An investigation of bread wheat meiosis via proteomics and gene-targeted approaches: the isolation and characterisation of four meiotic proteins

by

Kelvin Khoo Han Ping
B. Science (Hons.), The University of Adelaide

A thesis submitted for the degree of
Doctor of Philosophy

at

The University of Adelaide
Faculty of Sciences
School of Agriculture, Food and Wine
Waite Campus

August 2011

Table of contents

Table of contents ..... **III**

List of figures ..... **XV**

List of tables ..... **XIX**

Abstract ..... **XX**

Declaration ..... **XXIV**

Acknowledgements ..... **XXV**

Glossary of abbreviations ..... **XXVII**

Chapter 1 – Literature review **1**

1.1 Meiosis and meiosis in bread wheat 1

1.1.1 – A brief overview of meiosis 1

1.1.2 – Meiosis in bread wheat 5

1.2 – Meiotic processes 7

1.2.1 – Pre-meiotic chromosome interactions 7

1.2.2 – Telomere bouquet formation 9

1.2.3 – Homologous chromosome alignment 11

1.2.4 – Recombination, synapsis, and pairing 12

1.2.4.1 – Recombination and its effects on chromosome pairing 12

1.2.4.2 – Poor Homologous Synapsis 1 (PHS1) – a role in regulating recombination and homology searching? 15

1.2.4.3 – Synaptonemal complex formation and synapsis 17

1.2.4.4 – The roles of ASY1 and ZYP1 in SC formation and synapsis 19

1.2.5 – Homoeologous pairing of chromosomes in bread wheat.....22

1.3 – Proteomics: A global approach to cell biology.....26

1.3.1 – 2-dimensional gel electrophoresis (2DGE).....28

1.3.2 – A proteomics approach to plant meiosis.....28

1.4 – Rationale of the current study.....30

Chapter 2 – Identification of proteins with potential roles in meiosis via an optimised proteomics approach.....32

2.1 – Introduction.....32

2.2 Materials & Methods 34

2.2.1 – Staging and harvesting of wheat anthers and meiocytes 34

2.2.1.1 – Collection of meiotic tissue……....34

2.2.1.2 – Staging and harvesting of meiotic anthers and meiocytes……..34

2.2.2. – Optimising the protein extraction method from whole wheat anthers and meiocytes 37

2.2.2.1 – Trichloroacetic acid (TCA)-acetone extraction and precipitation ……37

2.2.2.2 – Urea-mercaptoethanol-NP40 extraction with acetone precipitation……38

2.2.2.3 – Phenol extraction with methanol/ammonium acetate precipitation ……38

2.2.3 – Protein Quantification.....39

2.2.4 – Visual assessment of protein quality via sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE).....40

2.2.4.1 – SDS-PAGE……..……..40

2.2.4.2 – Fixing, staining and storage of Bis-Tris gels.....41

2.2.5 – 2-Dimensional Gel Electrophoresis (2DGE) 41

2.2.5.1 – First dimension: Isoelectric focusing (IEF) of 2DGE samples...41

2.2.5.2 – Equilibration of IEF strips……..42

2.2.5.3 – Second dimension: SDS-PAGE……..….....43

2.2.6 – 2DGE gel staining.....43

2.2.7 – 2-Dimensional Fluorescence Difference Gel Electrophoresis (DIGE) 45

2.2.7.1 – Sample pH adjustment……45

2.2.7.2 – Preparation of CyDyes and protein labelling……..45

2.2.7.3 – IEF and SDS-PAGE of DIGE samples……...46

2.2.7.4 – Visualisation of DIGE protein spots……...46

2.3 – Results.....46

2.3.1 – Optimisation of a suitable protein extraction technique 46

2.3.2 – Optimisation of 2DGE gel resolution 50

2.3.2.1 – Horizontal resolution (1st dimension of 2DGE)…..……50

2.3.2.2 – Vertical resolution (2nd dimension of 2DGE)…….56

2.3.3 – Optimising a spot visualisation technique 56

2.3.3.1 – 2-dimensional fluorescence difference gel electrophoresis (DIGE)……57

2.3.3.2 – MS compatible eriochrome black T (EBT) silver staining ……59

2.3.4 – 2DGE of anther proteins from four pooled stages of meiosis 60

2.3.5 – 2DGE of PM-LP meiocyte-enriched samples 62

2.4 – Discussion.....67

Chapter 3 – Characterisation of proteomics candidates.....75

3.1 – Introduction.....75

3.2 – Materials & Methods.....76

3.2.1 – Mass-peptide identification of candidate spots.....76

3.2.1.1 – Preparation of candidate protein spots……76

3.2.1.2 – Mass spectrometry (MS)……77

3.2.2 – Analyses of peptides identified by mass spectrometry.....77

3.2.2.1 – Basic Logic Alignment Search Tool (BLAST) analyses of peptides obtained from mass spectrometry……77

3.2.3 – Isolation and amplification of candidate transcripts.....78

3.2.3.1 – Primer design……..78

3.2.3.2 – Collection and staging of meiotic anthers……...79

3.2.3.3 – RNA extraction and quantification……79

3.2.3.4 – Construction of cDNA and RACE libraries……80

3.2.3.5 – 5′ and 3′ RACE Polymerase Chain Reaction (PCR)……...80

3.2.3.6 – Agarose gel electrophoresis and gel purification of PCR products ……81

3.2.3.7 – Ligation of PCR products and bacterial transformation…….81

3.2.3.8 – Colony PCR screening: Identification of positive pCR®8/GW/TOPO® clones with candidate cDNA inserts…….82

3.2.3.9 – Sequencing reactions for positive pCR®8/GW/TOPO®-candidate insert clones……83

3.2.3.10 – Clean-up of sequencing reactions…….84

3.2.3.11 – Contig construction and sequence analysis of pCR®8/GW/TOPO®-candidate insert clones…….84

3.2.4 – Isolation and amplification of full length candidate open reading frames (ORFs).....85

3.2.4.1 – High fidelity amplification and purification of candidate ORFs from meiotic cDNA……85

3.2.4.2 – Production, screening and sequencing of pCR®8/GW/TOPO®-candidate ORF clones……86

3.2.4.3 – Contig construction and sequence analysis of pCR®8/GW/TOPO®-candidate ORF clones……..86

3.2.5 – Quantitative real-time PCR (Q-PCR) 86

3.2.5.1 – Q-PCR primer design……..86

3.2.5.2 – Testing of Q-PCR primers……..87

3.2.5.3 – Q-PCR…….87

3.2.6 – Chromosome location of candidates.....90

3.2.6.1 – Southern blot analyses using nullisomic-tetrasomic membranes.……90

3.2.6.1.1 – Pre-hybridisation of nullisomic-tetrasomic nylon membrane……..90

3.2.6.1.2 – Probe labelling and hybridisation……91

3.2.6.1.3 – Membrane washes……...91

3.2.6.1.4 – Autoradiography……..92

3.2.6.2 – Southern blot analyses using genotype membranes ……...92

3.2.6.2.1 – Genomic DNA extraction……92

3.2.6.2.2 – Enzymatic digestion of genomic DNA……93

3.2.6.2.3 – Capillary transfer of digested genomic DNA……..93

3.2.6.2.4 – Southern blot procedure……..94

3.2.6.3 – *In silico* mapping of the *KK06* gene candidate……...94

3.2.7 – Preparation of candidate ORF clones for heterologous protein expression 94

3.2.7.1 – Construction of pDEST17-candidate ORF clones……..94

3.2.8 – Protein production of candidates 96

3.2.8.1 – Transformation of BL21A1 *E. coli* cells with the pDEST17-candidate ORF clones……96

3.2.8.2 – Culturing of pDEST17-candidate ORF-transformed BL21-A1 cell lines…….96

3.2.8.3 – Induction of protein production and subsequent cell harvesting ……97

3.2.9 – Denatured extraction, purification and visualisation of proteins.....97

3.2.9.1 – Cell lysis and protein extraction……..97

3.2.9.2 – Purification of 6×His-tagged candidate proteins via nickel-affinity chromatography…….98

3.2.9.3 – SDS-PAGE for separation of proteins by molecular weight…..99

3.2.10 – Native extraction, purification and visualisation of proteins.....99

3.2.10.1 – Cell lysis and protein extraction……99

3.2.10.2 – Purification of 6×His-tagged candidate proteins via nickel-affinity chromatography……...100

3.2.10.3 – SDS-PAGE for separation of proteins by molecular weight..101

3.2.11 – Functional analyses of candidates.....101

3.2.11.1 – Competitive DNA-binding assay……101

3.3 – Results.....101

3.3.1 – Mass-peptide analyses of 2DGE protein spots 101

3.3.2 – Isolation and characterisation of the candidate gene transcripts.....105

3.3.3 – Characterisation of the candidate protein products.....108

3.3.4 – Chromosome location of the candidate genes 111

3.3.5 – Q-PCR analysis of candidate genes 114

3.3.6 – *In vitro* characterisation of proteins encoded by the candidate genes 116

3.3.6.1 – Extraction and isolation of protein products encoded by candidates transcripts……...116

3.3.6.2 – DNA-binding capabilities of KK01 and KK06 full-length proteins.……118

3.4 – Discussion 120

Chapter 4 – Poor homologous synapsis 1 (PHS1) interacts with chromatin but does not co-localise with ASYnapsis 1 (ASY1) during early meiosis in bread wheat 131

Chapter 4 - Addendum.....169

4.1 – Introduction.....169

4.2 – Materials & Methods 170

4.2.1 – Primer design 170

4.2.1.1 – *PHS1* primer design……..170

4.2.1.2 – *RAD50* primer design……...170

4.2.2 – Preparation of meiotic cDNA 171

4.2.2.1 – cDNA synthesis from purified meiotic RNA……171

4.2.3 – Preparation of *RAD50* and *PHS1* clones.....171

4.2.3.1 – High fidelity amplification and purification of *RAD50* and *PHS1* from meiotic cDNA……..171

4.2.3.2 – Ligation of *RAD50* and *PHS1* PCR products and bacterial transformation……..172

4.2.3.3 – Colony PCR: identification of positive clones…….....172

4.2.4 – Sequencing of *RAD50* and *PHS1* clones 173

4.2.4.1 – Sequencing PCR and clean-up……..173

4.2.4.2 – Contig construction and sequence analysis of clones……...173

4.2.5 – Construction of Y2H constructs 174

4.2.5.1 – Construction of pGADT7-*TaPHS1* activation domain prey plasmid…….174

4.2.5.2 – Construction of pGBKT7-*TaRAD50* binding domain bait plasmid ……..175

4.3 – Results.....176

4.3.1 – Construction of pGADT7-*TaPHS1* prey plasmid.....176

4.3.2 – Isolation and bioinformatic analyses of *TaRAD50* 177

4.3.3 – Construction of the pGBKT7-*TaRAD50* bait plasmid.....181

4.4 – Discussion.....182

Chapter 5 – The isolation and characterisation of bread wheat *Molecular ZIPper 1* (*TaZYP1*) 184

5.1 – Introduction.....184

5.2 – Materials & Methods 186

5.2.1 – Primer design 186

5.2.2 – Preparation of meiotic cDNA and 3′RACE libraries.....187

5.2.2.1 – Synthesis of libraries…….187

5.2.3 – Preparation of *ZYP1* clones 188

5.2.3.1 – High fidelity amplification and isolation of *ZYP1* from meiotic cDNA……...188

5.2.3.2 – Production and sequencing of pCR®8/GW/TOPO®-*ZYP1* clones……188

5.2.4 – Chromosomal location of *TaZYP1* 189

5.2.4.1 – Southern blot analysis using nullisomic-tetrasomic membranes ……..189

5.2.5 – Expression analysis of *TaZYP1* 189

5.2.5.1 – Quantitative real-time PCR (Q-PCR) 189

5.2.6 – Analyses of *TaZYP1* and its protein product 1900

5.2.6.1 – Comparative amino acid and conserved domain analyses of *Ta*ZYP1……189

5.2.7 – *Ta*ZYP1 protein production, extraction and purification 191

5.2.7.1 – Construction of pDEST17-*TaZYP1* ORF……..191

5.2.7.2 – Protein production of *Ta*ZYP1……..191

5.2.7.3 – Protein extraction, purification and SDS-PAGE analyses……191

5.2.8 – Mass-peptide identification of *Ta*ZYP1 191

5.2.8.1 – Sample preparation, MS, and peptide identification…….191

5.2.9 – Functional analysis of *Ta*ZYP1 192

5.2.9.1 – Competitive DNA-binding assay……..192

5.2.10 – Production of the anti-*Ta*ZYP1 polyclonal antibody.....192

5.2.10.1 – Polyclonal antibody production against whole *Ta*ZYP1 protein ……..192

5.2.10.1.1 – Protein sample clean-up……..192

5.2.10.1.2 – Determining protein sample quality and quantity……...193

5.2.10.1.3 – Separation of purified protein samples and polyclonal antibody production…….193

5.2.10.1.4 – Clean-up of polyclonal antibody raised against full-length *Ta*ZYP1 antigen……..194

5.2.10.2 – Polyclonal antibody production against a partial *Ta*ZYP1 peptide...……..194

5.2.10.2.1 – Partial peptide synthesis and polyclonal antibody production……194

5.2.11 – Immuno-localisation of *Ta*ZYP1 by fluorescence microscopy 195

5.2.11.1 – Fluorescence immuno-localisation of *Ta*ZYP1 and *Ta*ASY1 ……..195

5.3 – Results.....196

5.3.1 – Isolation and characterisation of the *TaZYP1* coding sequence 196

5.3.2 – Characterisation of the *TaZYP1* protein product *in vitro*.....199

5.3.2.1 – Heterologous expression of *Ta*ZYP1.....199

5.3.2.2 – *Ta*ZYP1 interacts with DNA.....200

5.3.3 – Transcript expression analysis of *TaZYP1*.....201

5.3.3.1 – Q-PCR analysis of *TaZYP1* in wild-type and the *ph1b* mutant ……..201

5.3.3.2 – Q-PCR analysis of *TaZYP1* in *Taasy1* mutant lines…….203

5.3.4 – Fluorescence immuno-localisation of *Ta*ZYP1 205

5.3.4.1 – Localisation profile of *Ta*ZYP1 in wild-type bread wheat…..205

5.3.4.2 – Characterising *Ta*ZYP1 localisation in the *ph1b* bread wheat mutant…….....207

5.3.4.3 – Characterising *Ta*ZYP1 localisation in *Taasy1* bread wheat mutants…….210

5.4 – Discussion.....213

Chapter 6 – General discussion.....223

6.1 – A large scale proteomics investigation of meiosis.....223

6.1.1 – Optimisation of 2DGE for studying wheat meiosis.....223

6.1.2 – Future directions for bread wheat meiosis using proteomics 224

6.1.3 – Proteomics and the identification of novel meiosis candidates 226

6.1.4 – Future research directions for the protein candidates identified.....228

6.2 – A gene-targeted approach to study meiosis: *TaPHS1* and *TaZYP1*.....229

6.2.1 – *TaPHS1*.....229

6.2.2 – A model for homology searching in wheat.....230

6.2.3 – Future directions of *Ta*PHS1 research.....234

6.2.4 – *Ta*ZYP1.....236

6.2.5 – Future directions of *Ta*ZYP1 research.....237

6.3 – The bigger picture.....238

References.....240

Appendix A.....254

Appendix B1.....255

Appendix B2.....339

Appendix B3 ..... **342**

Appendix C1 ..... **343**

Appendix C2 ..... **344**

Appendix D1 ..... **347**

Appendix E **349**

List of figures

Figure 1.1 – The five sub-stages of prophase I followed by the major stages of meiosis in regal lily (*Lilium regale*) ...…….2

Figure 1.2 – Cartoon model of ZIP1 molecules within the synaptonemal complex ……..……..22

Figure 2.1 – Meiotic divisions I and II as observed in bread wheat (*Triticum aestivum*) meiocytes using light microscopy …….....36

Figure 2.2 – Qualitative SDS-PAGE evaluation of proteins extracted from wheat anthers (T-IP) using three different extraction techniques (extraction protocol 1 – TCA/acetone extraction and precipitation; extraction protocol 2 – urea-mercaptoethanol-NP40 extraction with acetone precipitation; extraction protocol 3 – phenol extraction with methanol/ammonium acetate precipitation) ……..48

Figure 2.3 – Qualitative SDS-PAGE evaluation of total protein extracts from wheat anthers (T-IP) using the optimised protocol 1 ……49

Figure 2.4 – Suitability of protein extracts for use in 2DGE was assessed in pilot experiments using a 7 cm 2DGE gel format …….50

Figure 2.5 – Three different amounts of proteins were loaded onto the 11 cm gel format to determine the appropriate amount of protein required for adequate visualisation of the spots using the nanogram-sensitive CBB R250 staining technique used in this study ……..52

Figure 2.6 – Increasing total isoelectric focusing time resulted in increased spot resolution for IEF strips loaded with 100 or 150 µg of total protein ……53

Figure 2.7 – Horizontal resolution of the 11 cm 2DGE format was further optimised by utilising narrow range IEF strips …….55

Figure 2.8 – Optimising vertical resolution for 2DGE ……56

Figure 2.9 – Significant amounts of smearing were detected in DIGE gels with protein samples analysed in this study ……..58

Figure 2.10 – Sensitivity of protein spot visualisation was increased by using the EBT protocol post-CBB R250 staining ……59

Figure 2.11 – Spot numbers increased as meiosis progressed across the four pooled stages:- (A) PM-LP, (B) D-A, (C) TI-TII, and (D) T-IP ……...61

Figure 2.12 – Changing the tissue type analysed from whole anthers (A) to meiocytes (B) reduced the number of spots present in the 2DGE gels ……62

Figure 2.13 – 2DGE analyses between PM-LP and TI-TII protein extracts of Chinese Spring (A and B respectively) and the PM-LP extracts of the *ph* mutants (*ph1b* – C, *ph2a* – D respectively) revealed three regions with differing spot patterns ……..64

Figure 2.14 – Magnified view of the three regions with spot differences seen in Figure 2.13 (A – red square; B – blue rectangle, C – green rectangle) ……65

Figure 2.15 – Comparisons of normalised spot volumes of the six candidate spots identified ……...66

Figure 3.1 – Successful isolation of the full-length ORFs of *KK01*, *KK03*, *KK06* and a partial product for *KK04* ……106

Figure 3.2 – Isolation of *KK02* and *KK05* candidates were unsuccessful even when degenerate primers were used ……...106

Figure 3.3 – Location of peptides identified through MS/MS within the predicted amino acid sequences of candidate transcripts for KK01, KK03, KK04 and KK06 ……..107

Figure 3.4 – Conserved domain analyses of the candidate proteins ……..110

Figure 3.5 – Chromosomal locations of *KK01*, *KK03*, *KK04* and *KK06* candidate genes within the wheat genome ……..112

Figure 3.6 – Q-PCR analysis of the candidate genes ……115

Figure 3.7 – SDS-PAGE analysis of the KK01 protein product ……116

Figure 3.8 – SDS-PAGE analysis of the KK03 protein product ……117

Figure 3.9 – SDS-PAGE analysis of the KK06 protein product ……117

Figure 3.10 – Protein products for KK01 and KK06 interact with DNA *in vitro* ……..119

Kelvin H.P. Khoo, Amanda J. Able & Jason A. Able (2011) Poor homologous synapsis 1 (PHS1) interacts with chromatin but does not co-localise with ASYnapsis 1 (ASY1) during early meiosis in bread wheat. *BMC Plant Biology* (*submitted*).

