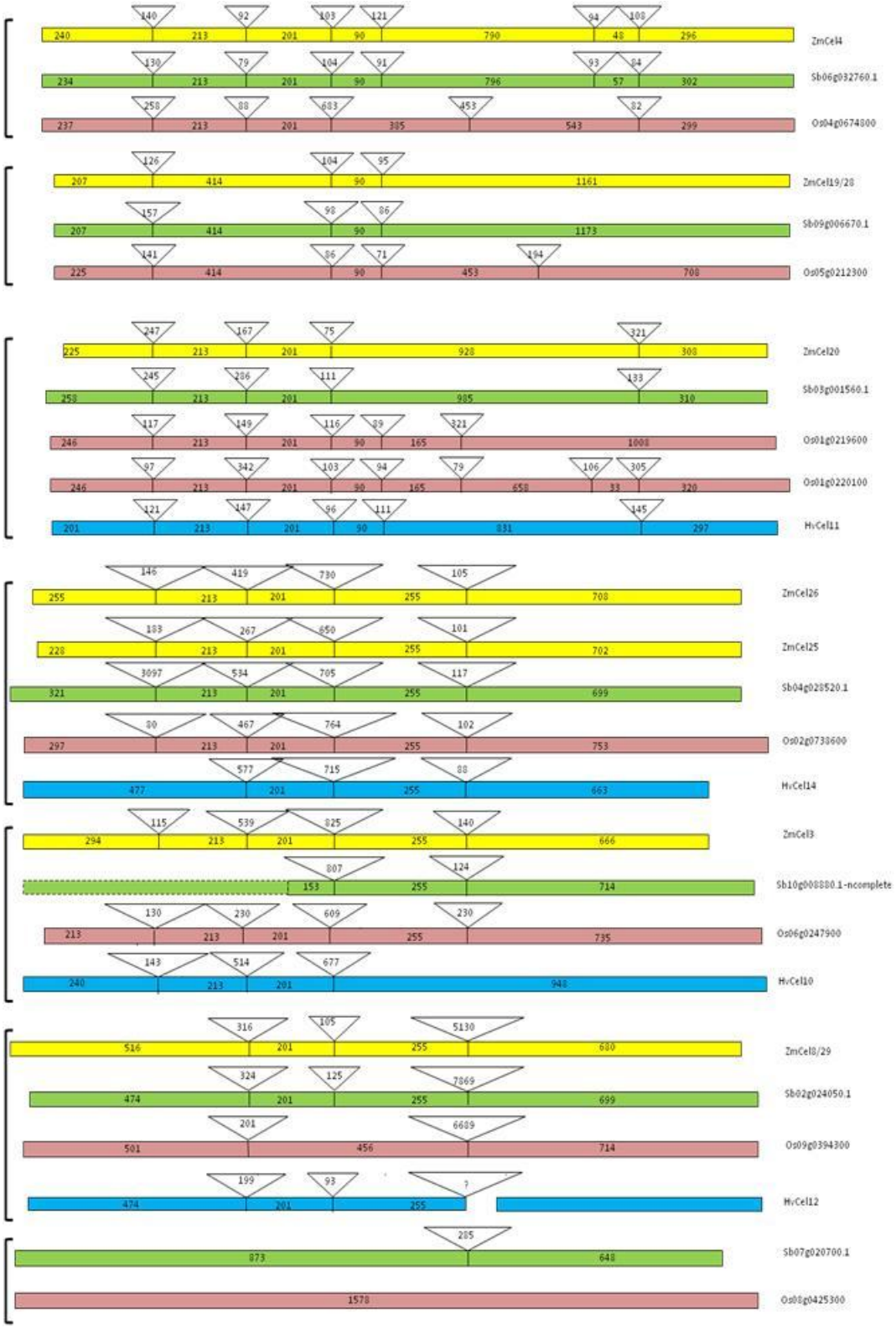
>Cel130995 amino acid_BAC60
msdvtgrfalaaavvavslvmasasaarhdygdalsksllyfesqrsgrlpynqrvrwrghsgltdgleqgvdlv
ggyydagdhvkfglpmaftvtmlswavleygdqiaaagegahalqsikwgtdyfikahtapnelwtqvgdg
dsdhycwqrpedmttsrraykvdaehpgsevaaetaaamaaasivfrragdahyahlllhhaqqlfefadkyr
gryddsvrvvksfypsasgykdellwaalwlhratggrdyldyvvsnaddfggtgwgvsefswdikypglqil
asklliеekelsseyrtvlekyrskgefyvcsnmnknpggaarnaprtpaglifirpwnnlqyvtnavflltvhsd
hlaalgeplrctvdedadtsdtggetaadvvvgaeevlafakaqadyilgsnpmetsylvgygeqyprrvhhra
astasltdekgfigctqgfdswysagernphdlvgavvggpdgedrfndqrgaymqteactyntapmvgvfsr
lmelhekqrlreqrdl*

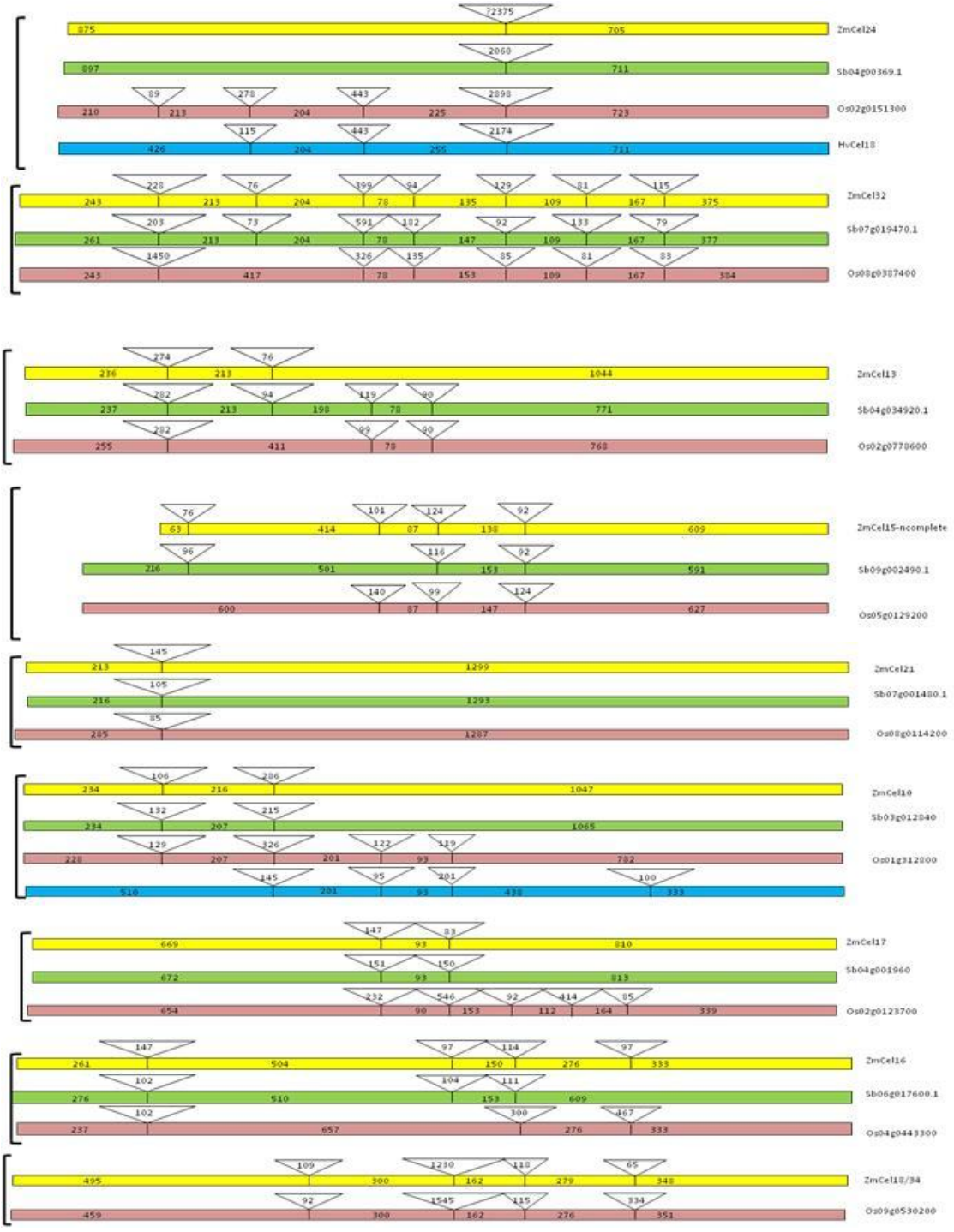
Appendix D

Shows the intron/exon structure of the barley, maize, rice and sorghum endo-β-1,4-D-glucanase gene family. Yellow bars represent maize, green=sorghum, blue= barley and brown=rice. Numbers within bars are exon nucleotide length (bps) triangles represent intron position and number within triangles represents intron length (bps). Brackets separate individual clades and include putative orthologues.




ZmCel2
Sb10g009270.1
HvCel8
HvCel4
ZmCel1
Sb04g028790.1
Os06g0256900
Os02g0733300
ZmCel24
Sb04g00369.1
Os02g0151300
HvCel18
ZmCel32
Sb07g019470.1
Os08g0387400
ZmCel13
Sb04g034920.1
Os02g0778600
ZmCel15-ncomplete
Sb09g002490.1
Os05g0129200


140 92 103 121 94 108
240 213 201 90 790 48 296
ZmCel4
130 79 104 91 93 84
234 213 201 90 796 57 302
Sb06g032760.1
258 88 683 453 82
237 213 201 385 543 299
Os04g0674800
126 104 95
207 414 90 1161
ZmCel19/28
157 98 86
207 414 90 1173
Sb09g006670.1
141 86 71 194
225 414 90 453 708
Os05g0212300
247 167 75 321
225 213 201 928 303
ZmCel20
245 206 111 133
258 213 201 985 310
Sb03g001560.1
117 149 116 89 321
246 213 201 90 165 1008
Os01g0219600
97 342 103 94 79 106 305
246 213 201 90 165 658 33 320
Os01g0220100
121 147 96 111 145
201 213 201 90 831 297
HvCel11
146 419 730 105
255 213 201 255 708
ZmCel26
183 267 650 101
228 213 201 255 702
ZmCel25
3097 534 705 117
321 213 201 255 699
Sb04g028520.1
80 467 764 102
297 213 201 255 753
Os02g0733600
577 715 88
477 201 255 663
HvCel14
115 539 825 140
294 213 201 255 666
ZmCel3
807 124
153 255 714
Sb10g008880.1-ncomplete
130 230 609 230
213 213 201 255 735
Os06g0247900
143 514 677
240 213 201 948
HvCel10
316 105 5130
516 201 255 680
ZmCel8/29
324 125 7869
474 201 255 699
Sb02g024050.1
201 6689
501 456 714
Os09g0394300
199 93 ?
474 201 255
HvCel12
285
873 648
Sb07g020700.1
1578
Os08g0425300


72375
875
705
ZmCel24
2060
897
711
Sb04g00369.1
89
278
443
2898
210
213
204
225
723
Os02g0151300
115
443
2174
426
204
255
711
HvCel18
228
76
399
94
129
81
115
243
213
204
78
135
109
167
375
ZmCel32
203
73
591
102
92
133
79
261
213
204
78
147
109
167
377
Sb07g019470.1
1450
326
135
85
81
83
243
417
78
153
109
167
384
Os08g0387400
274
76
236
213
1044
ZmCel13
282
94
119
90
237
213
198
78
771
Sb04g034920.1
282
99
90
255
411
78
768
Os02g0770600
76
101
124
92
63
414
87
138
609
ZmCel15-ncomplete
96
116
92
216
501
153
591
Sb09g002490.1
140
99
124
600
87
147
627
Os05g0129200
145
213
1299
ZmCel21
105
216
1293
Sb07g001480.1
85
285
1287
Os08g0114200
106
286
234
216
1047
ZmCel10
132
215
234
207
1065
Sb03g012040
129
326
122
119
228
207
201
93
782
Os01g312000
145
95
201
100
510
201
93
438
333
106
234
1047
ZmCel10
132
215
234
207
1065
Sb03g012040
129
326
122
119
228
207
201
93
782
Os01g312000
145
95
201
100
510
201
93
438
333
147
83
669
93
810
ZmCel17
151
150
672
93
813
Sb04g001960
232
546
92
414
85
654
90
153
112
164
339
Os02g0123700
147
97
114
97
261
504
150
276
333
ZmCel16
102
104
111
276
510
153
609
Sb06g017600.1
102
300
167
237
657
276
333
Os04g0443300
109
1230
110
65
495
300
162
279
348
ZmCel18/34
92
1545
115
334
459
300
162
276
351
Os09g0530200

Appendix E

The following tables show the data for the codon based substitution analysis using the Nei and Gojobori model and that using a maximum-likelihood model (PAML). These analyses are as described in Chapter 2, Methods 2.2.12.1 and 2.2.12.2.

	Nei and Gojobori Model - maize and sorghum						
Orthologous Pair	ds	SE	dn	SE	dn/ds	Ks	time since separation (years)
ZmCel7:ZmCel7B	0.226	0.026	0.029	0.005	0.13	6.94E-09	16294160
ZmCel7:Sb01g036480	0.206	0.025	0.029	0.005	0.14	6.67E-09	15442279
ZmCel7B:Sb01g036480	0.186	0.023	0.014	0.003	0.08	6.49E-09	14340786
ZmCel14:ZmCel30	0.27	0.03	0.01	0	0.04	5.55E-09	24324324
ZmCel14:Sb01g008860	0.22	0.03	0	0	0.00	5.05E-09	21782178
ZmCel30:Sb01g008860	0.33	0.03	0.01	0	0.03	5.5E-09	30000000
ZmCel6:Sb06g021440	0.15	0.02	0.05	0.01	0.33	3.5E-09	21428571
ZmCel12chr10:ZmCel12Chr2	0.07	0.01	0.01	0	0.14	3.5E-09	10000000
ZmCel12Chr10:Sb06g034260	0.14	0.02	0.02	0	0.14	3.6E-09	19444444
ZmCel12Chr2:Sb06g034260	0.13	0.02	0.02	0	0.15	3.5E-09	18571429
ZmCel25:ZmCel26	0.1	0.02	0.02	0.01	0.20	3.35E-09	14925373
ZmCel25:Sb04g028520	0.07	0.02	0.01	0	0.14	3.1E-09	11290323
ZmCel26:Sb04g028520	0.11	0.02	0.03	0.01	0.27	3.25E-09	16923077
ZmCel3:Sb10g008880	0.1	0.02	0.01	0	0.10	2.8E-09	17857143
ZmCel8/29:Sb02g024050	0.09	0.02	0.02	0	0.22	2.65E-09	16981132
ZmCel24:Sb04g003690	0.15	0.02	0.03	0.01	0.20	4.3E-09	17441860
ZmCel20:Sb03g001560	0.18	0.02	0.04	0.01	0.22	5E-09	18000000
ZmCel19:Sb09g006670	0.16	0.02	0.02	0	0.13	4.9E-09	16326531
ZmCel4:Sb06g032760	0.11	0.02	0.04	0.01	0.36	3.3E-09	16666667
ZmCel9:Sb10g030140	0.12	0.02	0.02	0.01	0.17	4.1E-09	14634146
ZmCel11:Sb02g030990	0.32	0.04	0.01	0	0.03	7.1E-09	22535211
ZmCel13:Sb04g034920	0.16	0.02	0.04	0.01	0.25	4.05E-09	19753086
ZmCel15:Sb09g002490	0.2	0.03	0.05	0.03	0.25	4.3E-09	23255814
ZmCel21:Sb07g001480	0.33	0.03	0.06	0.03	0.18	3.55E-09	46478873
ZmCel17:Sb04g001960	0.15	0.02	0.05	0.01	0.33	4.35E-09	17241379
ZmCel10:Sb03g012840	0.19	0.02	0.03	0.01	0.16	3.6E-09	26388889
ZmCel16:Sb06g017600	0.19	0.02	0.05	0.01	0.26	3.75E-09	25333333
ZmCel1:Sb04g028790	0.06	0.01	0.01	0	0.17	2.75E-09	10909091
ZmCel2:Sb10g009270	0.23	0.03	0.05	0.01	0.22	3.9E-09	29487179
ZmCel32:Sb07g019470	0.26	0.03	0.04	0.01	0.15	7.05E-09	18439716
ZmCel18/34:Os09g0530200	0.53	0.05	0.1	0.01	0.19	5.3E-09	50000000