Figure 1 - The PHS1 amino acid sequence is well-conserved across plant species ……162

Figure 2 – *TaPHS1* is located on chromosome group 7 of wheat …….163

Figure 3 – *Ta*PHS1 interacts with DNA *in vitro* ……164

Figure 4 – Q-PCR profiling of *TaPHS1* shows that it is expressed during meiosis ……165

Figure 5 – *Ta*PHS1 localisation during early meiosis in wild-type bread wheat ……...166

Figure 4.1 – Construction of the pGADT7-*TaPHS1* activation domain prey plasmid ……177

Figure 4.2 – Isolation and re-amplification of *TaRAD50* ……..179

Figure 4.3 – Conserved domain and phylogenetic analyses of *Ta*RAD50 ……180

Figure 4.4 – Excision of both *TaRAD50* inserts from their respective pGEM®T-Easy clones …….....181

Figure 5.1 – Isolation of the full-length coding sequence of *TaZYP1* ……...197

Figure 5.2 – *Ta*ZYP1 resides on chromosome group 2 ...……..197

Figure 5.3 – Phylogenetic analysis of *Ta*ZYP1 and its homologues …….198

Figure 5.4 – Conserved domain analysis of *Ta*ZYP1 ……198

Figure 5.5 – SDS-PAGE analysis of *Ta*ZYP1 ……...199

Figure 5.6 – Mass-peptide identification summary of *Ta*ZYP1 ……200

Figure 5.7 – *Ta*ZYP1 interacts with both ss- and dsDNA equally ……201

Figure 5.8 – Transcript expression analysis of *TaZYP1* ……202

Figure 5.9 – Q-PCR analysis of T_1 and T_2 *Taasy1* mutant lines ……204

Figure 5.10 – Immunolocalisation results using the polyclonal antibody raised against the full-length *Ta*ZYP1 antigen ……..205

Figure 5.11 –*Ta*ZYP1 localisation during sub-stages of prophase I in both wild-type bread wheat as well as the *ph1b* mutant ……..209

Figure 5.12 - Localisation of *Ta*ZYP1 and *Ta*ASY1 across five *Taasy1* RNAi knock-down mutants ……...212

Figure 6.1 – A model for *Ta*PHS1 homology searching in bread wheat …233-234

List of tables

Table 2.1 – Approximate physicochemical properties of six candidate spots identified via 2DGE analyses ……63

Table 3.1 – Comprehensive list of primers used in this study to isolate and characterise candidates identified to be differentially expressed via 2DGE experiments in Chapter 2 ……87

Table 3.2 – Identification of peptides obtained from 2DGE protein spots that were present at different levels isolated in Chapter 2 ……..103

Kelvin H.P. Khoo, Amanda J. Able & Jason A. Able (2011) Poor homologous synapsis 1 (PHS1) interacts with chromatin but does not co-localise with ASYnapsis 1 (ASY1) during early meiosis in bread wheat. *BMC Plant Biology* (*submitted*)

Table S1 - Detailed list of primers used in this study …….....168

Table 4.1 – Complete list of primers used in this study ...……..171

Table 4.2 – Summary of RAD50 sequence conservation 179

Table 5.1 – List of primers used to isolate and characterise *TaZYP1* …….....186

Abstract

During the early stages of meiosis, three key processes occur: chromosome pairing, synapsis and DNA recombination. Chromosomes are first replicated during interphase, after which they are aligned together in a non-random fashion to enable the installation of the synaptonemal complex (SC) along the chromosome axes leading to synapsis. Recombination machinery then enables strand invasion to occur, which then leads to the formation of chiasmata and ultimately, genetic recombination. Meiosis is further complicated in organisms with multiple genomes such as allohexaploid bread wheat (*Triticum aestivum* L.) which has three genomes (inherited from similar yet distinct progenitors), each with seven chromosomes. Thus a large number of proteins are likely to be required for the successful execution of this biological process.

The first approach in this study used proteomics to identify proteins that have possible roles during the early stages of wheat meiosis. Total protein samples isolated from staged meiocytes (specifically from pooled stages of pre-meiotic interphase to pachytene and from telophase I to telophase II) of wild-type Chinese Spring and the *Pairing homoeologous* deletion mutants, *ph1b* and *ph2a*, were analysed by 2-dimensional gel electrophoresis (2DGE). This resulted in identifying six differentially expressed protein spots (designated KK01 to KK06); from which three full-length coding sequences and one partial coding sequence of the candidate genes encoding these proteins were isolated (a putative speckle-type POZ protein, a pollen-specific SF21-like protein, a putative HSP70-like protein, as well as a partial hexose transporter peptide). Southern blot analysis revealed that these genes were spread across four different chromosome groups (2, 7, 5 and

1 respectively) with a copy on each of the three genomes (A, B and D). Q-PCR analysis of these four genes across the two pooled meiotic stages and various genotypes suggests that both *KK01* and *KK06* have roles during the early stages of meiosis and that they may be directly/indirectly regulated by a combination of elements within the *Ph1* and *Ph2* loci. The high level of *KK03* mRNA transcript detected in the later stages of meiosis is consistent with its role as a pollen-specific protein-encoding gene. In contrast, *KK04* expression suggests that it is post-transcriptionally regulated resulting in *KK04* being translated in the *ph2a* mutant. Both the speckle-type POZ protein and putative dnaK/HSP70 protein were also shown to interact with DNA *in vitro*.

The second approach of this study focused on isolating and characterising wheat homologues of two known meiotic proteins, namely *PHS1* and *ZYP1*. In the maize *PHS1* mutant *Zmphs1-0*, homologous chromosome pairing and synapsis are significantly affected, with homoeologous chromosome interactions occurring between multiple partners. More recently, co-immunolocalisation assays using anti-PHS1 and anti-RAD50 antibodies showed that both proteins had similar localisation patterns in the wild-type maize plants and that RAD50 localisation into the nucleus was affected by the absence of PHS1 thus implicating PHS1 as a regulator of RAD50 nuclear transport. In this study, the full-length coding transcript of wheat *PHS1* (*TaPHS1*) was isolated, sequenced and characterised. *TaPHS1* is located on chromosome group 7 with copies on the A, B and D genomes. Expression profiling of *TaPHS1* in both wild-type and the *ph1b* mutant during and post-meiosis show elevated levels of *TaPHS1* expression in the *ph1b* background. The *Ta*PHS1 protein has sequence similarity to other plant PHS1/PHS1-like proteins but also possesses a unique region of oligopeptide

repeat units. DNA-binding assays using both full-length and partial peptides of *Ta*PHS1 show conclusively that *Ta*PHS1 is able to interact with both single- and double-stranded DNA *in vitro*, even though no known conserved DNA-binding domain was identified within the *Ta*PHS1 sequence, indicating *Ta*PHS1 possesses a novel uncharacterised DNA-binding domain. Immunolocalisation data from assays conducted using an antibody raised against *Ta*PHS1 demonstrates that *Ta*PHS1 associates with chromatin during early meiosis, with the signal persisting beyond chromosome synapsis. Furthermore, *Ta*PHS1 does not appear to co-localise with the asynapsis protein – *Ta*ASY1 – possibly suggesting that these proteins are independently coordinated. Combined, these results provide new insight into the potential functions of PHS1 during early meiosis in bread wheat.

Similar to PHS1, Arabidopsis knock-down mutants of ZYP1 also display non-homologous chromosome interactions. ZYP1 has previously been characterised as a SC protein required for holding homologous chromosomes together in other species. In this study, the full-length coding sequence of the wheat *ZYP1* (*TaZYP1*) homologue was isolated, sequenced and characterised. Expression of *TaZYP1* analysed by Q-PCR across wild-type, *ph1b* and multiple *Taasy1* mutants during meiosis showed an approximate 1.3-fold increase in the *ph1b* mutant. In addition, DNA-binding assays demonstrate that *Ta*ZYP1 interacts with dsDNA under *in vitro* conditions while immunolocalisation (using an anti-*Ta*ZYP1 antibody) across wild-type, *ph1b* and *Taasy1* revealed the spatial and temporal localisation pattern of *Ta*ZYP1. Taken together, these results show that *Ta*ZYP1 plays an identical role to its homologues in other species as a SC protein and is affected by reduced levels of *Ta*ASY1 in wheat.

This body of work utilised a two-pronged approach to investigate meiosis in wheat with the overall outcome of identifying new meiotic proteins as well as characterising the wheat equivalents of two known meiotic proteins previously reported in other organisms. To this end, two previously uncharacterised wheat proteins with possible roles (involving interactions with chromatin) during meiosis have been successfully identified using the proteomics approach while both *Ta*PHS1 and *Ta*ZYP1 have been characterised with antibodies raised against both these proteins. The characterisation of *Ta*PHS1 and its DNA-binding capabilities, both *in vitro* and *in planta*, has shed light on a previously unknown function of the PHS1 protein while the localisation profile of *Ta*ZYP1 in *Taasy1* mutant lines has contributed to our understanding of how ASY1 levels can affect chromosome pairing in wheat.

Declaration

I declare that the work presented in this thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution to Kelvin Khoo Han Ping, and to the best of my knowledge and belief, this thesis does not contain any material previously written or published by another person, except where due reference is 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.

The author acknowledges that copyright of the published works contained within this thesis (as listed* below) resides with the copyright holder(s) of those works. I also give permission for the digital version of my thesis to be made available on the web, via the University 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.

* **Kelvin H.P. Khoo, Amanda J. Able & Jason A. Able** (2011) Poor Homologous Synapsis 1 (PHS1) interacts with chromatin but does not co-localise with ASYnapsis 1 (ASY1) during early meiosis in bread wheat. *BMC Plant Biology*, (*submitted*).

Kelvin Khoo Han Ping
May 2011

Acknowledgements

What a rollercoaster ride these past three and a half years have been. There are so many people I need to thank for helping me conquer this proverbial Everest.

First and foremost: Dr Jason Able – my principle supervisor. Thank you for your constant guidance and support throughout this project. Thank you for pushing me to be the best that I can be and for believing in my abilities. My achievements are a reflection of your commitment and dedication to this project. I am grateful for the passion for science that you have instilled in me and I hope to keep that the rest of my scientific career. I am sure that in the years to come, I will look back fondly on my years spent in the Able Meiosis Lab.

I would also like to acknowledge my co-supervisors and all of the other people that have guided me or provided a service: Dr Amanda Able, Dr Tim Chataway, Miss Margie Pallotta, Dr Neil Shirley, and Mr Jelle Lahnstein. Amanda, thank you for your supervision, guidance, and assistance throughout my project. Tim, thank you for your proteomics expertise and the use of your proteomics equipment. Margie, thank you for the use of your unbelievable nulli-tetra membranes. Neil, indeed you are the guru of all things Q-PCR. Thank you for your help and guidance in all things Q-PCR related. Jelle, thank you for the use of your proteomics equipment and for your useful experimental advice.

I would also like to acknowledge all of the members of the Able Lab over the years: Dr Scott Boden, Dr Wayne Crismani, Dr William Bovill, Dr Hayley Jolly, and Dr Damien Lightfoot. Thank you all for your guidance in experimental procedures, for taking the time to answer all my questions, for nuggets of wisdom and advice – in both science and life, and also for all the laughs we have had as a

group. It has been a pleasure working with all of you and I will always cherish these early years of my scientific career that I have spent with you all.

Thank you also to all the funding bodies that have made this research possible: The Australian Government, the University of Adelaide, the Farrer Memorial Trust, and the Grains Research and Development Corporation.

On a personal note, I would like to acknowledge my dear parents - Mum and Dad, for their ever-present love, support, guidance and prayers in my life. I thank you with utmost gratitude and sincerity for all the sacrifices that you have both made for my education, life and future. Though I know that I will never be able to repay my debt to you both, I hope that this achievement has made it worthwhile and that I have made you proud. To Alex and Adriana, thank you both for giving me a safe haven to rest in and for all the good times we have had together. To the whole Khoo family, I could never wish for a family better than ours - I love each and every one of you with all my heart.

I would also like to thank my close friends here in Adelaide that have made this path a little easier to walk. Indeed, it is not only in Liverpool that you never walk alone. Thank you to the whole Tan family, Hong Yau, Jingwen and Wan Chin for the bonds of friendship that we share. Last but not least, thank you Hong Pin, for being the loving, caring and patient person that you are. Thank you for turning good times into great times and bad ones into good ones. Thank you for always making me smile and for your unwavering support through these last few months.

Glossary of abbreviations

Abbreviation	Full term
2DGE	2-dimensional gel electrophoresis
3'	three prime
5'	five prime
9mer	9 base pair nucleotide
α-dCTP	alpha-deoxycytidine triphosphate
ºC	degrees Celsius
AFD1	*Absence of First Division 1*
Amp	ampicillin
At	*Arabidopsis thaliana*
ASY1	*ASYnapsis 1*
BCIP/NBT	5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium
BLAST	Basic Local Alignment and Search Tool
Bo	*Brassica oleracea*
bp	base pair
BSA	Bovine Serum Albumin
BTB	Bric-a-Brac, Tramtrack, Broad domain
BW26	Bob White 26 cultivar of bread wheat
CDK	*Cyclin Dependent Kinase*
cDNA	complimentary deoxyribonucleic acid
Ce	*Caenorhabdatis elegans*
CHAPS	3-[(3-Cholanidopropyl)Dimethylammonio]-1

CL	cell lysate
CT	cycle threshold
cv.	cultivar
D-A	diplotene to anaphase I pooled stage
Da	Dalton
DABCO	diazabicyclo-[2,2,2] octane
DAPI	4',6-diamidino-2-phenylindole
DIGE	2-dimensonal fluorescence difference gel electrophoresis
DMC1	*Disrupted Meiotic cDNA 1*
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
ds	double-stranded
DSB	double-stranded break
DTT	dithiothreitol
E	Expect value
EBT	eriochrome black T
EDTA	ethylene diamine tetra-acetic acid
ELP1	*Elongator Complex Protein 1*
EST	expressed sequence tag
FISH	fluorescent *in situ* hybridisation
FT	flow-through
g	gram
GAPDH	*GlycerAldehyde-3-Phosphate DeHydrogenase*
Ha	*Helianthus annuus*

His	histidine
hr	hour(s)
Hs	*Homo sapiens*
HSP70/70-2	Heat Shock Protein 70/70-2
Hv	*Hordeum vulgare*
HYP6	*Hypothetical 6*
IgG	immunoglobulin G
IEF	isoelectric focusing
IPTG	isopropyl-1-thio-P-D-galactoside
kb	kilobase
KCl	potassium chloride
kD	kilo Dalton
L	ladder/molecular weight marker
LB	Luria Bertani
µL	microlitre
µg	microgram
µM	micromolar
M	molar
MATH	Mephrin and TRAF homology domain
MALDI-TOF/TOF	Matrix-Assisted Laser Desorption Ionisation Time-of-Flight tandem mass-spectrometry
Mb	megabase
MCS	maleimidocaproyl-N-hydroxysuccinimide
MES	2-(N-morpholino)-ethane sulphonic acid
MFS	Major Facilitator Superfamily

mg	milligram
Mm	*Mus musculus*
mM	millimolar
min	minute(s)
MLH3	*Mut L Homologue 3*
MND1	*Meiotic Nuclear Divisions 1*
mRNA	messenger ribonucleic acid
MRE11	*Meiotic REcombination 11*
MRN	MND1-RAD50-NBS1 protein complex
MSH4/5	*MutS Homologue 4/5*
MS/MS	tandem mass spectrometry
MW	molecular weight
NaCl	sodium chloride
NBS1	*Nijmegen Break Syndrome 1*
NCBI	National Center of Biotechnology Information
ng	nanogram
Ni-NTA	nickel-nitrilotriacetic acid
nm	nanometre
NMR	nuclear magnetic resonance
NT	nullisomic-tetrasomic
ORF	open reading frame
Os	*Oryza sativa*
P	probability
PBS	phosphate buffered saline
PCR	polymerase chain reaction

Ph1/2	*Pairing homeoelogous 1/2*
PHS1	*Poor Homologous Synapsis 1*
pI	isoelectric potential
PM-LP	pre-meiotic interphase to pachytene pooled stage
PVP	polyvinyl pyrrolidone
PVPP	polyvinyl polypyrrolidone
Q-PCR	quantitative real-time PCR
r	correlation coefficient
R40	40 µg μL^{-1} RNAse in 1× TE
RAD50/51	*RADiation sensitive 50/51*
Rc	*Ricinus communis*
RNA	ribonucleic acid
RNAi	RNA interference
RNAse	ribonuclease
Rr	*Rattus rattus*
rpm	revolutions per minute
s	second(s)
Sb	*Sorghum bicolor*
Sc	*Saccharomyces cerevisiae*
SC	synaptonemal complex
SDS	sodium dodecyl sulphate
SDS-PAGE	SDS - polyacrylamide gel electrophoresis
SF21/21C1	*Sunflower 21/21 variant C1*
SMC	Structural Maintenance of Chromosomes domain
ss	single-stranded

SSC	standard saline citrate
SPO11	*SPOrulation-deficient 11*
TI-TII	telophase I to telophase II pooled stage
Ta	*Triticum aestivum*
Taq	*Thermus aquaticus*
TCA	trichloroacetic acid
T-DNA	transfer DNA
TE	Tris EDTA solution
T-IP	tetrad to immature pollen pooled stage
T_m	melting temperature
Tris	tris(hydroxymethyl)aminomethane
TUC	thiourea-urea-CHAPS
U	units
UV	ultra-violet
V	volts
Vhr	volt hours
Vv	*Vitus vinifera*
v/v	volume per volume
w/v	weight per volume
X-gal	5-bromo-4-chloro-3-indolyl-P-D-galactopyranoside
Y2H	yeast-2-hybrid
Zm	*Zea mays*
ZIP1/ZYP1/ZEP1	*Molecular Zipper 1*

Chapter 1 – Literature review

1.1 Meiosis and meiosis in bread wheat

1.1.1 – A brief overview of meiosis

Meiosis is a specialised cellular reduction-division process that is unique to sexually-reproducing organisms. It involves a single round of DNA replication followed by two consecutive rounds of chromosome segregation and cellular division, resulting in the generation of four haploid daughter cells from a single diploid cell. This process is important for the production of haploid gametes in sexually-reproducing species as it allows the correct amount of genetic material to be present in the diploid offspring. In addition, meiosis helps to generate genetic variation through the formation of new allelic combinations within offspring via homologous recombination and the independent assortment of chromosomes during during anaphase I (Pawlowski and Cande 2005).

The process of meiosis can be broken into two broad stages with the first round of chromosome segregation and cellular division termed meiosis I, and the second termed meiosis II. In both meiosis I and II, the meiotic cell progresses through five main stages – prophase, metaphase, anaphase, telophase and cytokinesis. Prophase I can be further broken down into five sub-stages – leptotene, zygotene, pachytene, diplotene and diakinesis (Figure 1.1 A-E). During these sub-stages of prophase I, an array of complex and vital cellular events such as homologous chromosome pairing, synapsis of homologous chromosomes, homologous recombination of DNA, and crossing-over occur (Hamant *et al.*

2006). Given that these events occur in prophase I, its study is important in the quest to thoroughly understand meiosis.

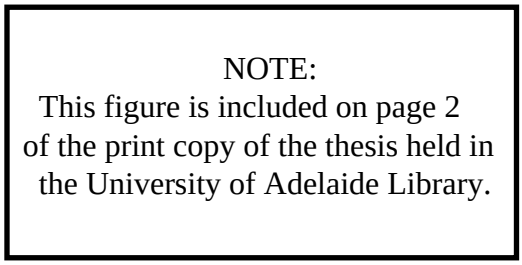


Figure 1.1 – The five sub-stages of prophase I followed by the major stages of meiosis in regal lily (*Lilium regale*). The five substages of prophase I (A – E): A) leptotene; B) zygotene; C) pachytene; D) diplotene; and E) diakinesis. This is followed by early metaphase I (F), late metaphase I (G), anaphase I (H), telophase I (I), interphase II (J); prophase II (K); metaphase II (L); anaphase II (M); telophase II (N); tetrads (O); and immature pollen (P). This figure has been adapted from McLeish and Snoad, Looking at chromosomes, 1958, Macmillan and Company Limited (Reproduced with permission of Palgrave Macmillan. This material may not be copied or reproduced without permission from Palgrave Macmillan).

At pre-meiosis, the centromeres and telomeres cluster and attach to the nuclear envelope. During leptotene (the first sub-stage of prophase I), chromosome condensation is initiated and followed by the addition of

proteinaceous structures, known as axial elements, along the chromosomes once the condensation process is complete (Zickler and Kleckner 1999). Structural changes of the chromatin between leptotene and zygotene are thought to initiate the formation of double strand breaks (DSBs) within the DNA that are required for homologous recombination to occur. Homologous chromosome pairing is initiated during zygotene when many essential proteins such as RAD51 (RADiation sensitivity protein 51), ASY1 (ASYnapsis 1) and ZYP1 (Arabidopsis synaptonemal complex protein 1) are recruited to the chromosomes (Armstrong *et al.* 2002, Higgins *et al.* 2005). The addition of the central element to the axial elements during the transition period into pachytene completes the formation of the proteinaceous synaptonemal complex (SC), thus allowing homologous chromosome pairing and synapsis to occur (Heyting 1996, Heyting 2005, Moses 1969, von Wettstein 1984). Completion of the SC and synapsis of the homologues in pachytene is then followed by diplotene where the SC disassembles causing the homologue juxtaposition effect to stop. However, the homologues stay connected as bivalents until metaphase I, joined together by the chiasmata that result from the crossing-over of the non-sister chromatid DNA strands (Page and Hawley 2004). Chromosome condensation and thickening continues into diakinesis, the last sub-stage of prophase I. Upon completion of condensation, the chromosomes are detached from the nuclear envelope where they were initially tethered by the telomere bouquet. Resolution of the DNA double Holliday Junctions (dHJs) formed during the recombination process is thought to occur during late metaphase I or the very early part of anaphase I prior to separation of the individual chromosomes of each chromosome pair to separate poles of the cell (Page and Hawley 2004, Petronczki *et al.* 2003). Completion of the first meiotic

division is followed by the second meiotic division, in which sister chromatids of each chromosome are separated to give rise to the four haploid daughter cells (Figure 1.1 O). Normal diploid chromosome numbers are restored upon fertilisation of the female gamete by a male gamete.