	PAML - maize and sorghum								
Orthologous Pair	S	N	ds	dn	dn/ds	ds A Zm/Os	ds B Sb/Os	Ks	Years since separation
ZmCel7:ZmCel7B	478	1199	0.20	0.03	0.15	0.67	0.62	6.45E-09	15451375
ZmCel7:Sb01g036480	486	1197	0.18	0.03	0.17	0.67	0.57	6.20E-09	14513788
ZmCel7B:Sb01g036480	482	1194	0.16	0.01	0.09	0.62	0.57	5.97E-09	13403703
ZmCel14:ZmCel30	373	1363	0.32	0.01	0.02	0.57	0.72	6.42E-09	24970810
ZmCel14:Sb01g008860	441	1295	0.22	0.00	0.02	0.57	0.54	5.52E-09	19668718
ZmCel30:Sb01g008860	412	1325	0.35	0.01	0.03	0.72	0.54	6.26E-09	28322823
ZmCel6:Sb06g021440	181	1699	0.48	0.04	0.08	1.61	1.37	1.49E-08	16222663
ZmCel12chr10:ZmCel12Chr2	153	1257	0.18	0.01	0.04	1.43	1.34	1.38E-08	6340599
ZmCel12Chr10:Sb06g034260	153	1256	0.34	0.01	0.04	1.43	1.34	1.38E-08	12218405
ZmCel12Chr2:Sb06g034260	154	1255	0.31	0.01	0.04	1.34	1.34	1.34E-08	11561191
ZmCel25:ZmCel26	57	884	0.49	0.02	0.04	1.26	1.54	1.40E-08	17653583
ZmCel25:Sb04g028520	61	881	0.29	0.01	0.02	1.26	1.32	1.29E-08	11397173
ZmCel26:Sb04g028520	60	881	0.51	0.02	0.05	1.54	1.32	1.43E-08	17746735
ZmCel3:Sb10g008880	70	863	0.36	0.01	0.03	1.49	1.17	1.33E-08	13438557
ZmCel8/29:Sb02g024050	122	1260	0.31	0.02	0.06	1.45	1.45	1.45E-08	10541379
ZmCel24:Sb04g003690	132	1329	0.48	0.02	0.04	2.33	2.02	2.17E-08	11029480
ZmCel20:Sb03g001560	147	1319	0.52	0.03	0.06	1.63	1.62	1.63E-08	15935079
ZmCel19:Sb09g006670	169	1301	0.37	0.02	0.04	1.39	1.57	1.48E-08	12651945
ZmCel4:Sb06g032760	133	1328	0.34	0.03	0.10	1.77	1.71	1.74E-08	9631118
ZmCel9:Sb10g030140	145	1274	0.33	0.02	0.06	1.25	1.35	1.30E-08	12556849
ZmCel11:Sb02g030990	426	1003	0.26	0.01	0.05	0.73	0.56	6.44E-09	20245494
ZmCel13:Sb04g034920	145	1294	0.46	0.04	0.08	2.42	1.95	2.18E-08	10439837
ZmCel15:Sb09g002490	173	1099	0.42	0.04	0.11	1.17	1.31	1.24E-08	16979453
ZmCel21:Sb07g001480	166	1253	1.17	0.04	0.04	1.71	1.75	1.73E-08	33850331
ZmCel17:Sb04g001960	167	1291	0.42	0.04	0.09	1.99	2.22	2.10E-08	10082295
ZmCel10:Sb03g012840	154	1232	0.53	0.03	0.05	1.71	1.41	1.56E-08	17052348
ZmCel16:Sb06g017600	152	1240	0.51	0.04	0.08	1.06	1.18	1.12E-08	22716292
ZmCel1:Sb04g028790	131	1336	0.18	0.01	0.05	0.76	0.78	7.67E-09	12008862
ZmCel2:Sb10g009270	179	1291	0.61	0.04	0.07	1.61	1.56	1.59E-08	19193223
ZmCel32:Sb07g019470	204	1158	0.49	0.03	0.07	1.40	1.16	1.28E-08	19113293

	Nei and Gojobori - barley and rice						
Orthologous Pair	ds	SE	dn	SE	dn/ds	Ks	Years since separation
HvCel6:Os03g0329500	0.55	0.06	0.10	0.01	0.19	6.32E-09	43467933
HvCel1:Os03g0736300	0.53	0.05	0.05	0.01	0.09	5.4E-09	49074074
HvCel5:Os04g0497200	0.26	0.03	0.09	0.01	0.35	3.8E-09	34210526
HvCel14:Os02g0738600	0.3	0.04	0.05	0.01	0.17	2.75E-09	54545455
HvCel10:Os06g0247900	0.34	0.04	0.06	0.01	0.18	3.25E-09	52307692
HvCel12:Os09g0394300	0.33	0.05	0.16	0.02	0.48	2.85E-09	57894737
HvCel18:Os02g0151300	0.34	0.03	0.12	0.01	0.35	4E-09	42500000
HvCel11:Os01g0219600	0.38	0.04	0.41	0.03	1.08	4.1E-09	46341463
HvCel3:Os09g0533900	0.55	0.06	0.06	0.01	0.11	8.4E-09	32738095
HvCel7:Os01g0312800	0.34	0.04	0.07	0.01	0.21	4.35E-09	39080460
HvCel4:Os02g0733300	0.35	0.03	0.11	0.01	0.31	3.15E-09	55555556
HvCel8:Os06g0256900	0.26	0.03	0.07	0.01	0.27	3.75E-09	34666667

	PAML - barley and rice									
Orthologous Pair	S	N	ds	dn	dn/ds	ds A	ds B	Ks A/B	Years since separation	S/N
HvCel6:Os03g0329500	438	1242	0.53	0.10	0.19	0.53	0.67	6E-09	44359896	0.26
HvCel1:Os03g0736300	379	1354	0.62	0.04	0.07	0.75	0.57	6.6E-09	47402844	0.22
HvCel5:Os04g0497200	103	1295	1.32	0.07	0.05	2.03	1.43	1.7E-08	38093449	0.07
HvCel14:Os02g0738600	71	871	1.26	0.06	0.05	1.11	1.26	1.2E-08	53269215	0.08
HvCel10:Os06g0247900	83	853	5.83	0.06	0.01	2.09	1.49	1.8E-08	162854031	0.09
HvCel12:Os09g0394300	58	550	3.02	0.13	0.04	1.49	1.45	1.5E-08	102652223	0.10
HvCel18:Os02g0151300	147	1302	1.45	0.10	0.07	1.80	2.33	2.1E-08	35154112	0.10
HvCel11:Os01g0219600	186	1277	1.04	0.13	0.12	1.33	1.63	1.5E-08	35277618	0.13
HvCel3:Os09g0533900	380	1048	0.50	0.06	0.13	0.79	0.73	7.6E-09	33127667	0.27
HvCel7:Os01g0312800	129	1262	1.43	0.05	0.04	3.84	1.71	2.8E-08	25761373	0.09
HvCel4:Os02g0733300	167	1293	1.40	0.08	0.06	1.34	0.92	1.1E-08	62118730	0.11
HvCel8:Os06g0256900	144	1311	0.90	0.05	0.06	0.91	1.61	1.3E-08	35670913	0.10
								Average	40269808	
								St dev	9924428	

Appendix F

Table 1: Data (log intensity) from the PLEXdb Affymetrix BB3: Transcription Patterns during Barley development experiment.

	ANTHER_MX	BRACTS_MX_	CAR5(caryopsis 5 DAP)_MX	CAR10(caryopsis 10 DAP)_MX	CAR16(caryopsis 16 DAP)_MX	COLEOPTILE_MX	COLEOPTILE_GP	CROWN_MX	CROWN_GP	DEM22(embryo)_MX	ENDOSPERM22_MX
Cel1	7	12	10	10	9	9	9	10	11	8	9
Cel2	7	7	7	6	7	7	7	8	7	7	7
Cel4	7	7	8	7	8	10	10	10	10	9	8
Cel5	9	10	10	9	10	9	9	9	10	8	9
Cel6	10	9	11	11	11	9	9	9	9	10	10
Cel7	8	7	8	7	7	8	8	8	7	7	7
Cel8	12	12	9	9	9	11	11	10	10	9	9
Cel9	8	9	9	9	9	9	7	9	7	9	9
Cel10	8	11	7	8	8	7	7	8	8	7	7
Cel11	7	8	12	7	7	7	7	6	7	8	8
Cel12	10	10	10	10	10	10	10	10	10	10	10
	ENDOSPERM22_MX	GEM_MX	GEM_GP	INFLORENSCENCE_MX	LEAF_MX	LEAF_GP	PISTIL_MX	RADULA_GP	RADULA_MX	ROOT_MX	ROOT_GP
Cel1	9	9	10	7	12	12	9	8	8	11	11
Cel2	7	7	7	7	7	7	7	7	7	7	7
Cel4	8	7	8	11	7	8	9	7	7	7	7
Cel5	9	10	10	7	10	9	10	11	12	11	11
Cel6	10	9	9	10	7	8	11	8	8	8	8
Cel7	7	8	8	8	7	7	8	8	8	8	8
Cel8	9	11	11	11	9	10	12	11	11	10	10
Cel9	9	9	7	10	8	7	9	7	8	8	7
Cel10	7	8	8	7	9	8	7	9	9	10	10
Cel11	8	7	7	7	6	7	12	7	7	7	7
Cel12	10	9	10	10	9	9	10	10	10	10	10

Table 2; Shows the normalized transcriptional results (copies mRNA per ul normalized cDNA) of 12 barley endo-β-1,4-D-glucanase genes across 16 barley tissues.