Meiosis has been extensively studied in model diploid organisms such as budding yeast, *Saccharomyces cerevisiae* (reviewed by Zickler and Kleckner 1998; and references therein), and Arabidopsis (*Arabidopsis thaliana*) (Armstrong *et al.* 2003, Armstrong and Jones 2003, reviewed by Hamant *et al.* 2006, and references therein). Such research has led to the discovery of many complex biochemical, molecular and cytological processes that occur during diploid meiosis. This collective data has been used to piece together the various pathways that lead to cell commitment into meiosis as well as those that occur during meiosis to form a chronological model (reviewed by Hamant *et al.* 2006; and references therein). However, research on the meiotic processes of more complex polyploid species has yet to reach parity with that of diploid meiosis, leading to a subsequent gap in the knowledge-base of such organisms.

Although meiosis is relatively well-conserved in sexually-reproducing species; given that polyploid organisms contain two or more genomes per cell, a higher level of complexity during this already complex cellular process is likely (Moore 2002). With many cropping commodities, and up to 70% of all flowering plants being polyploids (Bowers *et al.* 2003, Masterson 1994), the need to understand meiosis in complex organisms such as bread wheat should therefore not be underestimated.

1.1.2 – Meiosis in bread wheat

Bread wheat (*Triticum aestivum*) is one of the most commercially-important polyploid crops. The polyploidisation of bread wheat resulted through the fusion of three distinct, yet similar progenitor species: *Triticum urartu* (Riley 1975), *Aegilops speltoides* (Riley and Chapman 1958) and *Aegilops tauschii* (Riley 1975); each donating the A, B, and D genomes respectively. However, some conjecture still remains as to whether *A. speltoides* is the B genome progenitor. Each progenitor genome contributes seven chromosome pairs to the genome complement of bread wheat giving a total of 21 chromosome pairs (42 chromosomes) (Petersen *et al.* 2006).

Even though the A, B and D genomes within bread wheat originate from different progenitor species, the level of both gene content and gene order are highly conserved due to their relatedness. Inevitably, regions of varying genetic material do exist between the three genomes and normally occur in non-coding regions that consist largely of repetitive DNA sequences. Even so, the chromosomes of all three genomes genetically correspond with each other and are referred to as homoeologous chromosomes. This differs from homologous chromosomes which are chromosomes that share the same linear genetic sequence, gene order and repetitive DNA. In bread wheat, chromosomes 1A, 1B and 1D are homoeologues, while 1A and 1A are homologues.

Although bread wheat is an allohexaploid, it displays diploid behaviour during meiosis, whereby only homologous chromosome interactions are maintained. This occurs despite the high level of conservation of both gene order and content between corresponding chromosomes of the same groups but different genomes (e.g. between 1A, 1B, and 1D or between 4A, 4B, and 4D).

Aragón-Alcaide and colleagues (1997) previously reported the presence of chromosomal associations occurring between homoeologous chromosomes and even non-homologous chromosomes in pre-meiotic stages, albeit generally within the centromeric regions of the interacting chromosomes. However, these associations did not persist beyond pre-meiotic interphase suggesting the presence of cellular surveillance mechanisms that detect and actively prevent non-homologous chromosomal interactions from persisting beyond pre-meiotic interphase. Homoeologous chromosome pairing is detrimental as it results in multivalent associations and improper chromosome segregation at anaphase I. Consequently, it causes the production of non-viable gametes with unequal, and hence incorrect, haploid chromosome numbers. While research to uncover and understand the mechanisms that control homology searching/recognition between chromosomes as well as chromosome pairing within polyploids is occurring, many questions still remain unanswered.

Although researchers studying chromosome pairing in various diploid organisms can use single-copy probes targeted at specific sites of a chromosome pair to study the pairing behaviour of homologous chromosomes (Ding *et al.* 2004, Fung *et al.* 1998, Kerzendorfer *et al.* 2006, Scherthan *et al.* 1996, Weiner and Kleckner 1994), this technique is not suitable for use in polyploid organisms such as bread wheat as polyploidisation has led to the presence of multiple copies of each gene. In the simplest case, at least two other alleles of a given gene are located on each of the two homoeologous chromosome pairs in addition to the two alleles found on the homologous chromosome pair, equating to six possible sites being visualised using this method (Martínez-Pérez *et al.* 1999); thus making interpretation difficult with regards to whether homologous or homoeologous

chromosomes are involved. Further complications arise when the large amount of genetic duplication that has occurred in the bread wheat genome is taken into account. Previous studies have shown large regions of gene duplication on both homoeologous and non-homologous chromosomes which could further increase the number of sites that are possibly visualised using a single-copy site-specific probe (Foote *et al.* 1997).

1.2 – Meiotic processes

Due to the complexity of meiosis, many cytogenetic processes have to occur before and during the process of meiosis to ensure the formation of normal viable gametes. The most important processes are pre-meiotic chromosome interactions; telomere bouquet formation; homologous chromosome alignment; formation of the proteinaceous SC; and recombination of DNA, synapsis and pairing. Needless to say, a plethora of proteins are required to function either in protein complexes or as individual molecules to ensure that these key processes are initiated and completed.

1.2.1 – Pre-meiotic chromosome interactions

While observations of pre-meiotic chromosome interactions have been previously reported in some species, it does not appear to occur in the pre-meiotic interphase stage of all sexually-reproducing organisms. Some species that do display pre-meiotic chromosome interactions include allohexaploid wheat and other members of the *Triticeae* family (Abranches *et al.* 1998, Martínez-Pérez *et al.* 1999, Schwarzacher 1997), budding yeast (*S. cerevisiae*) (Jin *et al.* 1998) and fission

yeast (*Schizosaccharomyces pombe*) (Chikashige *et al.* 2006). In these organisms, the chromosome centromeres congregate at one pole of the cell while the telomeres are spread out at the opposite pole. This pre-meiotic chromosome arrangement based on the interphase cell polarity is known as the Rabl configuration. In contrast, humans (*Homo sapiens*) (Scherthan *et al.* 1998) and maize (*Zea mays*) (Dong and Jiang 1998) are among the many organisms that display neither the pre-meiotic chromosome interactions nor the Rabl chromosome arrangement.

While budding and fission yeast share a similar pre-meiotic chromosome arrangement to that seen in bread wheat, a further level of complexity is present during the pre-meiotic association in wheat that is not normally seen in other species. During the initiation of telomere clustering at one pole of the cell, the centromeres are seen to form seven distinct clusters at the other pole of the cell. Both these events occur simultaneously just prior to entry of the cell into prophase I (Martínez-Pérez *et al.* 2003). The seven centromere clusters are also observed in diploid wheat progenitors suggesting their formation is independent of both the presence of homologous chromosomes and the number of genomes present (Martínez-Pérez *et al.* 2003, Martínez-Pérez *et al.* 2000). One possible explanation for this elaborate centromeric cluster association is to bring homologues into close proximity of each other for efficient DNA homology base searching.

The seven homoeologous centromere clusters persist throughout the formation of the telomere bouquet (discussed further in section 1.2.2), after which the centromeres elongate and form complex tripartite structures during the final formation of the telomere cluster. These tripartite complexes are thought to allow

resolution of any non-homologous centromere associations to ensure that only homologous centromere associations persist (Martínez-Pérez *et al.* 2003). Upon complete formation of the telomere cluster, the seven homoeologous centromere groups dissociate and appear as 21 paired homologous centromere groups. In wheat, this signals the onset of prophase I, the condensation of the diffused chromatin, and further pairing interactions along the length of each homologous chromosome pair.

1.2.2 – Telomere bouquet formation

During pre-meiosis, chromosomes remain unpaired along their lengths with the exception of centromeric and telomeric regions. The telomeres attach to the nuclear envelope randomly, actively brought to one pole of the cell, and subsequently arranged to form a tight cluster termed the telomere bouquet (Bass *et al.* 1997, Scherthan *et al.* 1998). The timing of this clustering occurs during late pre-meiosis/early leptotene in hexaploid wheat (Martínez-Pérez *et al.* 2003) and Arabidopsis (Armstrong *et al.* 2001) but differs between species. Previous studies in maize report that telomeric clustering and bouquet formation occur during the leptotene-zygotene transition in maize (Carlton and Cande 2002, Harper *et al.* 2004), a much later stage compared to hexaploid wheat.

Telomere bouquet mutant analysis conducted in plants show that bouquet formation is not vital for the homologous chromosome pairing process to occur but seems to facilitate it (Hamant *et al.* 2006). By bringing the telomeres into close proximity with each other, the cluster may serve to keep homologous sequences in a small volume within the nucleus thus making the association of homologous telomeric and sub-telomeric regions easier. This is widely thought to

be part of the first step towards homologous chromosome recognition. Overall this results in more efficient pairing and synapsis of the homologues (Hamant *et al.* 2006, Martínez-Pérez *et al.* 1999). Indeed, maize mutant analysis conducted by Golubovskaya and colleagues (2002) shows that bouquet formation and homologous recombination are independent processes; but when coupled together lead to more efficient homologue pairing.

Telomere bouquet formation in fission yeast (*S. pombe*) has been proposed by Chikashige *et al.* (2006) to involve oscillation of chromosomes and the whole nucleus to bring the chromosomes together during what is known as the horsetail stage (Ding *et al.* 2004). While extensively studied in yeast, knowledge of the underlying mechanisms controlling telomere bouquet formation in cereals is relatively limited. One known difference between bouquet formation in fission yeast and plants is that cytoskeletal microtubules may play a different role.

Although Chikashige *et al.* (2006) hypothesise the use of cytoskeletal microtubules in the fission yeast model, no bouquet inhibition in the presence of microtubule depolymerising drugs such as amiprophos methyl (APM) was observed in rye (*Secale cereale*) (Cowan and Cande 2002). However colchicine, a known microtubule polymerisation inhibitor, was shown to inhibit bouquet formation even though the cytoskeletal microtubules were still present 15 hours post-colchicine treatment (Cowan and Cande 2002). This result suggests that in rye, and possibly plants in general, bouquet formation is not dependent on cytoplasmic microtubules and may occur via a different unknown cellular mechanism. This discrepancy serves to highlight that while meiotic processes in diploid organisms may be well-understood, the use of diploid meiosis models to

explain the same events in more complex organisms such as rye and wheat is not ideal.

1.2.3 – Homologous chromosome alignment

The formation of the homoeologous centromere clusters and their resolution into homologous centromere pairs, which occurs simultaneously with telomere bouquet formation, brings chromosomes within close proximity of one another. As previously discussed in section 1.2.2, this may lead to more efficient pairing of homologues by confining the chromosomes to a small volume within the nucleus thus making it relatively easier for the homologues to find and pair with each other. Upon entering into leptotene, rough alignment of homologous chromosomes is initiated in preparation for synapsis which occurs during zygotene (Schwarzacher 1997). The mechanism by which the rough alignment of homologous chromosomes occurs is not known though a number of mechanisms have been proposed. Wilson and colleagues (2005) suggested that it may occur by bringing specific DNA sequences located on homologues together at specific points, termed ‘rendezvous’ locations. The presence of highly-similar sequences such as telomeric and centromeric repeats, as well as highly-transcribed regions of DNA along both homologous chromosomes, is thought to facilitate the alignment process. Ribosomal RNA (rRNA) sequences present on certain chromosomes may also be utilised in this process (reviewed by Wilson *et al.* 2005; and references therein).

1.2.4 – Recombination, synapsis, and pairing

Upon completion of rough chromosome alignment, further processes are initiated to bring homologous chromosomes into closer contact with one another. Intimate physical connections are required for correct alignment of chromosomes during metaphase I and subsequent segregation of the homologues during anaphase I. Previous work has shown that the mechanisms used to achieve this are species-dependent. In organisms such as budding yeast (Peoples *et al.* 2002), Arabidopsis (Grelon *et al.* 2001), rye (Mikhailova *et al.* 2001) and maize (Ronceret *et al.* 2009); the stable pairing of homologous chromosomes during meiosis is dependent on the mechanisms that control recombination and/or synapsis. This is in contrast to organisms such as fission yeast (Nabeshima *et al.* 2001) and worm (*Caenorhabditis elegans*) (Dernburg *et al.* 1998) that use processes independent of recombination and synapsis.

1.2.4.1 – Recombination and its effects on chromosome pairing

Homologous recombination of DNA is initiated by the formation of programmed double-stranded breaks (DSBs) of DNA. Previous research has shown that at least ten gene products are required for this process, namely: SPO11, RAD50, MRE11, NBS1, MEI2, MER2, REC102, REC104, REC114 and SKI8 (Hamant *et al.* 2006, Keeney 2001, Nag *et al.* 2006).

Szostak and colleagues (Szostak *et al.* 1983) initially proposed the classical recombination double-strand break repair (DSBR) model based on their findings in yeast and animal cells. In this model, recombination is initiated by a DSB that normally occurs in the non-sister chromatids of bivalents during meiosis. The DSB is caused by SPO11 (SPOrulation 11), a type II topoisomerase with in-built

transesterase activity, with the assistance of other proteins (Klapholz *et al.* 1985). In the plant kingdom, SPO11 has been identified in several species; however, several homologues of SPO11 have typically been identified within any given species. Arabidopsis is one such example, where three homologues of SPO11 have been isolated. Even so, only one (*At*SPO11-1) has been shown to be involved in catalysing or causing the DSB formation in meiotic recombination to date (Grelon *et al.* 2001, Hartung and Puchta 2000, Hartung and Puchta 2001).

In Arabidopsis, the DSB is then resected from 5' to 3' by the *At*RAD50-*At*MRE11-*At*NBS1 (*Arabidopsis thaliana* RADiation sensitivity protein 50 - *Arabidopsis thaliana* Meiotic RECombination 11 - *Arabidopsis thaliana* Nijmegen Breakage Syndrome 1) protein complex forming 3' single-stranded overhangs on the ends of the DNAs (Bleuyard *et al.* 2004, Daoudal-Cotterell *et al.* 2002, Waterworth *et al.* 2007). Recombinase proteins such as RAD51 (RADiation-sensitive mutant 51) and DMC1 (Disrupted Meiotic cDNA 1), as well as other associated proteins are then recruited and loaded onto the single-stranded DNA (reviewed by Shinohara and Shinohara 2004; and references therein). Both RAD51 and DMC1 promote strand invasion, with DMC1 being vital for facilitating inter-homologue recombination (Schwacha and Kleckner 1997). A multitude of proteins interact to regulate this recombination process and help in the regulation of RAD51 and DMC1 activity (reviewed by Hamant *et al.* 2006; and references therein). In addition to its function as a recombinase; previous research conducted using nuclear magnetic resonance (NMR) imaging of the DNA-RAD51 nucleoprotein filament structure (Nishinaka *et al.* 1998), as well as *in vitro* DNA pairing assays, have shown that RAD51 is capable of homology searching and can promote homologous chromosome pairing over regions of

DNA several kilobases in length (Eggler *et al.* 2002).

Upon completion of DMC1/RAD51 loading onto the 3' single-stranded DNA, one strand then invades the intact homologous double-stranded DNA of the partnering chromatid causing the formation of the D-loop. DNA repair synthesis of the invading strand commences using the intact DNA strand as the template. The second broken strand of DNA from the DSB follows the same process resulting in the formation of the double Holliday junction (dHJ). This process of strand invasion and the formation of the dHJ has been postulated as the mechanism by which the chromosomes are brought into intimate juxtaposition with one another thus bringing the homologues close enough to facilitate the installation of the proteinaceous elements of the synaptonemal complex (SC) and complete synapsis (Chen *et al.* 2004, Wilson *et al.* 2005). Resolution of the dHJ occurs during late metaphase I/early anaphase I by the cutting of the DNA on alternative strands. Crossover or non-crossover of genetic material is thus dependant on which strands are cut.

Alternatively, Bishop and Zickler (2004) hypothesised that the choice of which DNA strands are cut could be pre-determined in the early stages of recombination. They hypothesised the Early Crossover Decision model (ECD) that suggests there may be two different pathways after the D-loop formation, with the decision of which pathway is to be taken made much earlier. The first pathway leads to formation of the dHJ followed by its resolution resulting in a crossover event while the second pathway involves the repaired invading DNA strand dissociating from the intact template strand, re-annealing to the other broken end, and DNA repair synthesis rejoining the two strands. This restores the chromosome leading to a non-crossover event (Bishop and Zickler 2004, Borner

et al. 2004, Haber *et al.* 2004).

In either model, a vast number of proteins are required to function individually as well as in protein complexes to ensure that the processes of homologous recombination, homologue juxtaposition, and homologue pairing are achieved correctly. While some of these proteins have already been identified, many yet remain to be identified. Of those that have been identified, the roles of many of these proteins and/or complexes are yet to be fully understood.

1.2.4.2 – Poor Homologous Synapsis 1 (PHS1) – a role in regulating recombination and homology searching?

In addition to the already complex array of proteins required for recombination, the recent discovery of the maize gene *PHS1* has added another level of complexity to the regulation of recombination events in plants. First discovered in a *Mutator* transposon-mutagenised maize population, analysis of the *phs1*-knockout mutant (*Zmphs1-0*) showed that *PHS1* had a meiosis-specific function with the mutant plants being sterile (Pawlowski *et al.* 2004). Based on sequence conservation searches of the databases, PHS1 appears to be a plant-specific gene with no reported homologues in any other non-plant species (Pawlowski *et al.* 2004). Detailed microscopic and fluorescent *in situ* hybridisation (FISH) analyses of the *Zmphs1-0* mutant meiocytes showed that the mutant had decreased levels of synapsis with improper alignment of the chromosomes in the synapsed regions observed. In addition to the mis-alignment, synapsis also occurs between multiple chromosome partners; highlighting the severity of the non-homologous chromosome interactions in the absence of the PHS1 protein (Pawlowski *et al.* 2004). Interestingly, the number of RAD51 recombination foci was also dramatically decreased in the *Zmphs1-0* mutant although the levels of RAD51

protein within the cells were equivalent to wild-type. This indicated that *ZmPHS1* regulated recombination during early meiosis and could possibly have a role in homology searching (Pawlowski *et al.* 2004).

The only other report on PHS1 details further characterisation in maize and Arabidopsis (Ronceret *et al.* 2009). Based on the FISH data obtained, Ronceret and colleagues (2009) concluded that PHS1 appeared to function in a similar manner regardless of genome size and complexity. Screening the maize and Arabidopsis *phs1* mutant meiocytes in dual 3-dimensional immuno-fluorescence localisation assays with antibodies against various meiotic proteins such as ASY1 and ZYP1 revealed that chromosome axis formation and installation of the SC components were similar to wild-type albeit with delayed ZYP1 loading observed in some instances (Ronceret *et al.* 2009). In contrast to ASY1 and ZYP1, the number of RAD50 foci within the nucleus was severely decreased in the *phs1* mutants while cytoplasmic accumulation of RAD50 was observed. Using the same technique, both PHS1 homologues were localised to the cytoplasm during the early stages of meiosis with peak cytoplasmic accumulation of the PHS1 protein seen during zygotene. In contrast to the Arabidopsis meiocytes where PHS1 was localised solely to the cytoplasm, observations of the maize meiocytes revealed that some foci clustered along the nuclear envelope during zygotene and a few foci were observed within the nuclei during pachytene (Ronceret *et al.* 2009). Combining the localisation data of PHS1 and RAD50, Ronceret and colleagues (2009) hypothesised that PHS1 regulates the transport of RAD50 into the nucleus (thereby indirectly regulating recombination and homology searching). Indeed this data explained the loss of RAD51 foci observed by Pawlowski and colleagues (Pawlowski *et al.* 2004), as the MRN complex (of

which RAD50 is a part) is required for the resection of the DSB before RAD51 can be recruited to the 3′ overhangs.

While these findings have added significantly to the current knowledge base, they have also raised some significant questions. Firstly, does PHS1 act as a direct shuttling protein for RAD50 and other, as yet unknown, meiotic proteins or does it regulate this process indirectly? What other proteins does it regulate? And does it have a direct role in homology searching? Seeing as there are subtle differences between the localisation profiles of the PHS1 homologues in both Arabidopsis and maize, do these differences translate into different functions for these proteins or suggest different regulation mechanisms in various plant species? Given that the absence of PHS1 leads to dramatically reduced levels of recombination, synapsis between multiple mis-aligned chromosomes, and non-homologous chromosome interactions, PHS1 is a prime candidate for further research, especially so in a complex organism such as bread wheat which has three highly similar, yet distinct genomes.

1.2.4.3 – Synaptonemal complex formation and synapsis

Synapsis is the formation of the tripartite SC along the entire length of the homologous chromosomes. Once the homologues have been juxtaposed, the process of synapsis is initiated. During leptotene, axial elements are attached to the non-sister chromatids (Zickler and Kleckner 1999). When pairing occurs in the later stages of SC formation, the axial elements come together and are then referred to as the lateral elements. The assembly of the central transverse filaments between the lateral elements completes the structure of the SC (reviewed by Page and Hawley 2004). Previous research has shown that the

process of synapsis starts from the telomeric and centromeric regions of the homologues first (Pawlowski and Cande 2005, Schwarzacher 1997, Zickler and Kleckner 1999).

The presence of recombination nodules (sites where recombination occurs) is thought to be linked with initiation of chromosome synapsis, thus linking the processes of synapsis and recombination (Pawlowski and Cande 2005). In yeast experiments, the recombinase enzyme, RAD51, localises within these nodules and has been shown to interact with ZIP3 (also known as CST9), a SC protein (Agarwal and Roeder 2000). ZIP3 itself has been shown to interact with other recombination proteins such as MSH4 (MutS Homologue 4) and MSH5 (MutS Homologue 5), which localise to late recombination nodules.

While both recombination and synapsis occur simultaneously during prophase I, they remain separate processes. However, recombination and synapsis could occur synchronously to stabilise chromosome associations during the process of recombination by holding the homologues in very close proximity. During pachytene, synapsis and recombination are completed. At this stage, the SC disassembles and the homologues are held together by the physical linking of the chiasmata that were formed during the recombination process (reviewed by Page and Hawley 2004, Zickler and Kleckner 1999). These chiasmata persist throughout metaphase I holding the homologues together as they align along the metaphase plate. Coupling the resolution of the chiasmata at late metaphase I/ early anaphase I with the removal of cohesion proteins such as SYN1 and AFD1 (by a yet unknown plant protease) from the chromatids, allows the homologous chromosomes to segregate and separate correctly to opposite poles of the cell. In asynaptic *zyp1* Arabidopsis mutants, homologous chromosome pairing and

bivalent formation are severely disrupted resulting in non-homologous chromosome associations, and in some cases, the formation of multivalents (interactions between more than two chromosomes) (Higgins *et al.* 2005). This serves to highlight the importance of the synaptonemal complex and the need for complete synapsis, due to significant downstream effects in the later stages of meiosis. However, even after decades of analysis, the role of the SC in pairing and recombination is still poorly understood and continues to be debated (reviewed by Hamant *et al.* 2006).

1.2.4.4 – The roles of ASY1 and ZYP1 in SC formation and synapsis

Two key proteins identified to have roles in synapsis and pairing of homologous chromosomes in plants are ASYnapsis 1 (ASY1) (Armstrong *et al.* 2002, Boden *et al.* 2007) and Molecular Zipper Protein *1* (ZYP1) (Higgins *et al.* 2005). Both have been shown to be directly or indirectly vital for synapsis and pairing to occur. ASY1 is also indirectly involved in homologous recombination as synapsis is an important event that leads to homologous recombination (Armstrong *et al.* 2002, Boden *et al.* 2007). In wheat, ASY1 has been shown to associate with axial elements of homologous chromosomes in addition to promoting homologous chromosome interactions (Boden *et al.* 2009). Analysis of *Taasy1* knock-down mutants have shown that lowered levels of *Ta*ASY1 leads to homoeologous chromosome interactions caused by the reduced promotion of homologous chromosome interactions (Boden *et al.* 2009). Gene expression analysis of *TaASY1* in the *pairing homoeologous 1* (*ph1*) mutant (*ph1b*) revealed that *TaASY1* expression is controlled by this locus and that absence of this locus results in a dramatic increase of the *TaASY1* transcript. These increased levels of *TaASY1*

lead to an excess of *Ta*ASY1 protein being produced, which subsequently leads to both homologous and non-homologous chromosome interactions being observed (Boden *et al.* 2009).

Unlike ASY1, which is associated with the axial elements of the SC, the yeast ZYP1 homologue (known as Molecular ZIPper 1, ZIP1) is a coiled-coil protein with globular N- and C- termini and forms the transverse filaments of the SC (Figure 1.2) (Sym *et al.* 1993, Sym and Roeder 1995). Studies of the yeast ZIP1 protein show that the C-terminus of ZIP1 interacts with chromatin, while the N-terminus stretches out into the central region of the SC and interacts with the N-termini of other ZIP1 molecules (thus forming the central element of the SC) (Figure 1.2) (Dong and Roeder 2000). While ZIP has been extensively studied in yeast, the ZIP1 homologues of plants have only been recently discovered.

The Arabidopsis *ZYP1a* and *ZYPb* genes were the first plant homologues of ZIP1 discovered in addition to being the first two SC proteins isolated from plants (Higgins *et al.* 2005). The discovery of these genes has since paved the way for the isolation and/or characterisation of other ZYP1 homologues in both *Secale cereale* (*Sc*ZYP1) (Mikhailova *et al.* 2006) and *Oryza sativa* (*Os*ZEP1) (Wang *et al.* 2010). In the analysis of the *Atzyp1* mutants by Higgins and colleagues (2005), they observed non-homologous chromosome pairing interactions that resulted in non-homologous recombination of DNA. This led them to suggest that ZYP1, and possibly the SC as a whole, may act to ensure that homologous recombination proceeds correctly in plants. This role is in addition to the other roles, including physically holding homologous chromosomes together to stabilise the homologues during the recombination process.

Localisation of the ZYP1 homologues of both Arabidopsis and rye have shown that ASY1 and ZYP1 do not co-localise but instead form a tripartite signal with ZYP1 being sandwiched by the ASY1 signal on either side. Such results confirm that these ZYP1 homologues share the same function as their yeast homologue even though the plant ZYP1 homologues share very little sequence similarities with their fungal homologue (Higgins *et al.* 2005, Mikhailova *et al.* 2006). Although the plant ZYP1 homologues appear to have the same functions during the early stages of meiosis, differences in the localisation profiles of the ZYP1 homologues of various species have previously been reported. One example of this is the formation of ZYP1 foci seen in the leptotene stage of both Arabidopsis (Higgins *et al.* 2005) and rice (Wang *et al.* 2010), while ZYP1 signal appears as linear tracts in rye meiocytes of the equivalent stage (Mikhailova *et al.* 2006). In a recent study of the rice ZEP1 homologue, Wang and colleagues found that unlike any previously reported ZYP1 homologues, the rice ZEP1 protein is re-loaded onto chromosomes two more times post-prophase II, and that ZEP1 is able to re-load onto the chromosomes normally even at the dyad and tetrad stages of pollen formation. This led Wang and colleagues (2010) to hypothesise that the ZEP1 protein may have a role in normalising chromosome behaviour after it has fulfilled its SC functions during the early stages of meiosis. Such differences in localisation profiles and functions highlight the need for species-specific studies to be conducted on the ZYP1 homologues. Considering that rice and wheat are closely related, the wheat ZYP1 homologue appears to be a prime candidate for characterisation using a gene-targeted approach as this protein may also reveal significant new data in terms of its function in a complex hexaploid species.

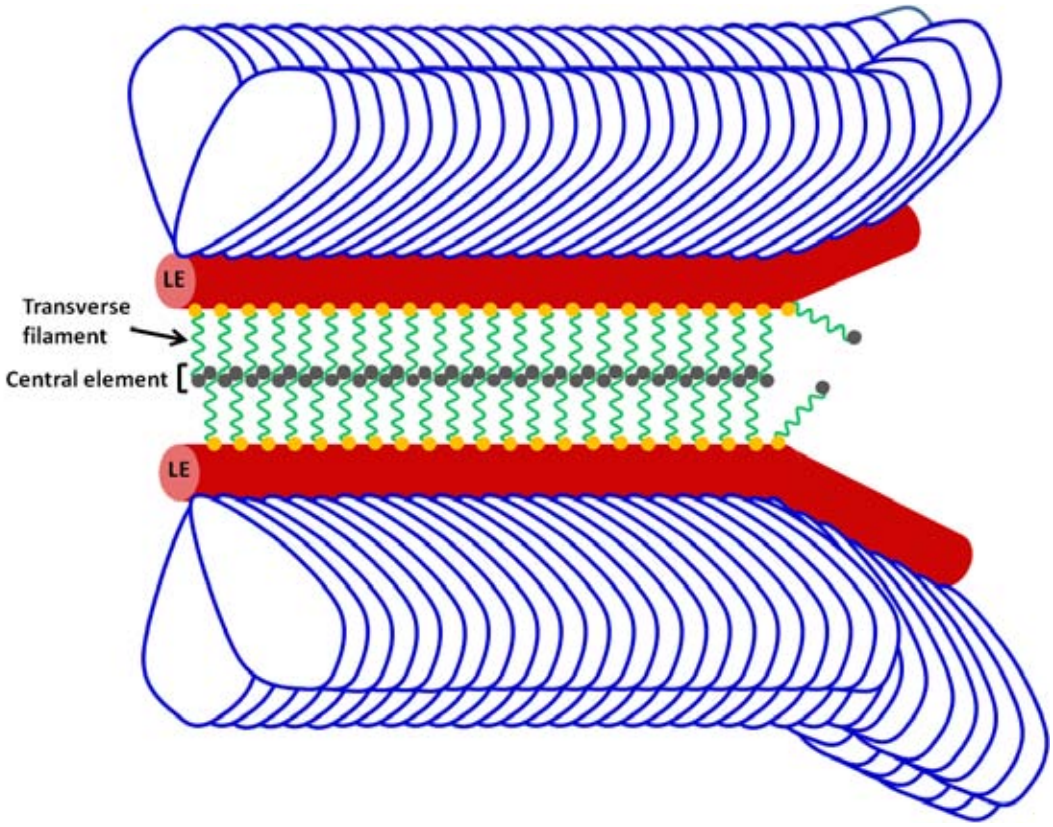


Figure 1.2 – Cartoon model of ZIP1 molecules within the synaptonemal complex. The ZIP1 protein molecule has a coiled-coil (green coils) central region capped at both ends by globular structures (C-terminus - yellow circles, N-terminus - grey circles). The C-termini of ZIP1 proteins interact with the lateral elements (LE, red rods) that are installed onto the chromatin (blue loops) while the N-termini interact with one another forming the central element of the SC. In doing so, the ZIP1 protein molecules form the transverse elements of the SC and hold the homologous sister chromosomes together.

1.2.5 – Homoeologous pairing of chromosomes in bread wheat

As previously discussed in section 1.1.2, bread wheat arose from hybridisation events between three independent progenitor species. While the early assumption that the three progenitor species were strongly differentiated was used to explain why hexaploid bread wheat displays diploid behaviour during meiosis, it was proven wrong when nullisomic-tetrasomic (compensation of a loss of one chromosome pair with another) experiments showed otherwise (Sears 1952, Sears 1976). This established that the three genomes were actually very closely-related

in both gene content and order. It also raised new questions as to how bread wheat chromosomes pair in a diploid-like manner. This resulted in the hypothesis that cellular mechanisms must exist to regulate and control chromosome pairing in hexaploid bread wheat in an effort to actively prevent non-homologous chromosome pairing interactions that result in multivalent associations, improper chromosome segregation at anaphase I, and thus production of non-viable gametes with incorrect haploid chromosome numbers (Riley and Chapman 1958).

The genetic control of chromosome pairing in bread wheat has previously been shown to be dependent on a series of promoting and suppressing *p*airing *h*omoeologous (*Ph*) loci. The two main suppressors of homoeologous pairing are the *Ph1* and *Ph2* loci. The *Ph1* locus was first identified and located on the long arm of chromosome 5B (Riley and Chapman 1958, Sears and Okamoto 1958) while a second locus, termed *Ph2*, was identified and located on the short arm of chromosome 3D (Mello-Sampayo 1968, Mello-Sampayo 1971, Mello-Sampayo and Canas 1973). Of the two main suppressor loci, the *Ph1* locus displays the strongest regulation of homologous chromosome pairing (Moore 2002, Sears 1976). An X-ray irradiated deletion mutant of *Ph1*, designated *ph1b* (originally identified by Sears (1977)), has increased homoeologous pairing at metaphase I. Mapping of the deleted region has shown a loss of approximately 70 Mb, predicted to possibly account for the loss of 200 or more genes (Gill *et al.* 1993). However, recent work by Griffiths and colleagues (2006) has significantly reduced the size of the *Ph1* locus to a region spanning only 2.5 Mb. A sub-telomeric insertion from chromosome 3A as well as seven *Cyclin-dependent kinase* (*Cdk*)-like genes that share sequence similarities to the *Cdk2* genes of both human and mouse (Griffiths *et al.* 2006, Martinez-Perez and Moore 2008) reside

within this refined *Ph1* locus (Griffiths *et al.* 2006). The presence of these *Cdk*-like genes within the *Ph1* locus is a significant finding as a previous study of the mouse *Cdk-2* deletion mutant has shown that homologous chromosome synapsis and recombination are severely disrupted in the absence of *Cdk-2* (Ortega *et al.* 2003, Viera *et al.* 2009). Furthermore, a study conducted by Al-Kaff and colleagues (2008) in wheat has shown that these *Cdk*-like genes within the *Ph1* region function to suppress the expression of their equivalents located on the chromosome 5A and 5D. In the absence of the *Ph1* locus, the *Cdk*-like equivalents of the A and D genome are expressed at levels that are significantly higher than normal (Al-Kaff *et al.* 2008). This led Al-Kaff and colleagues to suggest that a possible role of the *Ph1* locus is to coordinate the overall regulation of *Cdk*-like genes within the wheat genome. As previously discussed in section 1.2.4.4., one gene that appears to be directly regulated by the *Ph1* locus is *TaASY1*, where the absence of the *Ph1* locus results in a 20-fold increase in *TaASY1* transcript expression during early meiosis (Boden *et al.* 2009). This mis-regulated expression of *TaASY1* by the *Cdk*-like genes on the A and D genome may be the reason for the homoeologous chromosome pairing observed in the *ph1b* mutants.

Similar to the initial studies of the *Ph1* locus that focused on mutant analyses, two deletion mutants of *Ph2* have also been identified, namely *ph2a* (Sears 1977) and *ph2b* (Wall *et al.* 1971). The *ph2a* mutant was generated by X-ray irradiation (Sears 1977) with mapping of the deleted region showing a loss of approximately 80 Mb, predicted to account for the loss of at least 215 genes (Sutton *et al.* 2003). However, this figure could be as high as several thousand given the research reported in Sutton *et al.* (2003) which relied upon a comparative genetics approach with the completed rice genome sequence. The

ph2b mutant, which was generated by EMS (ethyl-methane sulfonate) mutagenesis, is thought to be either an INDEL (insertion/deletion) or a point mutation (substitution) (Wall *et al.* 1971). However, compared to the *ph1b* mutant, the *ph2* mutants exhibit only a moderate level of homoeologous chromosome pairing (Sears 1976).

While a significant amount of research has been conducted to uncover the mechanisms by which suppression of homoeologous pairing is achieved via the *Ph1* and *Ph2* loci, much is still left unknown; more so in the case of the *Ph2* locus. Analysis of the *ph1b* and *ph2b* wheat mutants by Martinez and colleagues (2001) showed that synapsis is affected in the *ph2b* mutant, whereas synapsis progresses normally in the *ph1b* mutant. Scoring multivalent associations and the timing at which these associations were corrected, revealed that only *Ph1* has a role in the correction of homoeologous associations. Taken together, these results suggest that the *Ph1* and *Ph2* loci work to promote homologous pairing and suppress homoeologous pairing via separate molecular pathways. More interestingly, these results raise the question as to whether the *Ph2* locus is a true pairing homoeologous locus or a synapsis protein that was initially mis-characterised (Martinez *et al.* 2001).

With the domestication of wheat over thousands of years, the selective pressures of plant breeding have led to a significant narrowing of the gene pool resulting in a loss of genetic diversity. The ability of the *ph* mutants to perform homoeologous pairing suggests that by understanding the molecular mechanisms of these loci, these loci may prove to be useful tools for increasing genetic diversity in current bread wheat lines. This could be achieved by the introgression of alien genetic material from closely-related plant relatives by either regulating

the *Ph* loci or by circumventing the effects of these loci. However, due to the presence of the seven *Cdk*-like genes within the *Ph1b* region, mutant analysis of the *Ph1b* region using RNAi mutants may prove unfeasible due to the relatively high levels of conservation between *Cdk* genes that could cause several of the (or all seven) *Cdk*-like genes to be silenced even if only one is targeted. One possible approach to overcoming this problem is to use proteomics to further study the *ph* mutants.

1.3 – Proteomics: A global approach to cell biology

Although DNA is viewed as the essential building block of life, the cell uses the translated gene products to perform all its required physiological and biochemical processes. Thus, while there is much to gain through studying the transcript levels of genes, the information gained from studies at the level of translated gene products may prove equally, if not more, important. Previous research on various species such as bacteria, yeast, and humans have shown the limitations of studies conducted solely at the transcript level, with many examples in the literature showing that levels of gene transcription are not indicative of the levels of the translated product (Bitton *et al.* 2008, Chen *et al.* 2002, Futcher *et al.* 1999, Ghaemmaghami *et al.* 2003, Gygi *et al.* 2000, Gygi *et al.* 1999a, Gygi *et al.* 1999b, Tian *et al.* 2004, Unwin *et al.* 2006, Waters *et al.* 2006). In the worst case scenario, little or no correlation could be made between transcription and translation profiles of some genes (Gygi *et al.* 1999a). However, this is not to say that a coupled proteomic and mRNA-based approach (e.g. microarray analysis) does not work. Previous research by Greenbaum and colleagues (Greenbaum *et al.* 2002) showed that in some cases, a coupled approach yields two sets of data

that complement each other, allowing correlations to be made, and is thus able to provide even more in-depth information than could have been gained by either approach independently.

Generally however, the poor correlation between transcript and protein levels in addition to the non-predictive nature of the two processes serves to highlight the need for analysis at the end point of the central dogma. While mRNA expression is important and cannot be dismissed, too little is currently known about the cell to accurately predict when an increase in mRNA species results in protein production and functional activity. Multiple explanations have been suggested as to why such significant differences exist between the transcript and protein levels. Of the more plausible reasons, it has been suggested that some genes may encode multiple protein isoforms due to differential splicing and translational efficiencies (reviewed by Peck 2005; and references therein). In addition, post-transcriptional regulation of mRNA and post-translational modifications of proteins that determine protein half-life may contribute to the large discrepancies often observed (Varshavsky 1996).

Due to the constant interaction of the cell with its surroundings as well as its own cellular processes, the proteome of a cell is in a constant state of dynamic flux. The study of cell biology via a proteomics approach could therefore assist in presenting a direct and realistic representation of the cellular environment at the specific time of analysis. To this end many different tools have been developed to study the proteome including 2-dimensional gel electrophoresis (2DGE) coupled with Matrix-Assisted Laser Desorption Ionisation Time-of-Flight tandem mass-spectrometry (MALDI-TOF/TOF). 2DGE allows separation and quantification of proteins while the MALDI-TOF-TOF allows identification of specific proteins.

1.3.1 – 2-dimensional gel electrophoresis (2DGE)

2DGE is the separation of proteins in two consecutive dimensions; firstly by their isoelectric potentials (pI) using an isoelectric focusing (IEF) procedure, and secondly by their molecular weights (MW) using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Since the first documentation of its use in 1975 (Klose 1975, O'Farrell 1975), the technology has significantly improved making it one of the most promising tools for proteomics (Fey and Larsen 2001, Görg *et al.* 2000, Harry *et al.* 2000).

Previous research has shown that 2DGE has the ability to resolve thousands of individual proteins from a given sample (Klose and Kobalz 1995), making it an efficient high-throughput method to study cellular protein profiles at a global scale. With the addition of sample pre-fractionation, specialised pH gradient isoelectric focusing (IEF) strips, and a new ultra-sensitive fluorescent dye sample labelling technique known as 2-dimensional fluorescence difference gel electrophoresis (DIGE), the resolving power of this already powerful proteomics tool has been further enhanced.

1.3.2 – A proteomics approach to plant meiosis

While proteomics has previously been used to study meiotic processes, most of the work done has focused on animal systems. An example of this is the use of 2D-DIGE coupled with MALDI-TOF/TOF to analyse the proteomic profiles of male rat cells at different stages of spermatogenesis (Rolland *et al.* 2007). The authors were able to identify 123 unique proteins that were differentially-

expressed in male rat germ cells using this approach. More recent work in cows dissected the proteomic profile of male spermatozoa using differential detergent fractionation multi-dimensional protein identification technology (DDF-Mud PIT) and identified 125 putative biomarkers for bovine fertility (Peddinti *et al.* 2008).