	HvCel1	HvCel2	HvCel3	HvCel4	HvCel5	HvCel6	HvCel7	HvCel8	HvCel10	HvCel11	HvCel12	HvCel14
Leaf tip	5982	145	1800	48	41312	525	32	194	65	24	48	7327
Leaf base	69486	255	239255	266	155855	631	185	4773	4426	64	411	40396
Root tip	1968	3242	35967	1853	332937	2157	30105	15342	4096	854	261	14228
Root base	6788	737	27166	407	141078	750	6598	5780	2982	300	252	20316
Anther at pre-anthesis	4496	3708	9177	255	104564	58052	950	11309	2711	718	1159	2178
Pistil at pre-anthesis	3289	3725	21160	1832	21058	12714	19871	6435	179	46044	507	937
Whole flower at anthesis	3613	5465	16229	1794	39051	14777	68865	5769	323	67597	1078	2067
Spike	9003	7421	68050	3295	23248	5157	3045	779	655	57	852	2482
Whole grain 3-5dap	4060	2190	13233	1520	12652	12959	36719	330	131	79001	391	106
Whole grain 8-10 dap	1644	229	5012	93	8230	22635	239	97	1242	1489	206	323
Embryo 22dap	748	107	407	24	4410	2344	0	0	14	12	93	48
Coleoptile 3 days germ	989	799	6497	531	5802	273	1652	524	36	1	135	26
Stem	15582	417	70413	30	8498	503	98	68	204	102	98	10266
Peduncle	64009	2134	220424	489	43527	1156	1049	334	608	382	417	55830
Mature leaf	3278	12	12211	0	8939	113	0	18	54	30	161	1971
Scutellum	1411	129	3469	57	48092	1488	82	4	12	4	83	1284
	Standard deviation											
	6327	135	743	24	1021	98	56	48	37	24	24	1250
	3470	44	31029	44	46501	195	82	541	2916	80	135	1795
	821	550	2621	379	12838	329	2315	1474	3486	403	301	2578
	1190	189	3622	68	1394	92	917	624	1025	83	68	949
	2858	1078	1565	175	6235	9859	452	4244	4217	175	488	599
	152	463	3345	139	1025	1229	1540	649	47	3843	119	21
	605	516	1948	170	842	332	3057	2073	83	5303	166	239
	1444	705	1875	603	827	190	158	54	111	5	58	419
	2167	99	1081	110	361	803	847	10	138	1764	55	10
	117	43	409	28	543	2072	324	57	2039	116	40	34
	161	26	223	8	76	119	0	0	20	4	26	11
	31	699	437	3	153	18	134	76	13	0	21	3
	5657	98	1469	20	2187	20	102	27	84	26	53	710
	13913	570	5545	41	4360	115	268	83	95	109	74	1722
	900	21	1558	0	1061	10	0	31	18	27	18	430

Table 3: Shows raw QPCR transcript data for the four control genes used for normalization of the raw endo-β-1,4-D-glucanase transcript data collected for barley stem series.

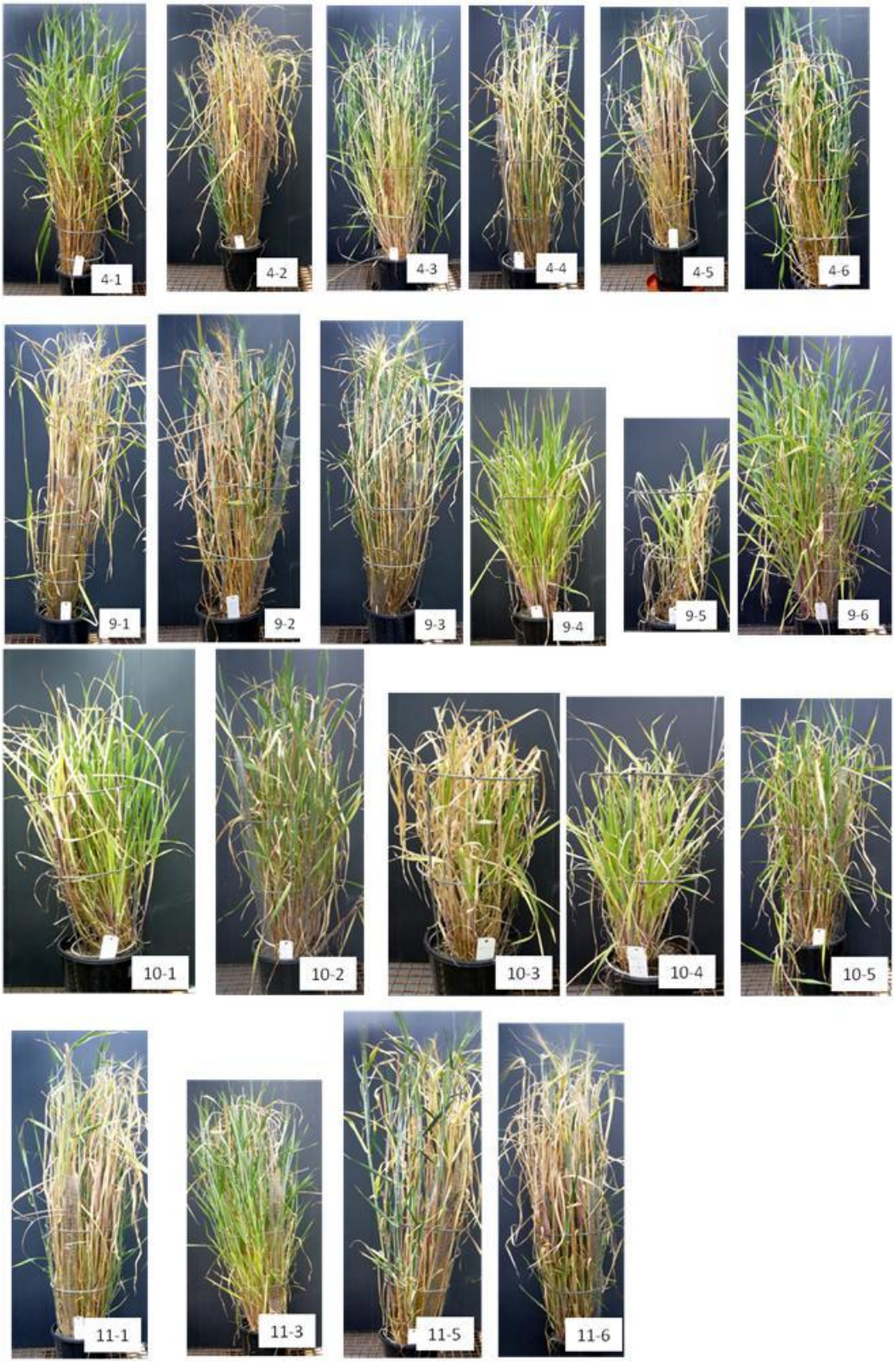
Raw data	Control1	Control2	Control3	Control4
	GAPHv	Cyclophilin	TubulinHv	HSP70Hv
B1	552253	1317497	2528010	54763
B2	4148043	3954827	47981293	648897
B3	865647	2667803	14999583	46057
B4	8086277	4087790	41810280	276650
B5	6558193	3753777	23462623	75847
B6	7169633	3922033	21695513	180590
B7	189933	1388010	309227	29370

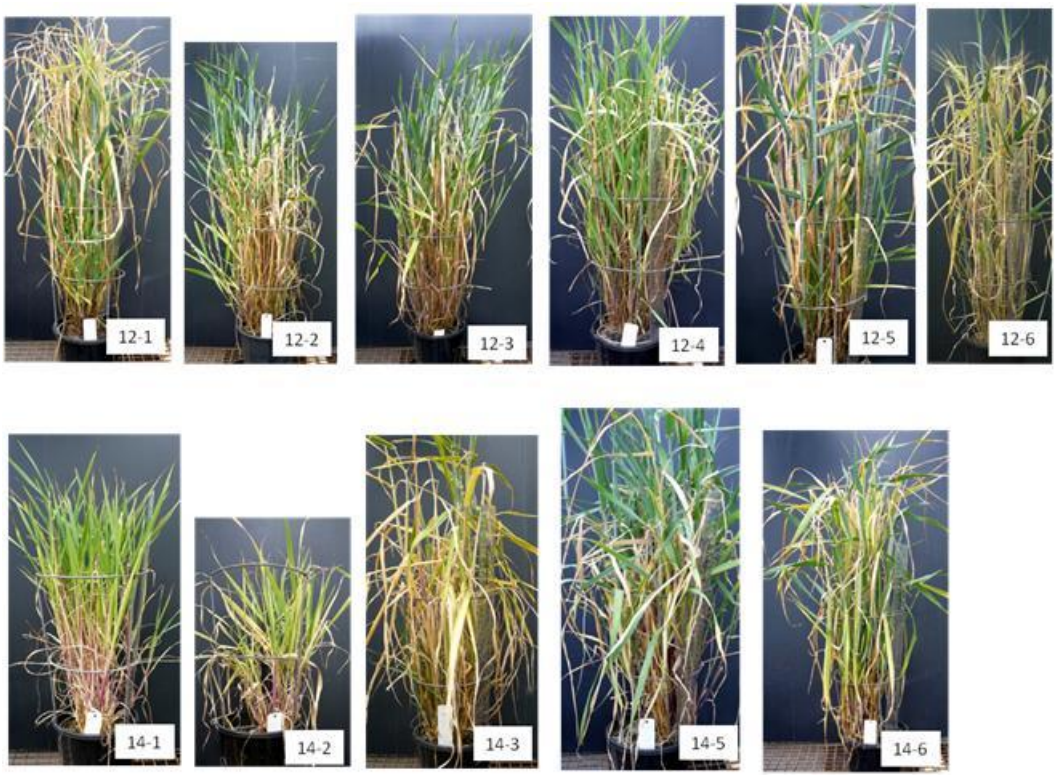
Table 4: shows normalized QPCR data for the barley stem series including global normalization factor (GNF) used to normalize the raw data. The GNF is derived from the three control genes showing the least between gene transcript variation across the series. Copies of mRNA transcript per ul normalised cDNA was derived by dividing the raw transcript data for the endo-β-1,4-D-glucanase genes by the global normalisation factor (GNF).