Only two previous reports attempting to dissect the proteomic profiles of plant meiocytes at different stages of meiosis have been reported. Work by Sánchez-Morán *et al.* (2005) used a coupled 2DGE-MALDI-TOF/TOF approach to dissect the stage-specific proteomic profiles of Chinese cabbage (*Brassica oleracea*) double-haploid plants during meiosis. The results showed that between 18% to 33% of the proteins identified through this method have putative roles in meiosis based on their sequence similarities with known meiotic proteins found in other species. Importantly, the authors showed that a proteomics approach to studying meiosis is a viable, albeit labour-intensive method with regards to the isolation of sufficient suitable meiotic material required for such analysis.

More recently, Phillips and colleagues (2008) utilised a combination of proteomics tools, which they collectively termed translation proteomics, to dissect meiosis in rye (*S. cereale*). Immunocytology using antibodies of known meiotic proteins and a 2DGE-Western blot showed that ASY1 and ZYP1 proteins probably had different functions when compared to their respective Arabidopsis homologues while there also appears to be multiple forms (active/inactive) present within the cell.

While a comprehensive chronological transcriptome profile of meiosis in allohexaploid bread wheat has already been constructed via microarray analysis (Crismani *et al.* 2006), there are no reported studies on the meiotic proteome of allohexaploid bread wheat. With only two previous studies on the plant meiotic

proteome published and the emerging awareness of the non-predictive nature of the transcriptome profile on the proteome profile, any study that aims to add to this limited knowledge-base is of significant importance.

1.4 – Rationale of the current study

While a significant volume of research has been conducted in model organisms such as yeast and Arabidopsis, much less is known about the proteins that are involved in the complex biochemical pathways that occur during meiosis in hexaploid bread wheat. With plant meiosis research, especially in more complex species lagging farther behind, this project has used two different approaches to study meiosis in hexaploid bread wheat.

The first component of this study involves using a broad-scale proteomics approach to identify and characterise proteins that are differentially expressed between stages of meiosis as well as between wild-type and the *ph* mutants. Through using such a strategy it is anticipated that novel meiotic proteins will be identified and characterised, and with the addition of the *ph* mutants, proteins that are differentially-expressed may also be detected.

The second component of this study involves a gene-targeted approach to isolate and characterise the wheat homologues of PHS1 and ZYP1. These candidates have been selected for characterisation to determine their function(s) and localisation profiles during early meiosis in order to further the current understanding of these proteins. In addition, the presence of ZYP1 loading in wild-type and *Taasy1* knock-down mutants will be investigated to determine whether reduced levels of *Ta*ASY1 affects ZYP1 function.

Broadly, the overall aim of this study is to increase the current knowledge base of bread wheat meiosis with a view that such results will eventually contribute to the development of molecular tools capable of introducing new genetic material into current wheat breeding lines. Such success will significantly accelerate plant breeding programs and allow for the generation of superior crop lines.

Chapter 2 – Identification of proteins with potential roles in meiosis via an optimised proteomics approach

2.1 – Introduction

Although meiosis research using yeast (*Saccharomyces* spp.) has been ongoing for decades, the focus of our understanding for this tightly coordinated process has only shifted to more complex organisms (such as plants) in the last 10 years. Consequently, the meiosis community has added significantly to our knowledge of specific gene function in plants such as *Arabidopsis thaliana*, *Zea mays* (maize) and *Triticum aestivum* (bread wheat). Examples include *SPOrulation 11-1* (*SPO11-1*) (Grelon *et al.* 2001), *ASYnapsis 1* (*ASY1*) (Boden *et al.* 2009, Boden *et al.* 2007, Caryl *et al.* 2000, Ross *et al.* 1997), *Molecular ZIPper 1* (*ZYP1*) (Higgins *et al.* 2005), *Disrupted Meiotic cDNA 1* (*DMC1*) (Klimyuk and Jones 1997) and the plant-specific *Poor Homologous Synapsis 1* (*PHS1*) (Pawlowski *et al.* 2004, Ronceret *et al.* 2009).

While many of these studies have relied on various classical genetic approaches, a possible alternative approach to study meiosis, especially in complex organisms, is proteomics. Proteomics has the potential to provide a more accurate profile of the cell as protein species present at a particular given time-point are studied rather than the mRNA transcripts. Recent studies performed by various investigators in multiple fields of research have shown that gene expression levels do not always correspond directly to the levels of their protein

products within the cell, further highlighting the importance of proteomic data in truly understanding cellular processes (reviewed by Waters *et al.* 2006).

Two dimensional gel electrophoresis (2DGE) is one of the many proteomics techniques available today. However, 2DGE has not been widely used to study meiosis in plants. Amongst the reported attempts are those by Sánchez-Morán *et al.* (2005) and Phillips *et al.* (2008). Using protein samples extracted from *Brassica* anthers, Sánchez-Morán and colleagues (2005) showed that 2DGE was a viable and powerful tool for studying meiosis. Work by Phillips and colleagues (2008) showed that translational proteomics based on 2DGE experiments can also be used to further investigate post-translational modifications to protein species isolated from plant meiotic tissue.

The objective of the work presented in this chapter was to optimise a 2DGE proteomics approach for investigating meiosis in bread wheat with the aim of identifying proteins that have roles in meiosis. Among the many parameters investigated and subsequently optimised were protein extraction techniques to find a suitable method for total protein extraction from meiotic bread wheat tissue, the amount of total protein required for adequate spot detection, various spot visualisation techniques as well as horizontal and vertical resolution parameters for successful 2DGE gels. The optimised method was then used to detect differences between protein profiles of four pooled stages of meiosis. Comparative analyses of the protein profiles for tissue from the earlier stages of meiosis in wild-type and *pairing homoeologous* mutants (*ph1b* and *ph2a*) were also performed to help elucidate the mechanisms by which homologous pairing is achieved.

2.2 Materials & Methods

2.2.1 – Staging and harvesting of wheat anthers and meiocytes

2.2.1.1 – Collection of meiotic tissue

Spikes were collected from Chinese Spring, *ph1b* and *ph2a* wheat plants grown in a controlled-environment room programmed with a 16 hr daylight 8 hr night photoperiod at approximately 23°C.

2.2.1.2 – Staging and harvesting of meiotic anthers and meiocytes

The anthers were staged and grouped into four pooled stages (pre-meiotic interphase to pachytene, PM-LP; diplotene to anaphase I, D-A; telophase I to telophase II, TI-TII; and tetrads to immature pollen, T-IP) prior to protein extraction (refer to Figure 2.1 for classification of stages used in this study). Florets and anthers were dissected under a Leica MZ 6 stereo dissecting microscope (Leica Microsystems, North Ryde, New South Wales, Australia). Using PrecisionGlide^{TM} syringes (Becton Dickinson Medical (S) Pty Ltd., Tuas Avenue, Singapore), the central floret was removed and the remaining spike tissue kept on ice. Primary and secondary anthers were removed from this floret, placed on clear glass microscope slides (Livingstone International Pty Ltd., Rosebury, New South Wales, Australia), stained for 5 min with 2% aceto-carmine (see Appendix A) at room temperature, and then cut in half longitudinally. Cover-slips (HD Scientific Supplies Pty Ltd., Sunshine, Victoria, Australia) were placed over the anther halves on the microscope slides and then gently squashed to release the meiocytes. These were observed using a compound microscope (Olympus MODEL CK40-RPSL, Richmond, South Australia, Australia) at 400× to 1000×

magnification. Anthers from alternating florets along the spike were staged this way with the last five florets at each end of the spike being discarded. For whole anther samples, the anthers from identical stages were pooled, snap-frozen and stored at -80°C until required for extraction. For meiocyte-enriched samples, anthers from identical stages were cut on one end and the meiocytes extruded into an ice-cold drop of phosphate buffered saline (PBS). The drop was then put into a 1.5 mL tube and treated the same as the pooled whole anther samples.

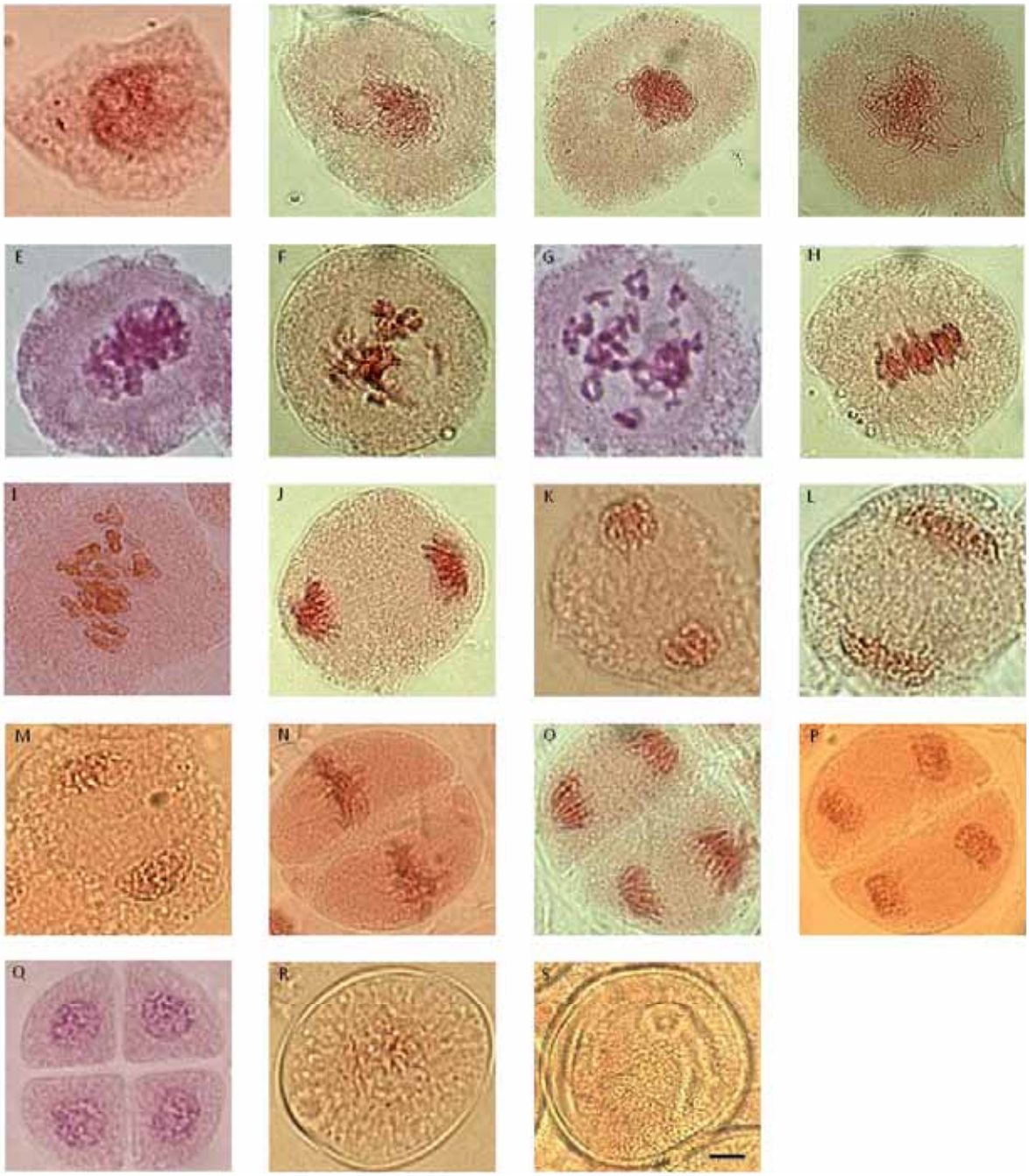


Figure 2.1 – Meiotic divisions I and II as observed in bread wheat (*Triticum aestivum*) meiocytes using light microscopy. Prophase I is represented by (A) pre-meiotic interphase, (B) leptotene-zygotene transition, (C) zygotene, (D) pachytene, (E) diplotene, (F) diplotene-diakinesis transition and (G) diakinesis. The remaining stages of meiosis I are represented by (H) metaphase I, (I) early anaphase I, (J) late anaphase I and (K) telophase I. The second meiotic division is represented by (L) interphase II, (M) prophase II, (N) metaphase II, (O) anaphase II, (P) telophase II and (Q) tetrad formation. (R) and (S) represent immature and mature pollen respectively. The four pooled stages analysed in this study were PM-LP (A-D), D-A (E-J), TI-TII (K-P), and T-IP (Q-R). Scale bar, 5 µm.

2.2.2. – Optimising the protein extraction method from whole wheat anthers and meiocytes

2.2.2.1 – Trichloroacetic acid (TCA)-acetone extraction and precipitation

This protocol was previously described by Sánchez-Morán *et al.* (2005) for extracting proteins from Chinese cabbage (*Brassica oleraceae*) anthers. Pooled whole anther samples (40 mg) were vortex-ground using metal ball-bearings and a MS 3 Digital vortex (IKA Works, Petaling Jaya, Selangor, Malaysia) in the presence of liquid nitrogen. The homogenate was resuspended by vortexing for 20 s in 2 mL of acetone containing 10% w/v trichloroacetic acid (Sigma) that was pre-chilled to -20°C. The sample was incubated overnight at -20°C then centrifuged at 13000 rpm at 4°C for 15 min. The supernatant was removed and the pellet was washed with 2 mL of acetone containing 20 mM DTT by incubating the pellet for 1 hr at -20°C. The pellet was then collected by centrifugation at 13000 rpm at 4°C for 15 min. The wash and centrifugation steps were repeated twice more with 30 min incubations. A final wash using 90% (v/v) acetone in nanopure water was performed to remove excess salts from the sample.

After the final centrifugation, the pellet was air-dried in a laminar-flow cabinet and resuspended with 200 µL of 7 M thiourea-2 M urea-2% (w/v) 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS) resuspension buffer (TUC buffer). The proteins were resuspended by 4 hr of slow vortexing at room temperature. Cellular debris was cleared from the resuspended solution by centrifugation at 9000 rpm for 15 min at room temperature. The supernatant was inserted into a fresh tube and the centrifugation step repeated to remove any residual debris. The resulting supernatant isolated was the cleared protein sample that was used for the 2DGE assay.

This protocol was subsequently further optimised by pre-chilling all tips, tubes and extraction solutions to -20°C prior to use. In addition, an additional wash step using 90% acetone (v/v) in nanopure water was added to reduce the salt concentration within the precipitated protein pellets. Two further centrifugation steps were also added post-resuspension of the samples in TUC buffer to help remove any residual tissue debris from the samples.

2.2.2.2 – Urea-mercaptoethanol-NP40 extraction with acetone precipitation

This protocol was previously described by Islam *et al.* (2003) for extracting proteins from endosperm tissue. Pooled whole anther samples (40 mg) were vortex-ground as previously described (Section 2.2.2.1). The homogenate was digested for 1 hr in 300 µL of lysis buffer [8 M urea, 2% (v/v) NP-40, 0.8% (v/v) ampholyte (Amersham Biosciences, Roosendaal, Netherlands), 5% (v/v) 2-mercaptoethanol and 5% (w/v) polyvinylpyrryolidone-40]. The resuspended sample was centrifuged at 9000 rpm for 10 min prior to a 2 hr incubation of the pellet in 80% (v/v) acetone at -20°C. The sample was centrifuged at 9000 rpm for 10 min and the pellets isolated and resuspended in 50 µL of lysis buffer. Residual cellular debris was removed from the resuspended sample by centrifugation at max speed for 10 min. The resultant supernatant was used for the 2DGE assay.

2.2.2.3 – Phenol extraction with methanol/ammonium acetate precipitation

This protocol was previously described by Carpentier *et al.* (2005) for extracting proteins from recalcitrant tissue. Pooled whole anther samples (40 mg) were vortex-ground as previously described (section 2.2.2.1). The homogenate was resuspended by vortexing for 30 s in 500 µL of ice-cold extraction buffer [50 mM

Tris-HCl pH8.5, 5 mM EDTA, 100 mM KCl, 1% (w/v) dithiothreitol (DTT), 30% (w/v) sucrose and one tablet of Roche CØmplete protease inhibitor cocktail (Roche Applied Science, Mannheim, Germany)]. Five hundred µL of ice-cold Tris-buffered phenol (pH 8.0) (Sigma) was added and the sample vortexed for 15 min at 4°C followed by a 3 min centrifugation step (9000 rpm, 4°C). The phenolic phase was collected and the extraction, vortexing and centrifugation steps repeated. The phenolic phase was collected and 2.5 mL of 100 mM ammonium acetate in methanol was added to precipitate the extracted proteins. The sample was incubated overnight at -20°C then centrifuged at 10000 rpm for 30 min at 4°C. The supernatant was removed and rinsed twice with 2.5 mL of ice-cold acetone/0.2% (w/v) DTT with 1hr incubations at -20°C between rinses. The pellet was air-dried in a laminar-flow cabinet then resuspended by vortexing for 1 hr at room temperature in 100 µL of lysis buffer [7 M urea, 2 M thiourea, 4% (w/v) CHAPS, 1% (w/v) DTT, 0.8% ampholyte (Amersham)].

2.2.3 – Protein Quantification

Protein was quantified using a method adapted from Bradford (1976). Triplicates of 1/10 dilutions of samples were made using nanopure water. Two hundred µL of Bradford reagent (BIO-RAD Laboratories Inc., Hercules, California, USA) was added to 800 µL of each sample and incubated for 7 min at room temperature before the absorbance was measured at 595 nm using a spectrophotometer (Metertech SP8001, Metertech Inc., Nankang, Taipei, Taiwan, Republic of China). To develop a standard curve, a Bovine Serum Albumin (BSA) dilution series was made with concentration values of 13.7 µg mL^{-1}, 10.275 µg mL^{-1}, 6.85 µg mL^{-1}, 3.425 µg mL^{-1} and 1.37 µg mL^{-1} using the BIO-RAD Protein Assay

Standard II (BIO-RAD), lyophilised BSA and nanopure water. Each dilution was done in triplicate to ensure that an average value could be obtained. Each standard was then spectrophotometrically quantified in the same manner as the samples. The values of the dilution series were used to generate a BSA linear regression curve from which the protein concentrations of the samples were predicted using Genstat (version 11.0, Numerical Algorithms Group).

2.2.4 – Visual assessment of protein quality via sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

2.2.4.1 – SDS-PAGE

Either 1 μg or 5 μg of each extract was added to 2.5 μL of NuPAGE® LDS Sample Buffer (4×) (Invitrogen, Carlsbad, California, USA) and 1 μL of NuPAGE® Reducing Agent (10×) (Invitrogen). When the protocol was further refined, 8 μg of extract was used instead. Samples were then made up to a total of 10 μL with nanopure water. The samples were heated at 70°C for 10 min before being loaded into 15-well NuPAGE® Novex® 4-12% Bis(2-hydroxyethyl)-imino-tris(hydroxymethyl)-methane (Bis-Tris) mini gels (Invitrogen). Bio-Rad Precision Plus Dual Colour Protein Standard (10 μL) was also loaded onto the gels to estimate protein size. Electrophoresis was performed using 2-(N-morpholino)-ethane sulphonic acid (MES) (Invitrogen) buffer according to the manufacturer's instructions to separate the protein species within the samples according to their molecular weights.

2.2.4.2 – Fixing, staining and storage of Bis-Tris gels

Upon completion of the SDS-PAGE in section 2.2.4.1, gels were removed from their cassettes and left to shake in 100 mL of fixing solution [10% (v/v) methanol, 10% (v/v) glacial acetic acid, 40% (v/v) ethanol, nanopure water] for 1 hr. The fixing solution was then removed and replaced with 100 mL sensitisation solution [10% (w/v) ammonium sulphate, 1% (v/v) glacial acetic acid, nanopure water]. The sensitisation solution was then replaced with 100 mL of staining solution [0.125% (w/v) Coomassie Brilliant Blue R250 (CBB R250), 5% (v/v) glacial acetic acid, 45% (v/v) ethanol, nanopure water] and left to stain overnight on a TASP-OS1 orbital shaker (Thermoline, Northgate, Queensland, Australia). De-staining of the gels was performed for 1 hr using de-stain solution I [5% (v/v) glacial acetic acid, 40% (v/v) ethanol, nanopure water], followed by de-stain solution II [3% (v/v) glacial acetic acid, 30% (v/v) ethanol, nanopure water] until the background of the gels were clear. Even though the protein bands could be seen clearly by eye under normal lighting conditions, a light-box was used to further illuminate the gels during analysis. Gel images were obtained by scanning using an Epson Perfection 4180 Photo scanner (Epson Australia Pty Ltd., Adelaide, South Australia, Australia). Analysed gels were stored in preservation solution [5% (v/v) glacial acetic acid, nanopure water] at 4°C.

2.2.5 – 2-Dimensional Gel Electrophoresis (2DGE)

2.2.5.1 – First dimension: Isoelectric focusing (IEF) of 2DGE samples

Carrier ampholytes (GE Healthcare), DTT and bromophenol blue were added to samples from section 2.2.3 to final concentrations of 1% (v/v), 65 mM and 0.02% (v/v) respectively. Each sample (at 75 µg, 100 µg or 150 µg of protein as

determined by Bradford assay as per Section 2.2.3) was made up to a final volume of 125 µL for 7 cm strips and 200 µL for 11cm strips using TUC buffer. The sample was loaded into a Ettan IPGphor strip holder (GE Healthcare Life Science, Uppsala, Sweden) of the appropriate length corresponding to the use of either a 7 cm or a 11 cm IPG Readystrip (BIO-RAD). During the optimisation of the 11 cm IPG Readystrip (BIO-RAD) IEF protocol, strips with pH 3-10, 3-6, 5-8 and 7-10 ranges were used. The IPG strip was lowered gel-side down into the strip holder and subsequently covered with DryStrip Cover Fluid (GE Healthcare). The strip holder cover was then attached and the holder set in place in the Ettan IPGphor (GE Healthcare). The IEF protocol for the 7 cm strips was performed at 20°C with a set maximum of 50 mA per strip. The strip was passively rehydrated for 6 hr then focused at 30 V for 8 hr (step-and-hold), 1000 V for 0.5 hr (gradient), 5000 V for 1 hr 20 min (gradient), 5000 V for 3000 Vhr (step-and-hold). During the optimisation of isoelectric focusing for the 11 cm strips, the same IEF protocol was repeated but with differing total Vhr used (35000 and 45000 Vhr) (Section 2.3.2.1). The optimised 11 cm IEF protocol was performed at 20°C with a set maximum of 50 mA per strip focused at 30 V for 13 hr (step-and-hold), 200 V for 1 hr (step-and-hold), 500 V for 1 hr (step-and-hold), 1000 V for 1 hr (step-and-hold), 8000 V for 30 min (gradient), 8000 V for 45000 Vhr (step-and-hold).

2.2.5.2 – Equilibration of IEF strips

Prior to performing the second dimension SDS-PAGE, the IEF strips were removed from the strip holders and dried gently to remove excess Cover Fluid. The strips were then equilibrated with equilibration buffer [8M urea, 50 mM Tris, 20% (v/v) glycerol and 2% (w/v) SDS; all components were mixed then adjusted

to a final pH of 8.8] containing 1% w/v DTT for 20 min on a platform mixer (Ratek Instruments Pty Ltd., Boronia, Victoria, Australia) before being rinsed with sterile nanopure water. The equilibration and rinsing process was repeated with equilibration buffer containing 2.5% (w/v) iodoacetamide.

2.2.5.3 – Second dimension: SDS-PAGE

The equilibrated IEF strips were slotted into BIO-RAD Criterion XT Bis-Tris 11 cm gels (4-12.5%, 10% or 12%) and sealed in place with 0.5% (w/v) agarose dissolved in MES buffer. The gels were electrophoresed in a Criterion 11 cm SDS-PAGE cell (BIO-RAD) under the following conditions and parameters: 23°C, 200V constant for 45 min in 1×MES buffer, starting at 200 mA and finishing at 110 mA per gel. The same was done for the 7 cm strips with NuPAGE® 4-12.5% ZOOM Bis-Tris gels (Invitrogen) with the exception that gel running parameters were as per the manufacturer's protocol.

2.2.6 – 2DGE gel staining

During the optimisation process, gels were initially fixed, stained and visualised using the CBB R250 protocol as per section 2.2.4.2. Upon completion of an optimised 2D protocol, optimised gels were stained using an eriochrome black T (EBT) silver staining procedure post-CBB R250 staining and de-staining. To stain each gel using the EBT technique, 100 mL of the following solutions were made fresh: sensitiser solution [0.006% (w/v) EBT, 30% (v/v) ethanol, silver solution [0.25% (w/v) silver nitrate, 0.037% (v/v) formaldehyde (added just before use)], developer solution [2% (w/v) potassium carbonate, 0.04% (w/v) sodium hydroxide, 0.002% (w/v) sodium thiosulphate, 0.007% (v/v) formaldehyde (added

just before use)], and stop solution [30% (v/v) ethanol, 10% (v/v) acetic acid]. To silver-stain the gels, de-stained CBB R250 gels were sensitised for 2 min with sensitiser solution, de-stained for 2 min with de-stain solution, washed twice for 2 min with deionised water, incubated in silver solution for 5 min, washed twice for 2 min with deionised water, and incubated in developer solution until the desired staining intensity was achieved. The developing reaction was quenched with stop solution. All steps were performed at room temperature on a TASP-OS1 orbital shaker (Thermoline).

Imaging of protein spots from CBB R250 gels and silver stain gels was done using either an Epson Perfection 4180 Photo scanner (as per section 2.2.4.2) or an ImageScanner (GE Healthcare) operated by the Lab-Scan 3.00 software (GE Healthcare). Intensity calibration of the Imagescanner was performed using an intensity stepwedge prior to gel image capture. Image analysis was performed using the ImageMaster 2D Elite 4.01 software (GE Healthcare). Gels were compared manually as well as using the automated spot detection mode. Once spot matching was completed, the background was subtracted using the 'average on boundary' method while spot normalisation was performed using the 'total spot volume' method. The normalised spot volume data for each spot was then used in a one-way analysis of variance (ANOVA) test in Genstat (version 11.0, Numerical Algorithms Group) to determine the least significant difference (LSD) ($p < 0.05$) between the mean volume for the spots in each treatment.

2.2.7 – 2-Dimensional Fluorescence Difference Gel Electrophoresis (DIGE)

2.2.7.1 – Sample pH adjustment

Due to the nature of the fluorescent dyes used in DIGE, the pH of the protein sample has to be adjusted prior to addition of the dyes to the sample. Protein sample (1 µL) was spotted onto a pH-Fix 4.5-10.0 indicator strip (Macherey-Nagel GmbH & Co.KG, Düren, Germany) to determine the pH of the sample. The sample pH was then adjusted to the pH 8.5 using 50 mM sodium hydroxide.

2.2.7.2 – Preparation of CyDyes and protein labelling

Prior to labelling, the CyDye 2-D Fluorescence Difference Gel Electrophoresis (DIGE) Fluor minimal dyes (GE Healthcare) were reconstituted then diluted according to the manufacturer's protocol using 99.8% anhydrous N,N-dimethylformamide (Sigma). Fifty µg of each protein sample (as determined by the Bradford assay as per section 2.2.3) was labelled with 400 pmoles of its respective CyDye (Cy3 or Cy5) and left to incubate for 30 min on ice in the dark. One µL of 10 mM L-lysine monohydrochloride (Sigma) was then added to each sample to stop the labelling reaction. The samples were incubated on ice for 10 min before they were snap-frozen with liquid nitrogen and kept in the -80°C freezer until required for isoelectric focusing. The internal standard protein sample, which is a mixture of all the protein samples required for the experiment, was labelled with the Cy2 dye and treated exactly like the other protein samples.

2.2.7.3 – IEF and SDS-PAGE of DIGE samples

The Cy2-labelled internal standard was mixed with the Cy2- and Cy3-labelled protein samples before carrier ampholytes (GE Healthcare), DTT and bromophenol blue were added to the sample to final concentrations of 1% (v/v), 65 mM and 0.02% (v/v) respectively. The mixed sample was made up to 200 µL using TUC buffer. The sample was applied to an 11 cm IPG Readystrip (pH5-8) (BIO-RAD) as per section 2.2.5.1 with the final IEF step modified to reach a total of 35000, 45000 or 55000 Vhrs. IEF strips were then equilibrated as per section 2.2.5.2 and the second dimension (SDS-PAGE) run as per section 2.2.5.3.

2.2.7.4 – Visualisation of DIGE protein spots

Gels were transported on ice to the Flinders Proteomics Centre where they were removed from their cassettes and placed on a low fluorescence glass plate before being scanned using a Typhoon 9400 Laser Scanner (GE Healthcare). Gel images were analysed using both DeCyder 2D (version 6.5; GE Healthcare) and ImageMaster 2D Elite 4.01 (GE Healthcare).

2.3 – Results

2.3.1 – Optimisation of a suitable protein extraction technique

In order to determine a suitable protocol for extracting total cellular proteins from wheat meiotic tissue, three different extraction protocols were investigated. Extraction protocol 1 was a TCA-acetone extraction and precipitation protocol while extraction protocol 2 utilised a urea-mercaptoethanol-NP40 extraction followed by acetone precipitation. Extraction protocol 3 consisted of a phenol

extraction step followed by methanol/ammonium acetate precipitation. Results of Bradford assays conducted on the protein extracts obtained indicated that protocol 2 (both soluble and insoluble PVP) yielded the most protein followed by protocol 1 and lastly protocol 3 (data not shown). However, SDS-PAGE analysis of the protein extracts revealed that the quality of the extracts were significantly different (Figure 2.2). Intact protein species (as evidenced by tight bands) were observed for extracts using protocol 1 and 3 with extracts using protocol 1 containing more bands over a broader range of molecular weights. Samples extracted using protocol 2 (soluble PVP) appeared smeary within the gel suggesting protein degradation had occurred. Although 1 or 5 µg of total protein (as quantified via the Bradford assay) were loaded into the gel lanes, no bands or smears were present in lanes loaded with extracts of protocol 2 (using insoluble PVP). This was most likely due to contaminating residual PVP debris giving false spectrophotometric readings of the samples in the Bradford assay. Based on this result, protocol 1 was chosen for further optimisation.

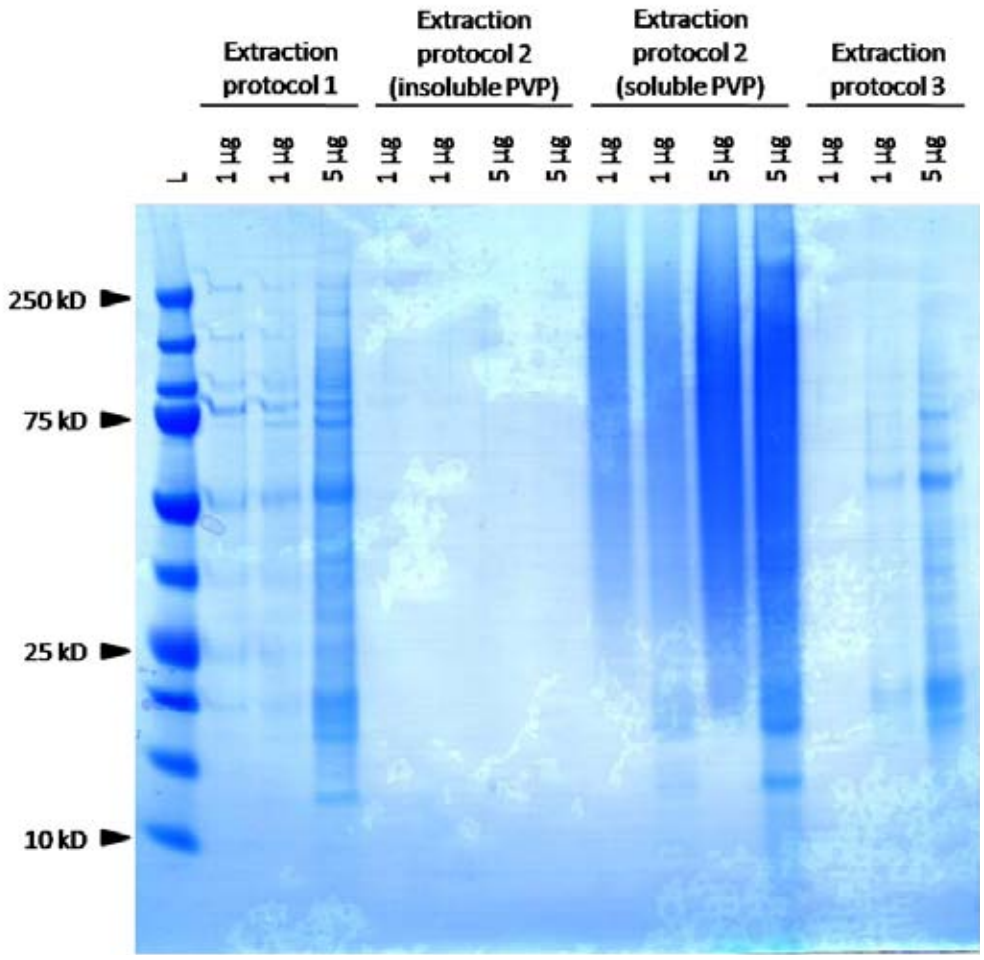


Figure 2.2 – Qualitative SDS-PAGE evaluation of proteins extracted from wheat anthers (T-IP) using three different extraction techniques (extraction protocol 1 – TCA/acetone extraction and precipitation; extraction protocol 2 – urea-mercaptoethanol-NP40 extraction with acetone precipitation; extraction protocol 3 – phenol extraction with methanol/ammonium acetate precipitation). Total protein extracts of three biological replicates from three technical replicates of extraction protocols 1 and 3 were loaded while four were loaded for protocol 2 (for both insoluble and soluble PVP respectively). The presence of defined protein bands in protocols 1 and 3 indicated that the two protocols worked better than protocol 2 which yielded no bands with insoluble PVP and degraded proteins with soluble PVP.

When the TCA-acetone extraction was further refined (section 2.2.2.1), far superior yields and much more defined protein bands in subsequent extractions were observed (Figure 2.3).

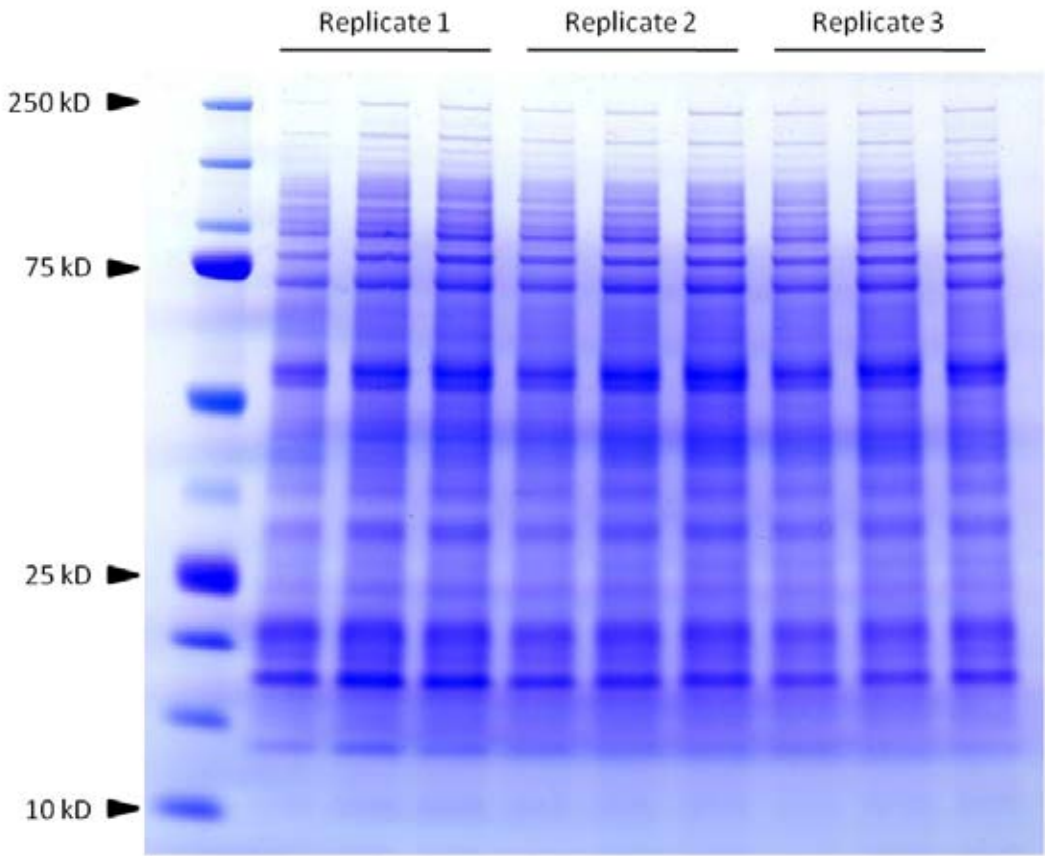


Figure 2.3 – Qualitative SDS-PAGE evaluation of total protein extracts from wheat anthers (T-IP) using the optimised protocol 1. Optimisation of extraction protocol 1 resulted in an increased number of protein species being extracted as evidenced by the increased number of bands (compared to Figure 2.2), higher protein yields, and minimal protein degradation. Total protein extracts of three biological replicates from three technical replicates are shown with 8 µg of total protein from each sample loaded.

Further assessment of the protein extracts involved pilot 2DGE experiments using 7 cm IEF strips (pH 3-10) to ensure samples obtained from the optimised extraction protocol 1 were suitable for use in 2DGE experiments (Figure 2.4).

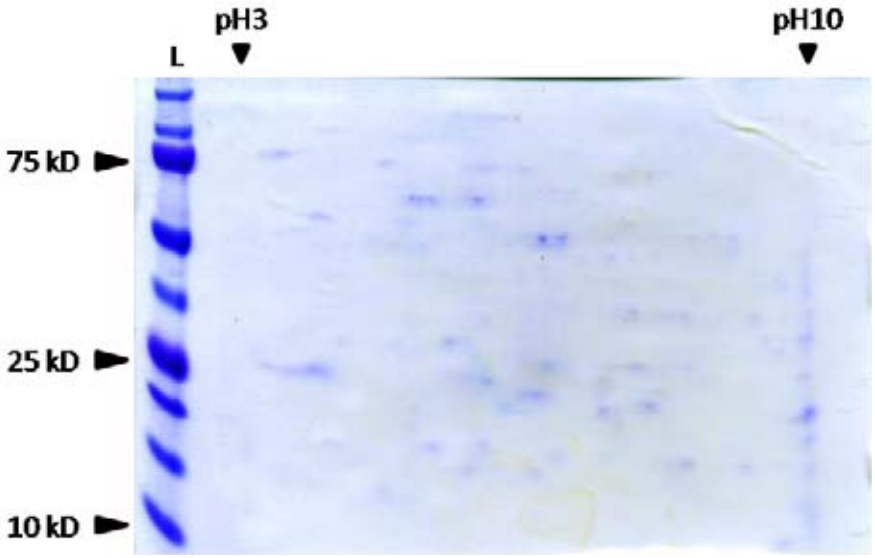


Figure 2.4 – Suitability of protein extracts for use in 2DGE was assessed in pilot experiments using a 7 cm 2DGE gel format. Fifty µg of total protein extracted from pooled T-IP wheat anthers was loaded onto an IEF strip (pH 3-10), focused in the first dimension for 8164 Vhr then separated on a 4-12% NuPAGE Bis-Tris gel. The presence of well-defined spots over a large molecular weight range indicated that the samples obtained from the optimised extraction protocol 1 were of suitable quality for 2DGE.

2.3.2 – Optimisation of 2DGE gel resolution

2.3.2.1 – Horizontal resolution (1st dimension of 2DGE)

In an effort to increase the horizontal resolution of the 2DGE gels obtained, an 11 cm IEF strip and gel format was adopted. In order to determine the appropriate amount of protein required for good separation and visualisation of protein spots in this gel format, different amounts of protein (50, 100 or 150 µg) were investigated (Figure 2.5). The results indicated that while 50 µg of protein (Figure 2.5 A) was insufficient; little difference was seen between gels loaded with 100 or 150 µg of protein (Figure 2.5 B and C). One difference observed was the over-representation of high abundance protein species in gels loaded with 150 µg of protein which led to larger protein spots that masked spots of low abundance proteins. One hundred µg of total protein was therefore deemed an appropriate amount for loading onto the 11 cm format. In addition, determining the

appropriate amount of protein required was important due to the large amount of time and effort required for staging and pooling the wheat anthers according to the four pooled stages as well as for extracting sufficient amounts of protein from the meiotic tissue.

Initial pilot experiments (refer to Figure 2.5) focused for the recommended 35000 Vhr resulted in some spots remaining diffuse and smeary, indicative of the samples being under-focussed. IEF parameters were then investigated to ensure that proper separation of the protein species within the 100 µg of protein extracts was achievable. This involved increasing the length of the IEF protocol to a total of 45000 Vhr. Comparisons between gels loaded with 100 or 150 µg of total protein focused for 35000 (Figure 2.6 A and C) and 45000 Vhr (Figure 2.6 B and D) revealed that at both protein loads, the prolonged IEF protocol resulted in superior protein spot separation and definition while also reducing background smears in the gels.