Global Normalised data														
	Global GNF	HvCel1	HvCel2	HvCel3	HvCel4	HvCel5	HvCel6	HvCel7	HvCel8	HvCel10	HvCel11	HvCel12	HvCel14	HvCel16
B1	0.403	15674	16706	222015	14509	37610	6973	793	4643	7048	58	2322	6387	405
B2	2.598	21235	1355	363094	604	13201	371	136	1936	4112	46	435	7834	73
B3	0.560	67810	1126	1732590	1441	51044	1364	405	1161	17797	578	250	116567	304
B4	2.470	26241	463	694892	529	9357	730	381	1031	1739	601	89	23767	204
B5	1.455	22828	71	536222	241	7659	520	39	408	2340	66	83	8168	60
B6	2.030	14793	110	298375	172	5746	412	46	721	1711	94	97	5812	31
B7	0.234	2354	2596	4607	4008	17060	4094	1241	8273	913	3252	2240	4080	1113
working SD data														
B1		1612	187	7006	2095	2321	3455	114	955	330	62	254	234	151
B2		2026	213	3641	64	3859	201	21	157	114	20	72	911	14
B3		1034	234	80653	758	2129	500	92	294	1168	144	64	12752	71
B4		8162	52	15049	203	219	44	60	99	124	43	31	2810	34
B5		794	56	18291	89	435	132	4	314	95	17	66	302	22
B6		409	30	11975	78	855	149	15	904	386	10	35	804	32
B7		428	321	624	1052	1547	236	154	11460	462	260	813	575	340

Appendix G

The following photos represent the surviving T1 plants at 5 months of age.

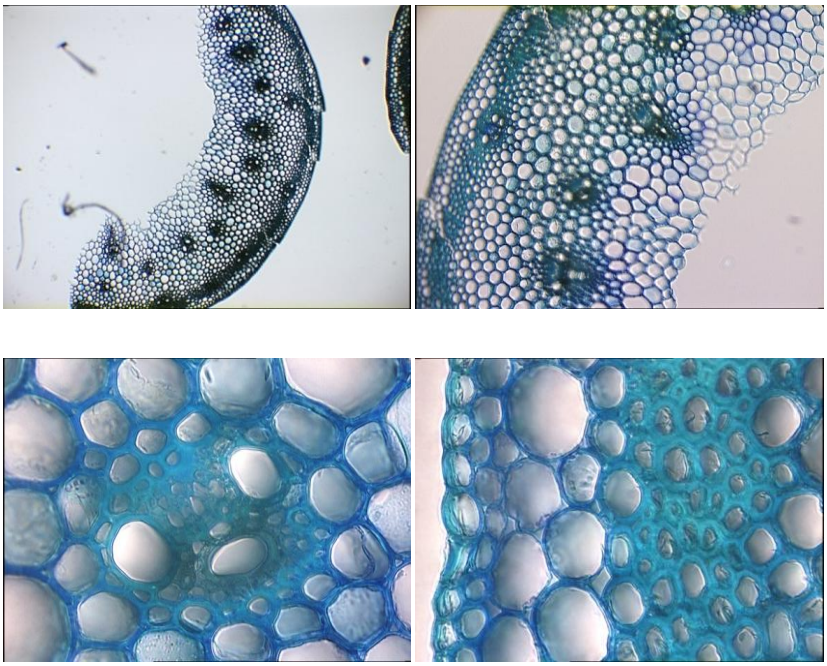



12-1
12-2
12-3
12-4
12-5
12-6
14-1
14-2
14-3
14-5
14-6

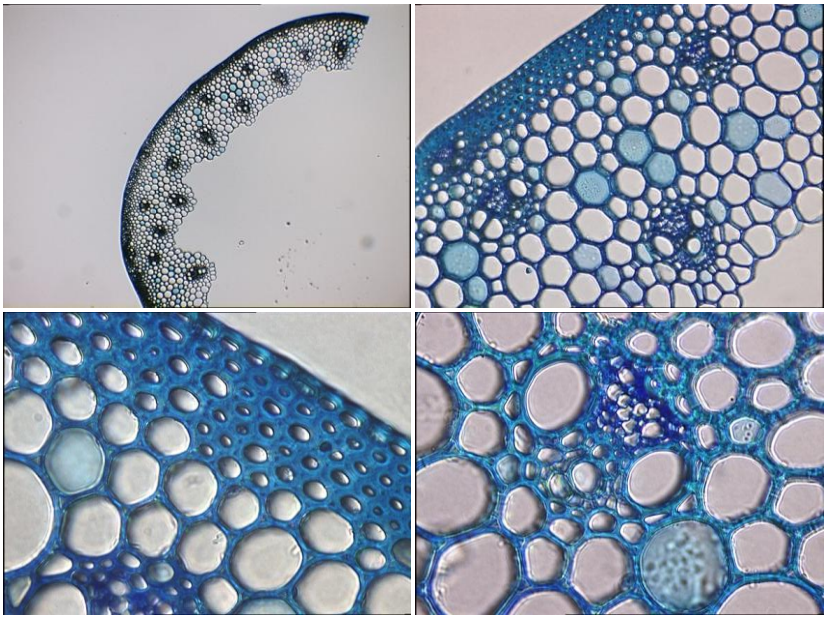
Appendix H

Images below show the glycerol preparations of the 12 T_1 plant stems that were stained with toluidine blue.

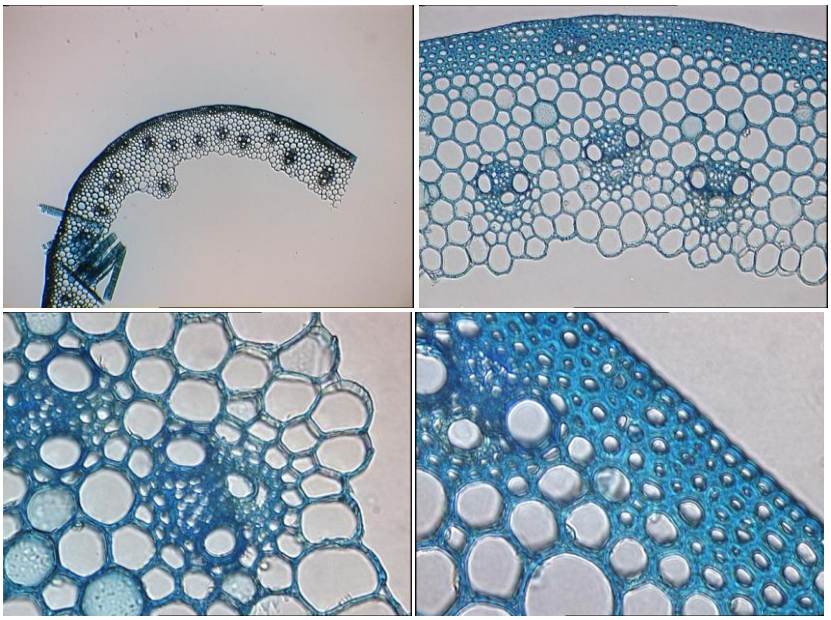
1. Glycerol preps of plant 235-12-1 2nd internode from the top



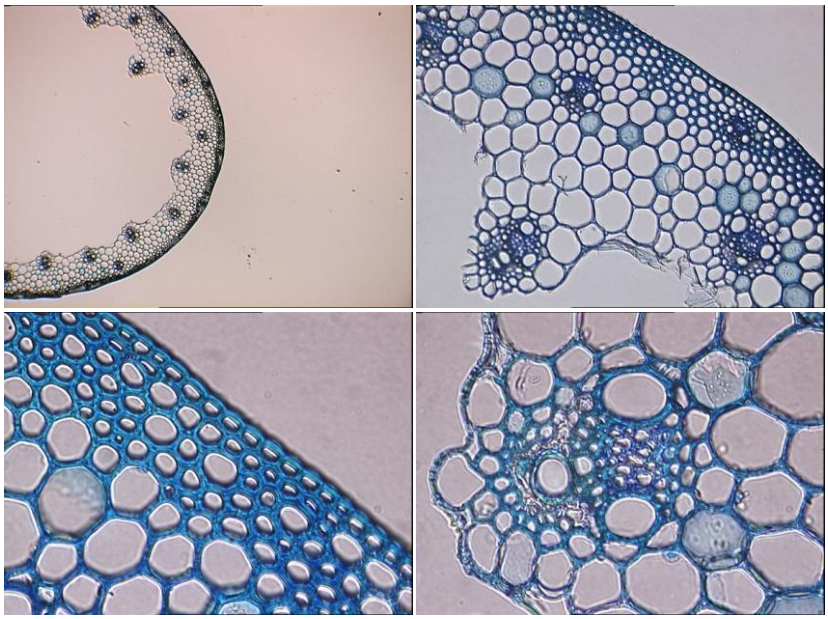
2. Glycerol preps 235-12-5- 2nd internode from the top