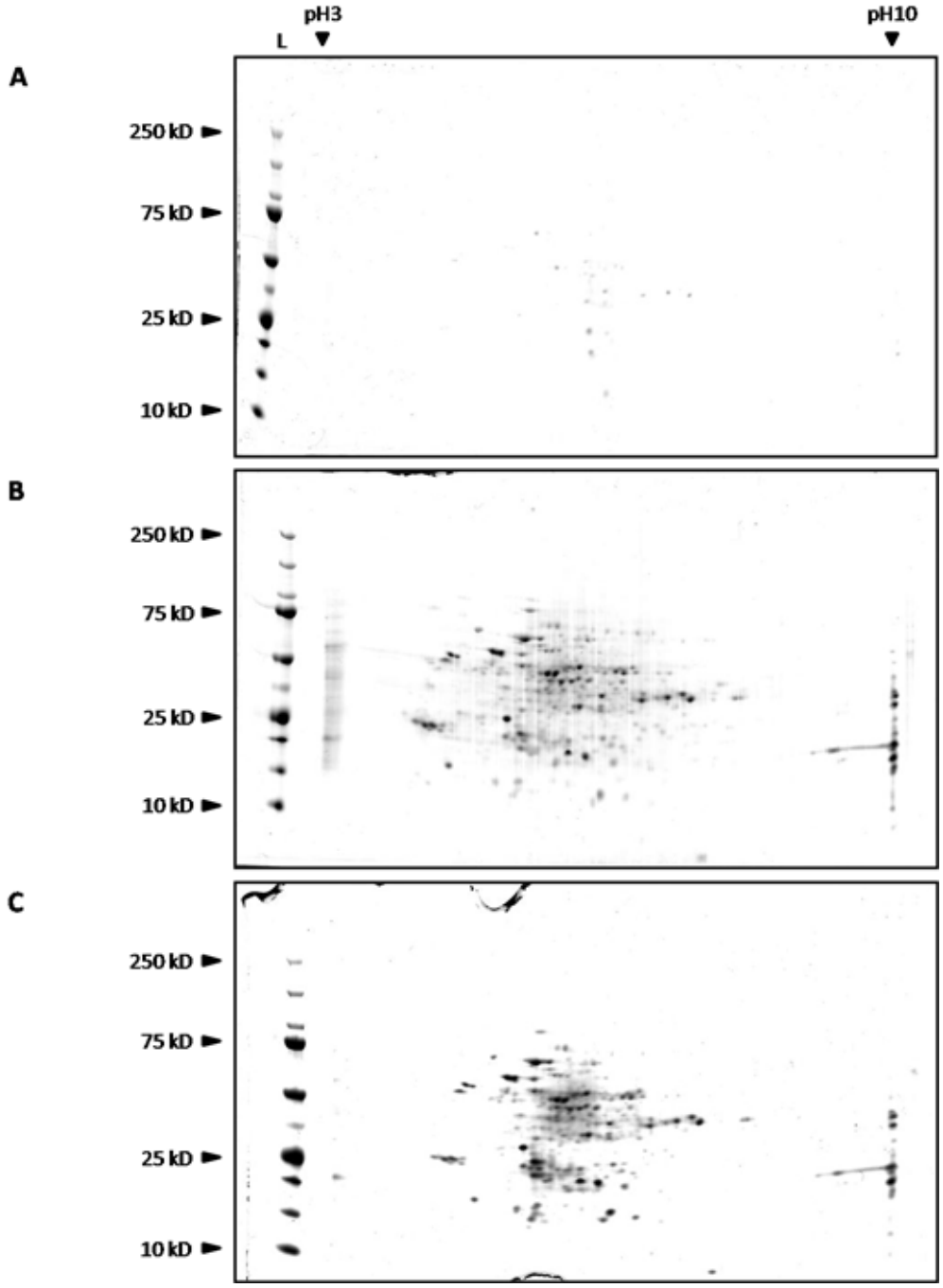


Figure 2.5 – Three different amounts of proteins were loaded onto the 11 cm gel format to determine the appropriate amount of protein required for adequate visualisation of the spots using the nanogram-sensitive CBB R250 staining technique used in this study. Gels with only 50 µg of protein (A) showed reduced loading, while little difference was seen between the spot patterns of gels loaded with 100 or 150 µg of protein (B and C respectively). Compared to gels loaded with 100 µg, gels with 150 µg displayed spots that appeared slightly less well-defined with spots of high abundance protein species (that appear as larger darker spots) masking spots of low abundance proteins. IEF strips used were pH 3-10, focused for 35000 Vhr.

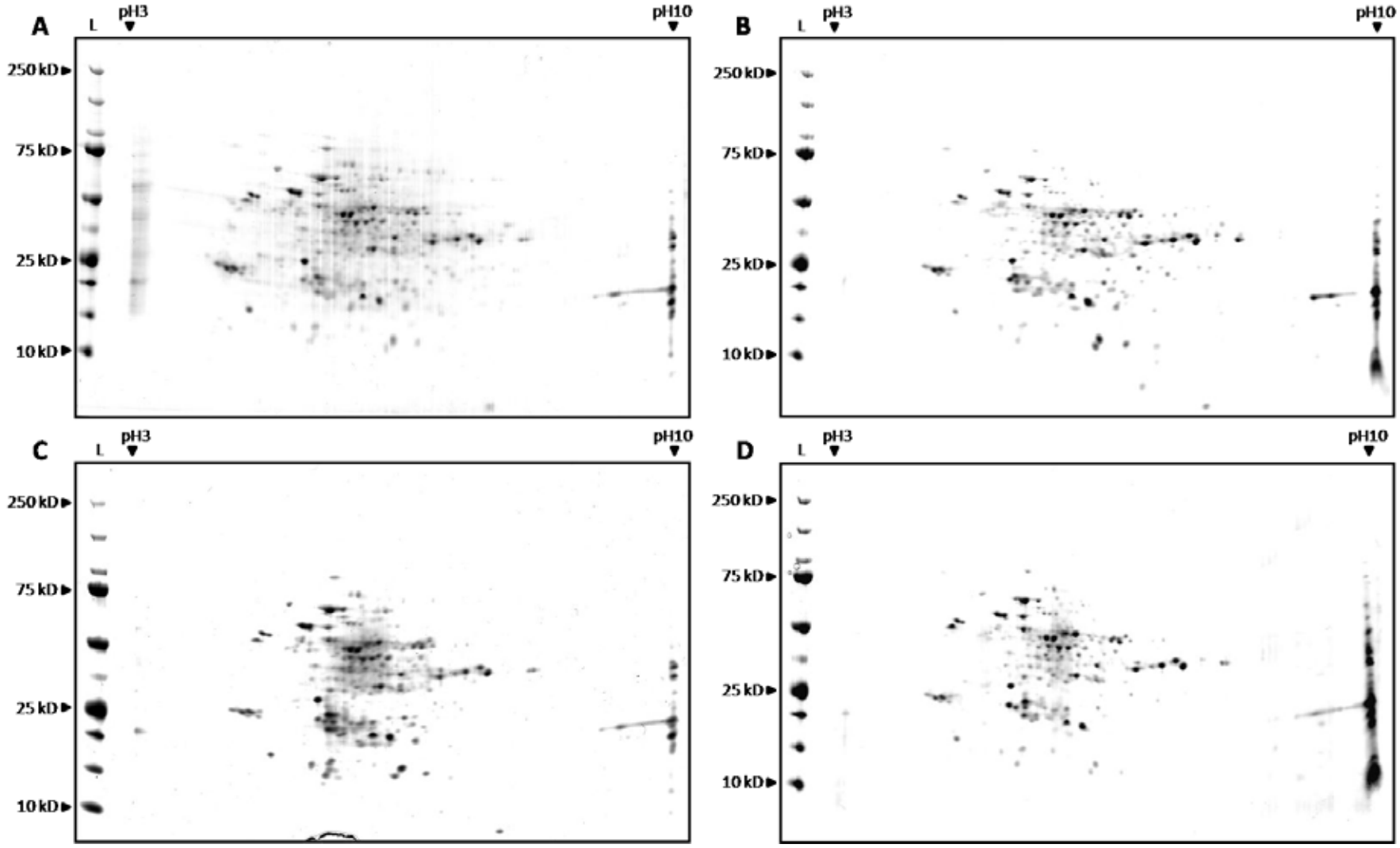


Figure 2.6 – Increasing total isoelectric focusing time resulted in increased spot resolution for IEF strips loaded with 100 or 150 µg of total protein. Gels loaded with 100 or 150 µg of protein focused for 35000 Vhr (A and C) resulted in poorer spot definition with higher amounts of background staining compared to their equivalents focused for 45000 Vhr (B and D). Although similar spot patterns were observed for gels loaded with 100 or 150 µg of proteins, over-representation of high abundance protein species in the 150 µg gels masked the presence of spots of low abundance protein species.

In an effort to further increase the horizontal resolution obtainable using the larger 11 cm format, narrow range pH strips were utilised. This involved using IEF strips with narrow pH ranges from pH 3-6, 5-8, and 7-10 (Figure 2.7 B, C, and D). Results showed that while some protein species had isoelectric potentials between pH 3-6 and 7-10, the majority of the proteins appeared to have isoelectric potentials within the pH 5-8 range. Comparisons between the relative horizontal resolutions afforded by broad range pH 3-10 (Figure 2.7 A) and narrow range pH 5-8 (Figure 2.7 C) strips showed that narrow range pH 5-8 strips gave far superior protein spot resolution while also encompassing the majority of protein species present within the samples analysed.

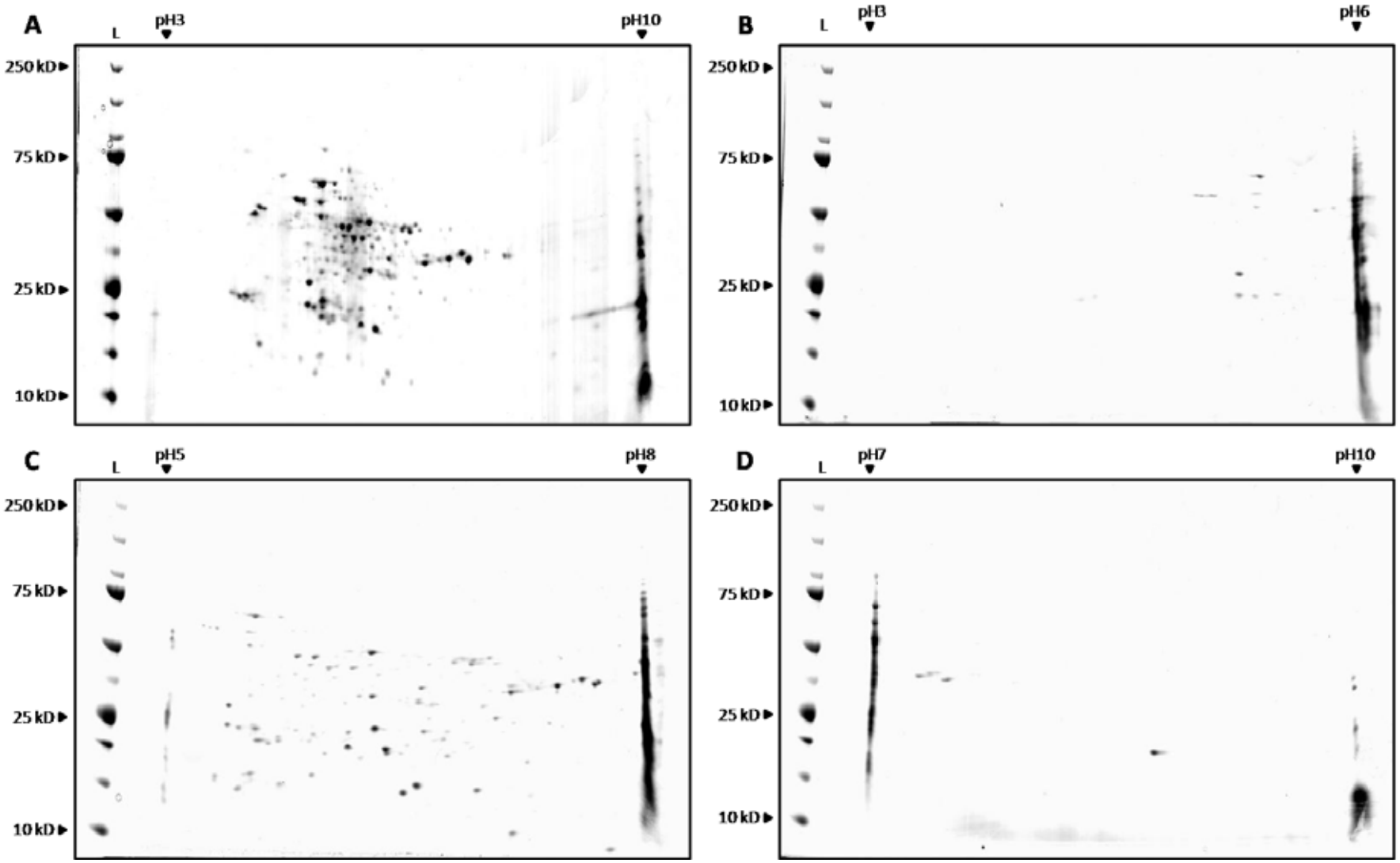


Figure 2.7 – Horizontal resolution of the 11 cm 2DGE format was further optimised by utilising narrow range IEF strips. IEF strips were loaded with 100 µg of protein (pooled stage T-IP) and focused for 45000 Vhr. Horizontal resolution of the pH 3-10 strip (A) was compared to those of narrow range strips (B – pH 3-6; C – pH 5-8; D – pH 7-10). The narrow range pH 5-8 strip encompassed the large majority of protein species within the analysed samples and was thus selected for use in the rest of the 2DGE experiments within this study.

2.3.2.2 – Vertical resolution (2nd dimension of 2DGE)

Optimisation of vertical resolution was achieved by loading 100 µg aliquots of the same T-IP protein extract onto 11 cm IEF strips that were focused then inserted into Bis-Tris gels of differing percentages (Figure 2.8 A and B). Vertical separation of the proteins by molecular weight was significantly better in 12% Bis-Tris gels compared to that of 10% gels. This was evident when comparing the distances between equivalent spots in each gel.

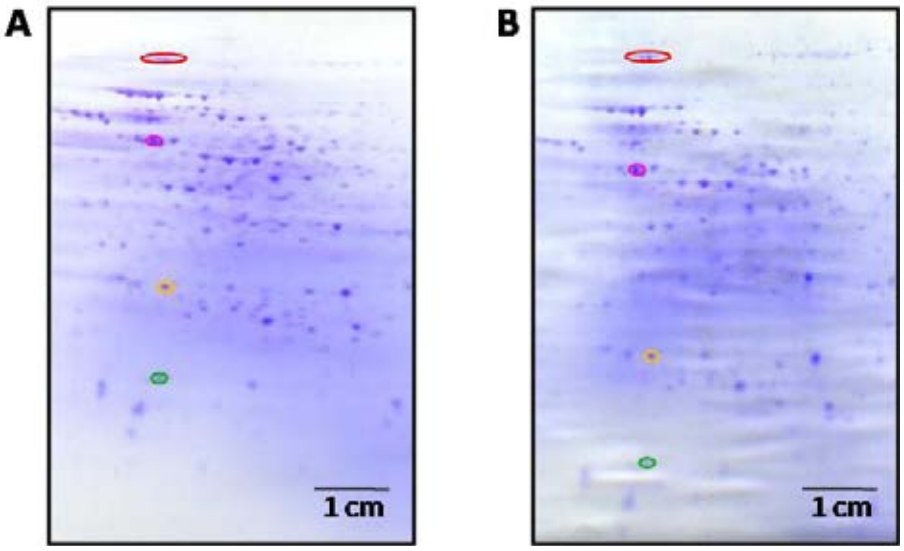


Figure 2.8 – Optimising vertical resolution for 2DGE. Aliquots (100 µg) of T-IP protein extracts were focused for 45000 Vhr then separated in the second dimension in 10% Bis-Tris gels (A) and 12% Bis-Tris gels (B). Four reference spots (denoted by the red, purple, yellow and green ovals) were identified and the distances between them measured for each gel. In all cases, the 12% gels gave greater vertical separation by molecular weight than the 10% gels.

2.3.3 – Optimising a spot visualisation technique

In addition to the nanogram sensitive CBB R250 spot visualisation technique used, other visualisation techniques such as the super-sensitive 2-dimensional fluorescence difference gel electrophoresis (DIGE) method and mass-

spectrometry (MS) compatible Eriochrome black T (EBT) silver staining were investigated. This was done to determine whether an increase in spot visualisation sensitivity and hence increased spot data, could be obtained from the 2D gels.

2.3.3.1 – 2-dimensional fluorescence difference gel electrophoresis (DIGE)

The use of fluorescent CyDyes in DIGE appeared to be a good alternative to classical visualisation techniques such as Coomassie staining and silver staining. This is because CyDyes allow for much more sensitive detection of protein spots resulting in less protein sample(s) needing to be loaded per gel. In addition to following the manufacturer's protocol, optimisation of the DIGE protocol included varying the length of the IEF protocol and reducing the ampholyte concentration from 2% to 1%. However, DIGE was not compatible with the samples in this study (Figure 2.9) with a significant amount of smearing present within the gels, most likely indicative of under- or over-focusing of the samples. This isoelectric focusing problem could not be remedied by decreasing (Figure 2.9 A) or increasing (Figure 2.9 C) the total Vhr of the IEF protocol even though the reduced ampholyte concentration ensured that the samples were focused at the maximum voltage of 8000 V.

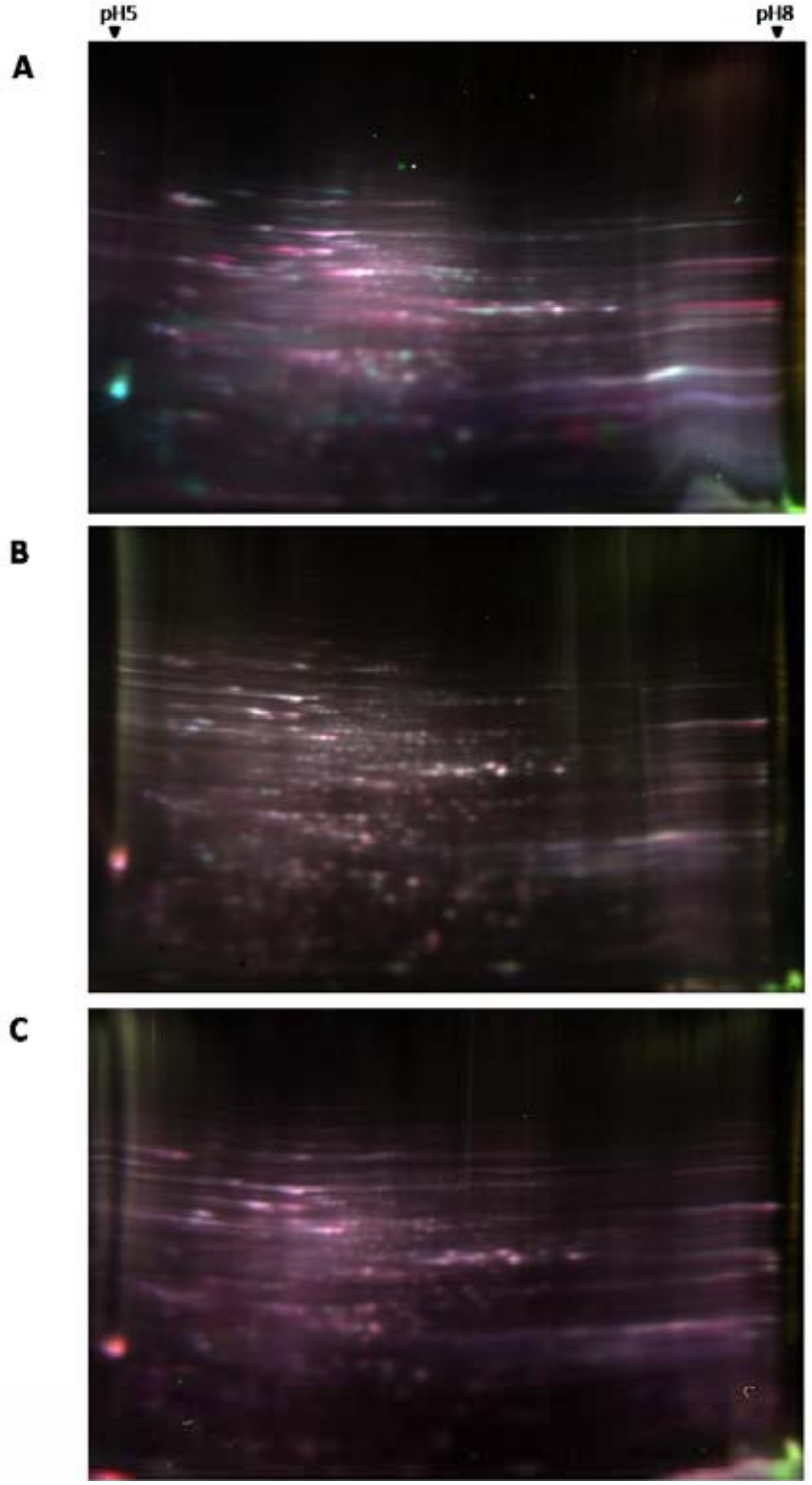


Figure 2.9 – Significant amounts of smearing were detected in DIGE gels with protein samples analysed in this study. One hundred µg of protein (33.33 µg of Cy2-tagged sample, 33.33 µg of Cy3-tagged sample, and 33.33 µg of an equivalent mixture of both samples tagged with Cy5 to be used as an internal standard), added with 1% ampholyte (v/v), was loaded onto each narrow range pH 5-8 11 cm IEF strip and then focused for (A) 35000 Vhr, (B) 45000 Vhr or (C) 55000 Vhr.