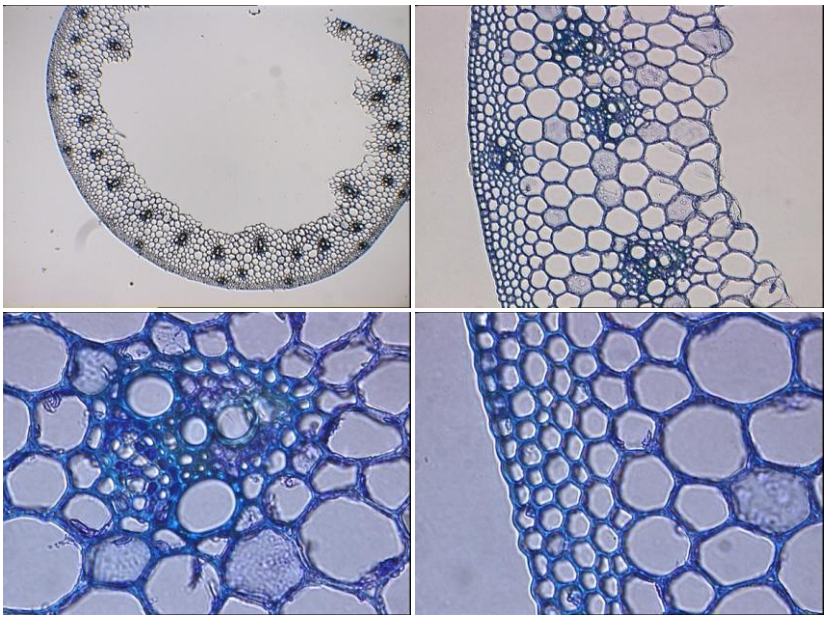
Glycerol preps of plant 235-12-5 4th internode from the top



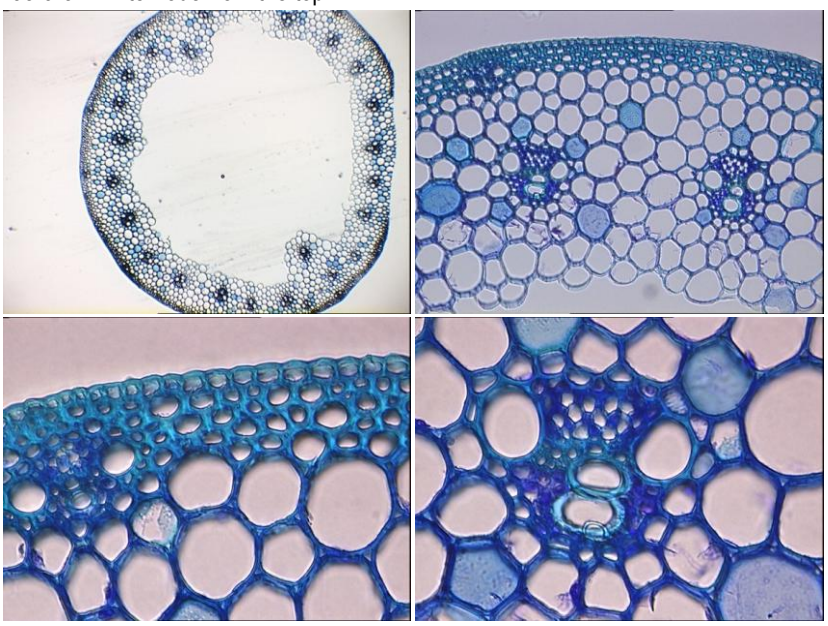
3. 235-4-5-2 2nd internode from the top



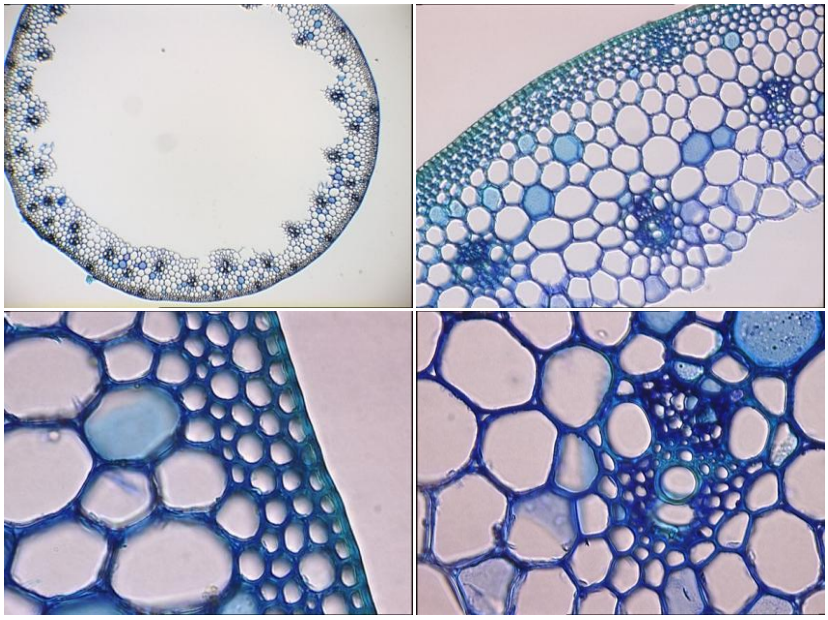
4. 235-9-5 2nd internode from the top



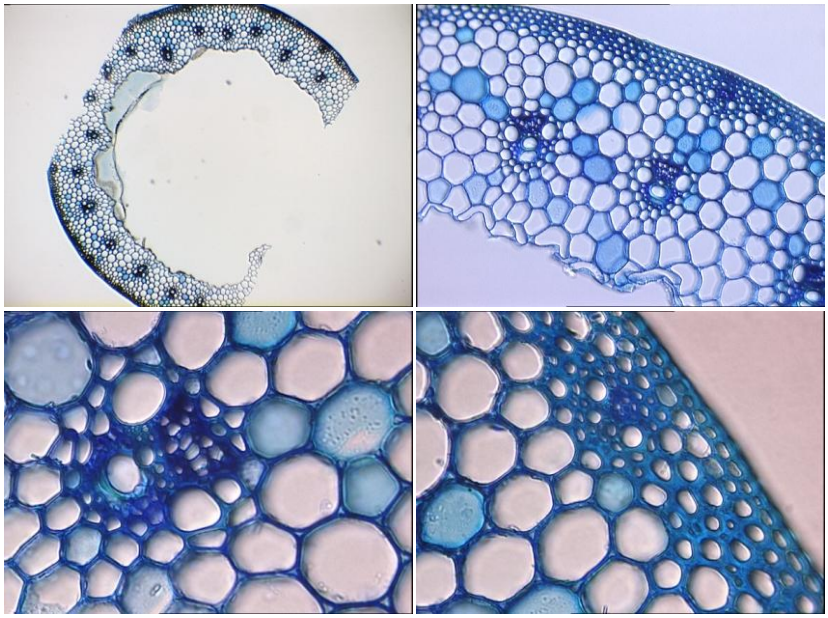
235-9-5 4th internode from the top



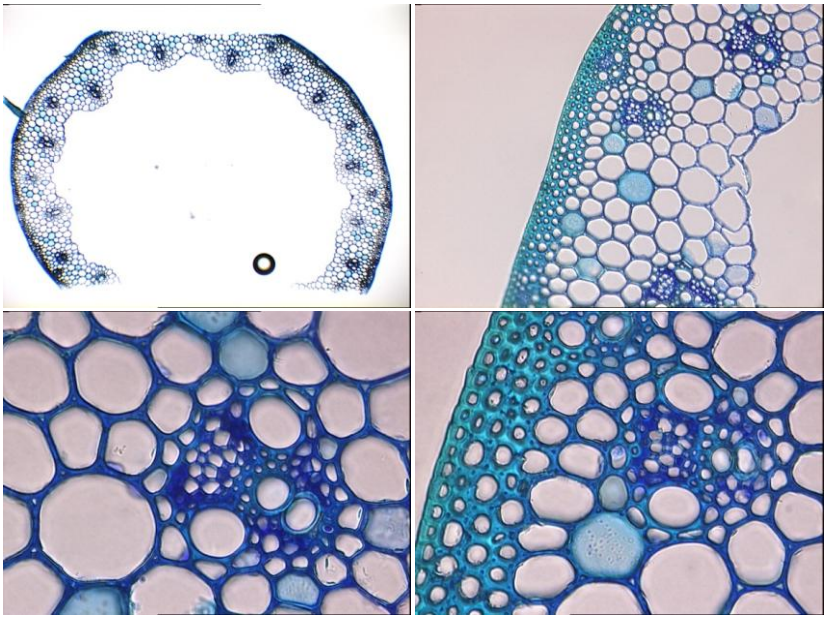
5. 235-9-4-2nd internode from the top



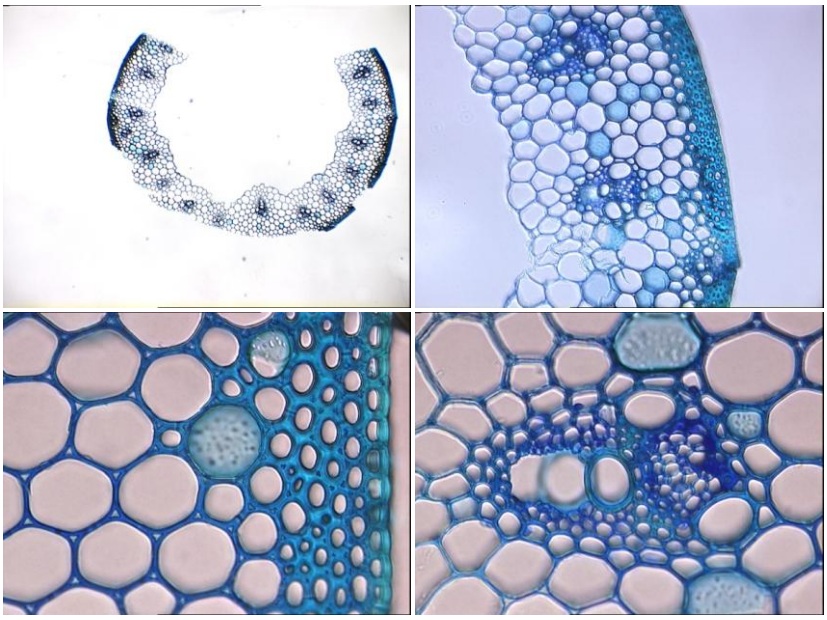
235-9-4-4th internode from the top



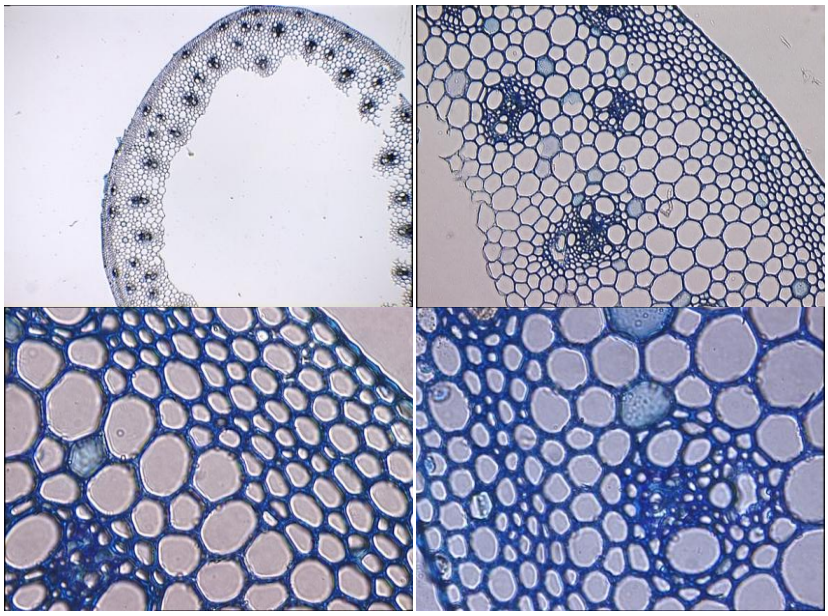
6. 235-10-2 2nd internode from the top



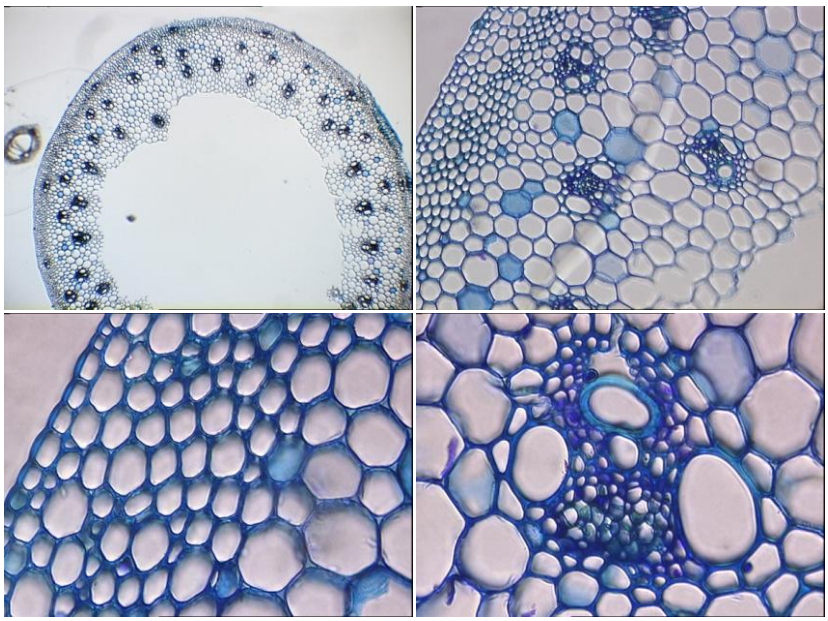
235-10-2 4th internode from the top



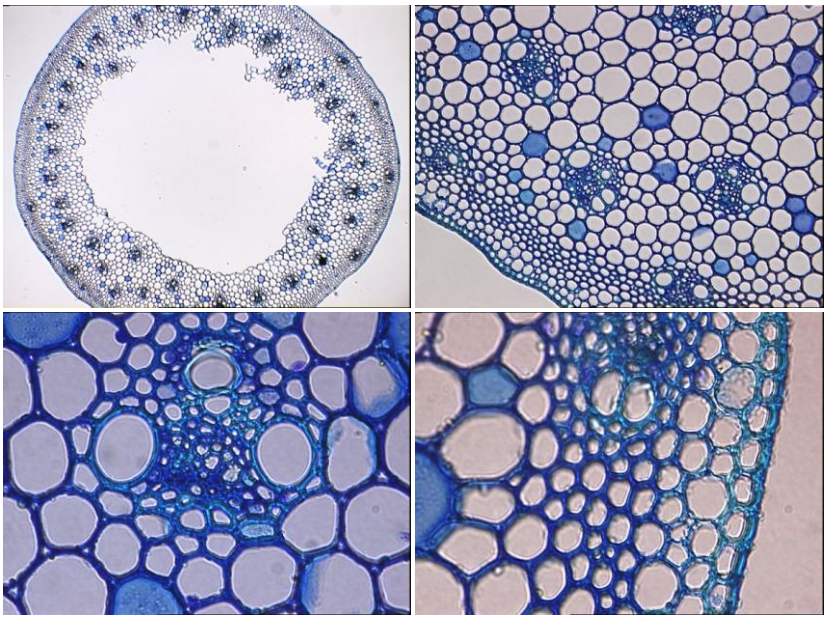
7. 235-10-3 2nd internode from the top



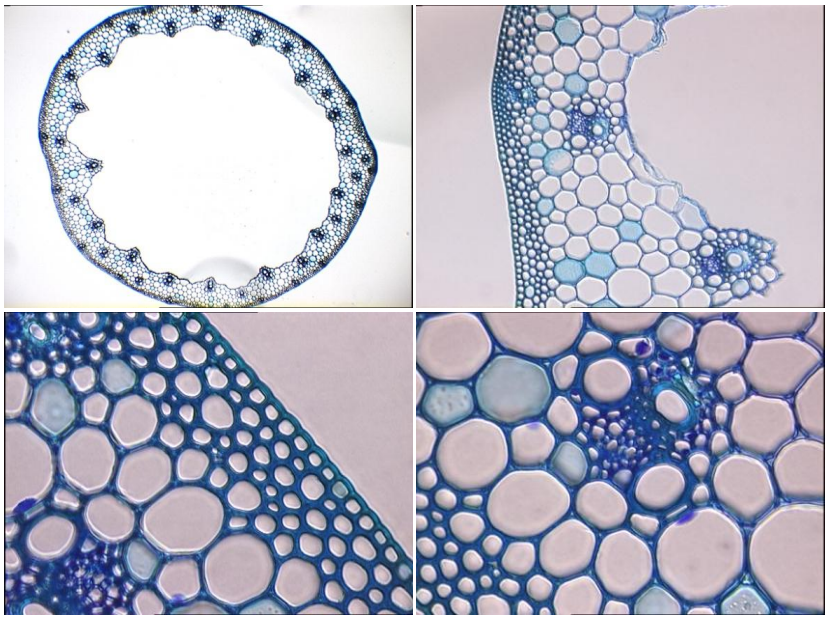
235-10-3-4th internode from the top



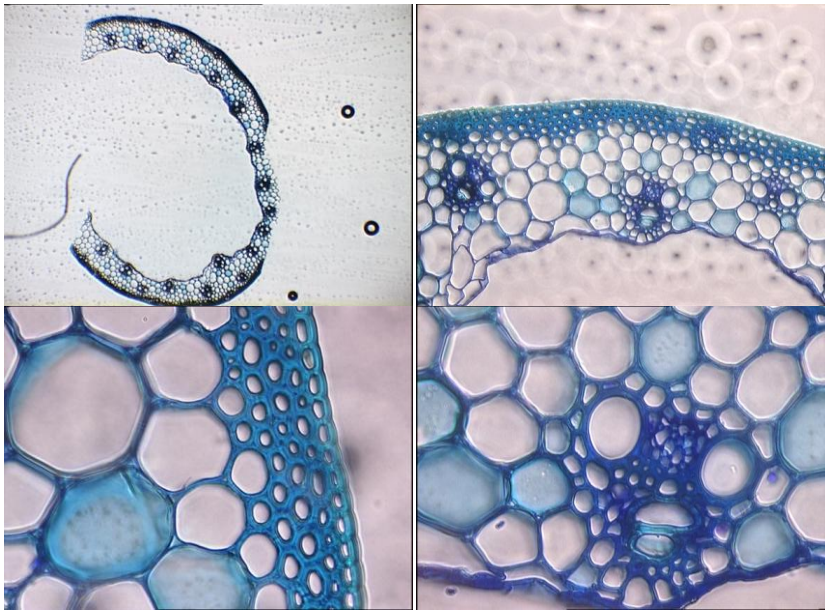
8. 235-10-4 2nd internode from the top



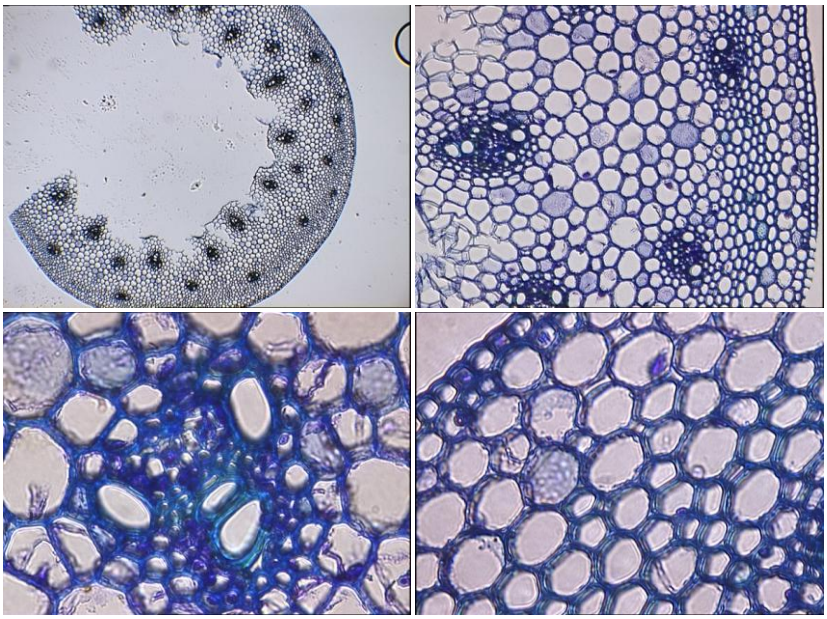
9. 235-11-3-2nd internode from the top



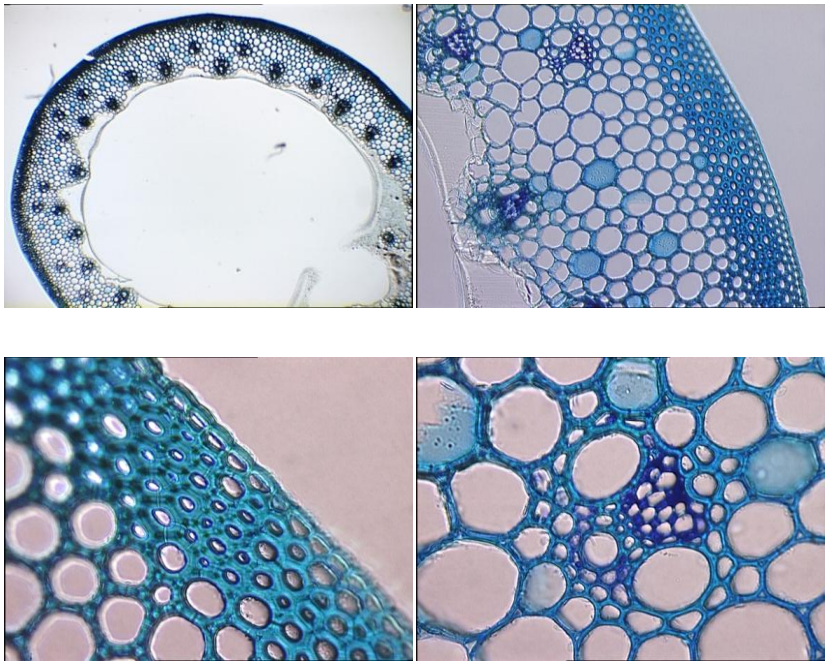
235-11-3-4th internode from the top



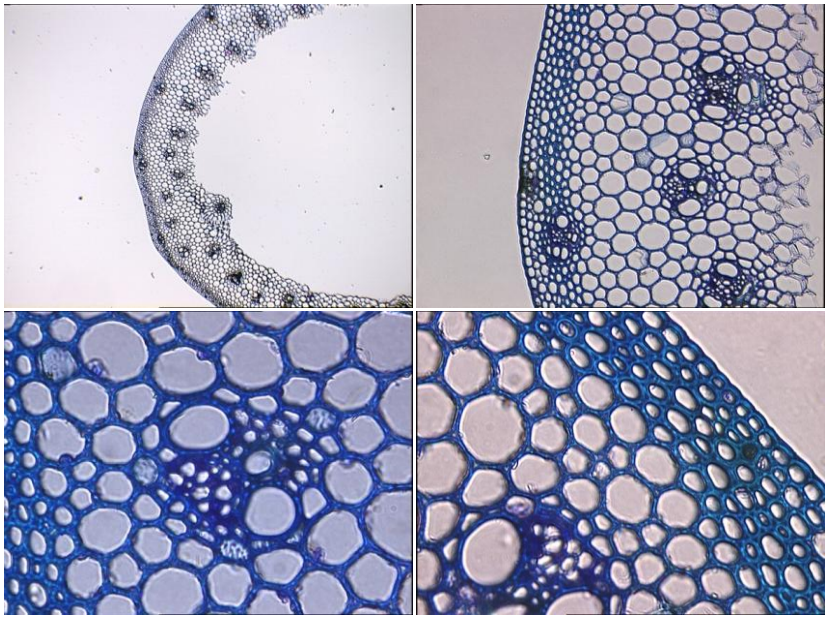
10. 235-14-1 2nd internode from the top



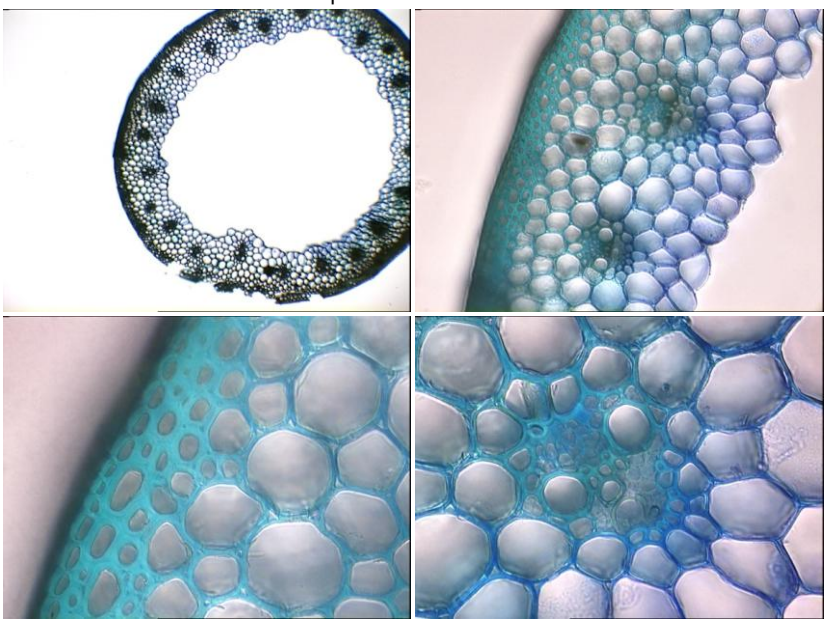
235-14-1-4th internode from the top



11. 235-14-3-2nd internode from the top



12. 235-14-5-2nd internode from the top



235-14-5-4th internode from the top