2.3.3.2 – MS compatible eriochrome black T (EBT) silver staining

Having determined that DIGE was not compatible, a MS compatible EBT silver stain protocol was investigated (section 2.2.6.2). Although less sensitive than DIGE, EBT silver staining is significantly more sensitive than the CBB R250 method. Consequently, only 75 µg of protein had to be loaded per gel instead of the 100 µg required for the CBB R250 method. This protocol involved staining the gels with CBB R250, de-staining and then staining the gels with the EBT silver stain. The resulting gels had a larger number of spots that were more defined (Figure 2.10 A and B).

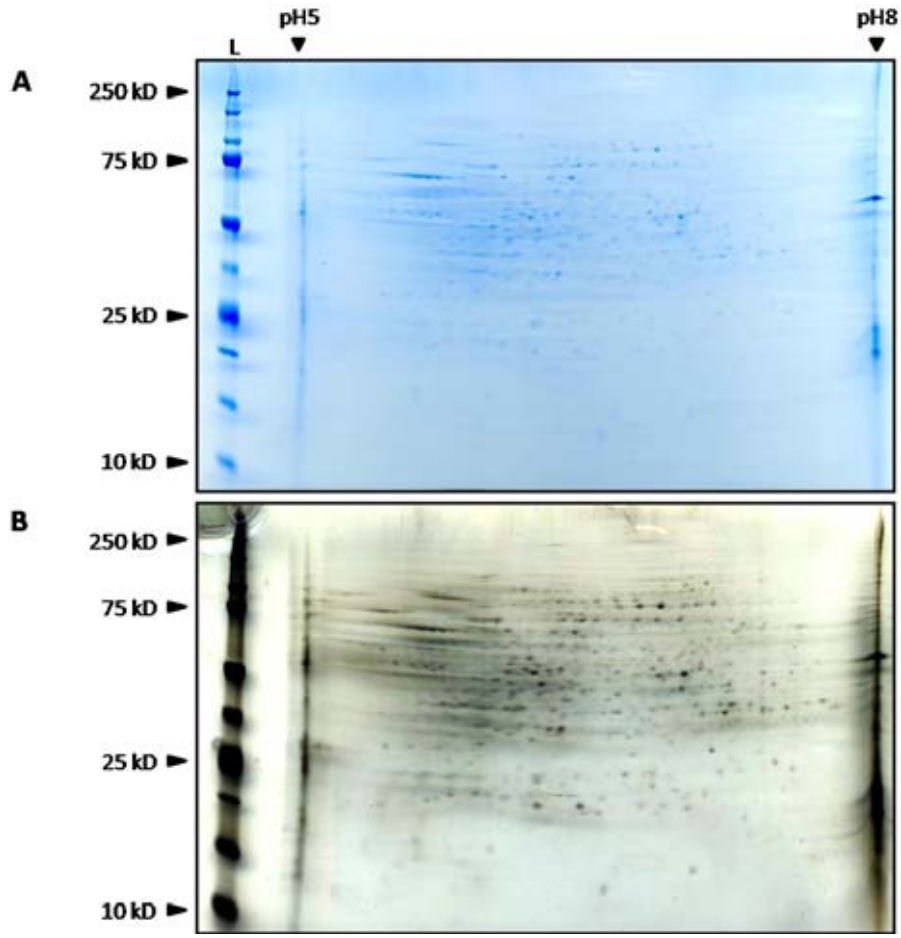


Figure 2.10 – Sensitivity of protein spot visualisation was increased by using the EBT protocol post-CBB R250 staining. Protein (75 µg) was loaded onto an 11 cm pH5-8 IEF strip and focused for 45000 Vhr then separated on a 12% Bis-Tris gel. The gel was stained with CBB R250 (A) and a scanned image obtained before being de-stained. The gel was re-stained with the EBT silver stain, and another scanned image obtained (B).

2.3.4 – 2DGE of anther proteins from four pooled stages of meiosis

The triplicate gels of each pooled stage showed consistent spot patterns in addition to an increase in spot number across the triplicate gels of the four pooled stages (Figure 2.11). The gels also showed a decrease in lower molecular weight proteins and a consequent increase in those with higher molecular weights. In addition, visual analyses of the gels showed that PM-LP and D-A gels had almost identical spot patterns. Thus comparisons between the pooled stages were only conducted between the PM-LP and TI-TII pooled stages.

Gel analysis was however complicated by the large number of spots present within each gel leading to the 2DGE software being unable to correctly detect and match equivalent protein spots accurately even with manual spot matching of numerous spots. To overcome this, the tissue type used for this study was refined from anthers to meiocytes in an effort to reduce 'contaminating' protein species extracted from mitotic cells that constitute the anther sac. A pilot experiment showed that protein extracts from meiocyte-enriched samples (Figure 2.12 A) resulted in a reduced number of protein spots compared to that of whole anthers (Figure 2.12 B).

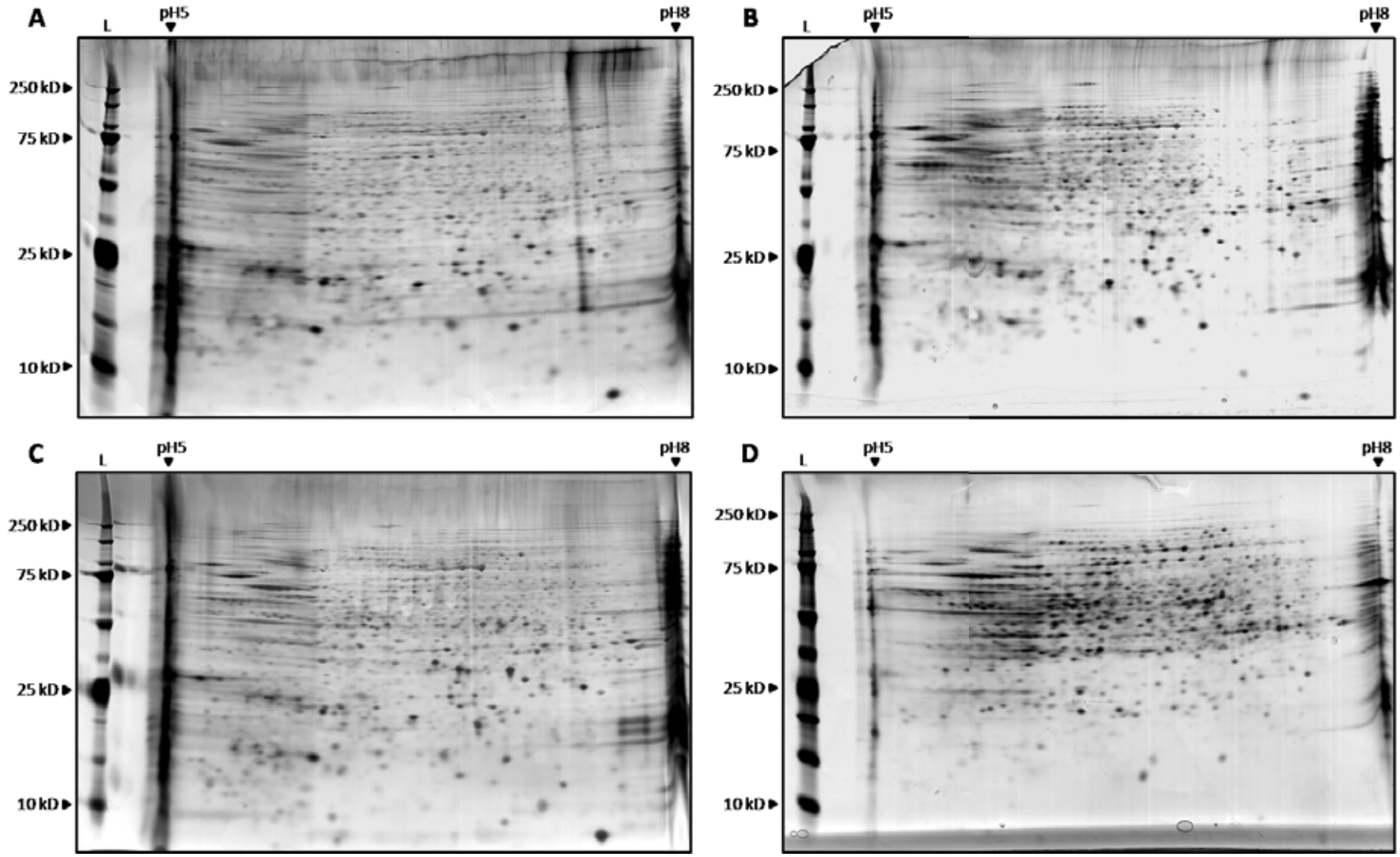


Figure 2.11 – Spot numbers increased as meiosis progressed across the four pooled stages:- (A) PM-LP, (B) D-A, (C) TI-TII, and (D) T-IP. Interestingly, the increase in spot number correlates with data previously reported by Crismani *et al*. (2006). In addition, a progressive increase in higher molecular weight proteins was paired with a decrease in lower molecular weight across the four pooled stages. Gels shown are representative of three replicate gels of three biological replicates of each pooled stage.

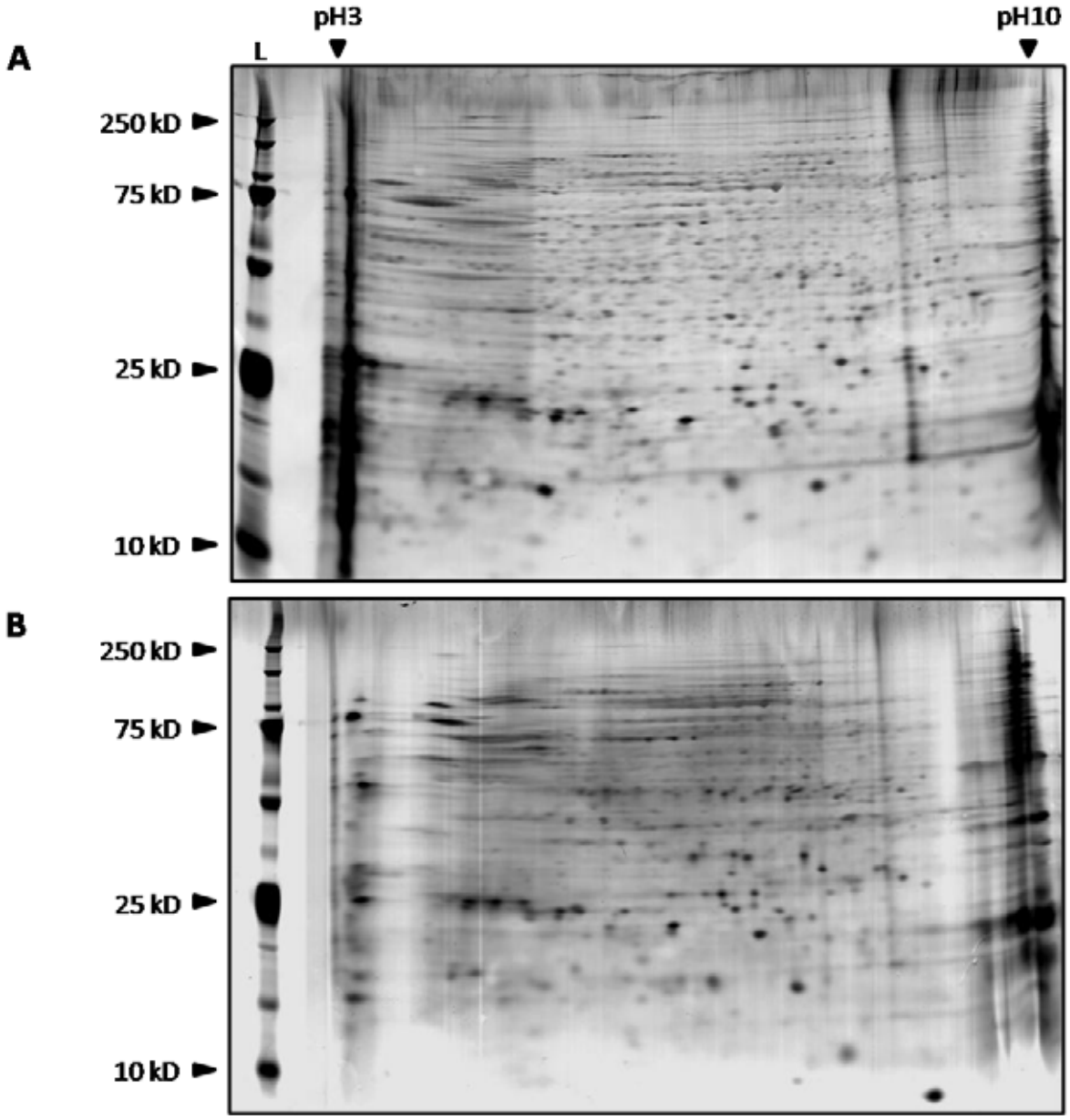


Figure 2.12 – Changing the tissue type analysed from whole anthers (A) to meiocytes (B) reduced the number of spots present in the 2DGE gels. Total protein was extracted from whole anthers and meiocyte-enriched samples (where meiocytes were extruded from the anther sacs) using the TCA-acetone precipitation method. Total protein (75 µg) was loaded onto each narrow range pH 5-8 IEF strip and focused for 45000 Vhr then separated on 12% Bis-Tris gels. Gels were then stained with CBB R250, de-stained and then stained with the EBT silver stain protocol.

2.3.5 – 2DGE of PM-LP meiocyte-enriched samples

Based on the results shown in section 2.3.4, further analyses were conducted using total protein extracts from meiocyte-enriched samples. Due to the specific nature of this study (that is, the identification of proteins required during the early stages of meiosis), and the extensive amount of time required to obtain meiocyte-enriched samples as well as the limited

availability of the tissue samples; only two pooled stages were analysed – namely PM-LP and TI-TII. For analyses between stages, wild-type Chinese Spring PM-LP and TI-TII samples were used while PM-LP samples of wild-type and *ph* mutants (*ph1b* and *ph2a*) were used to investigate differences that are present between the three genotypes during the early stages of meiosis.

Analyses of the gels conducted using both 2DGE analysis software and manual visual detection revealed three regions with differences (Figure 2.13), with six spots being differentially detected (named KK01, KK02, KK03, KK04, KK05, and KK06) (either between pooled stages, genotypes, or both) (Table 2.1, Figure 2.14 and Figure 2.15). The physicochemical properties of these regions are detailed in Table 2.1.

Table 2.1 – Approximate physicochemical properties of six candidate spots identified via 2DGE analyses. Total protein extracts from two pooled stages (PM-LP and D-A) of Chinese Spring were analysed to determine differences between stages of meiosis while PM-LP samples of both *ph* mutants were compared to that of Chinese Spring to identify possible spot differences between the genotypes that may help elucidate the mechanisms that control homologous pairing.

Spot name(s)	Notation of region in Figure 2.13	Notation of spot in Figure 2.14	Approximate molecular weight of region	Approximate pH of region
KK01	Red square	Red oval	20-25 kD	pH 6.0-6.5
KK02		Blue oval		
KK03		Orange circle		
KK04		Green circle		
KK05	Blue rectangle	Yellow circle	60-100 kD	pH 7.0-7.5
KK06	Green rectangle	Purple oval	50-100 kD	pH 7.5-8.0

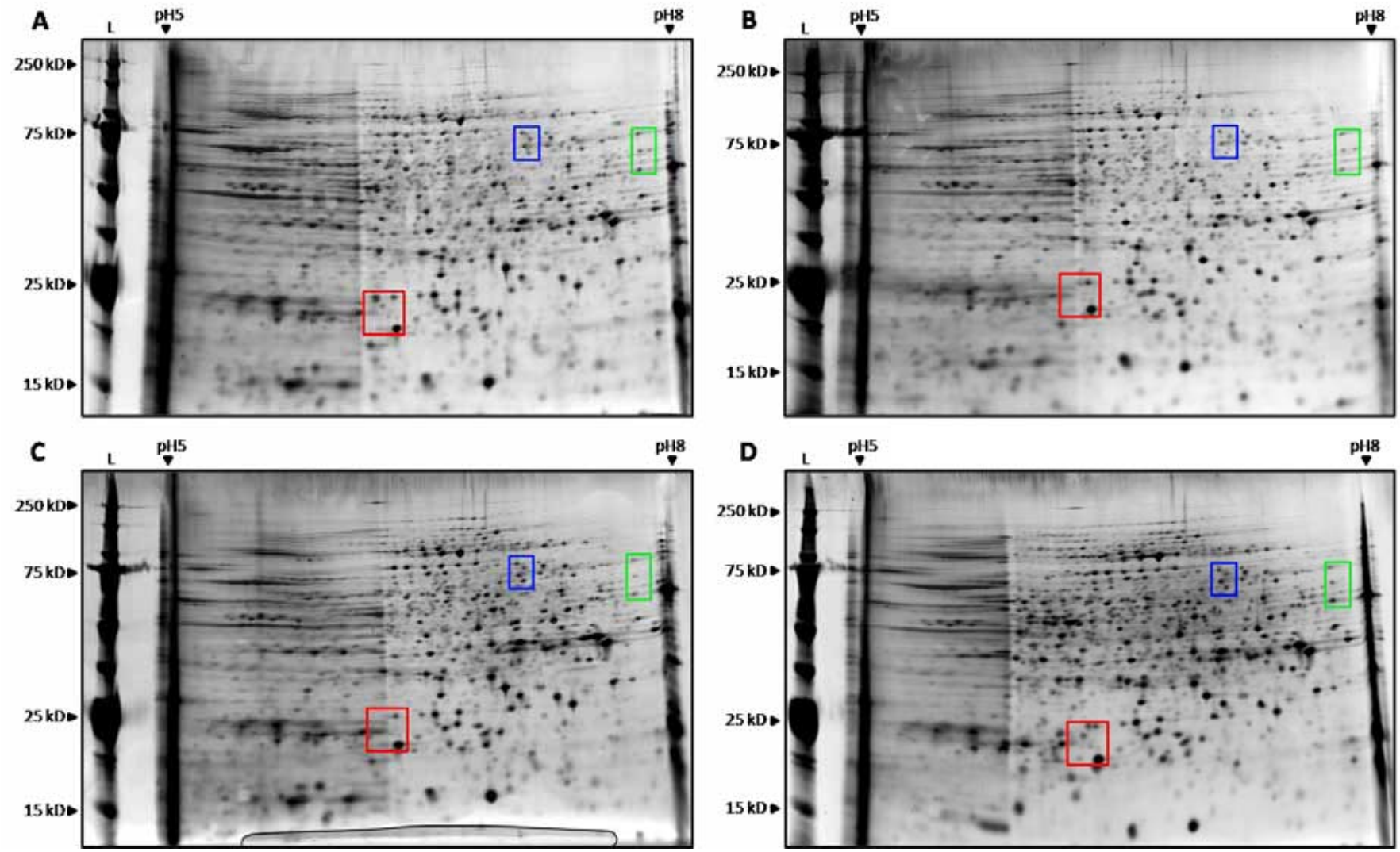


Figure 2.13 – 2DGE analyses between PM-LP and TI-TII protein extracts of Chinese Spring (A and B respectively) and the PM-LP extracts of the *ph* mutants (*ph1b* – C, *ph2a* – D respectively) revealed three regions with differing spot patterns. Six candidate spots (KK01 – KK06) were identified to be differentially detected either between pooled stages or genotypes, or both. KK01, KK02, KK03 and KK04 were located within the red square while KK05 and KK06 were located within the blue and green rectangles respectively. Gels shown are representative of three technical replicates of three biological replicates.

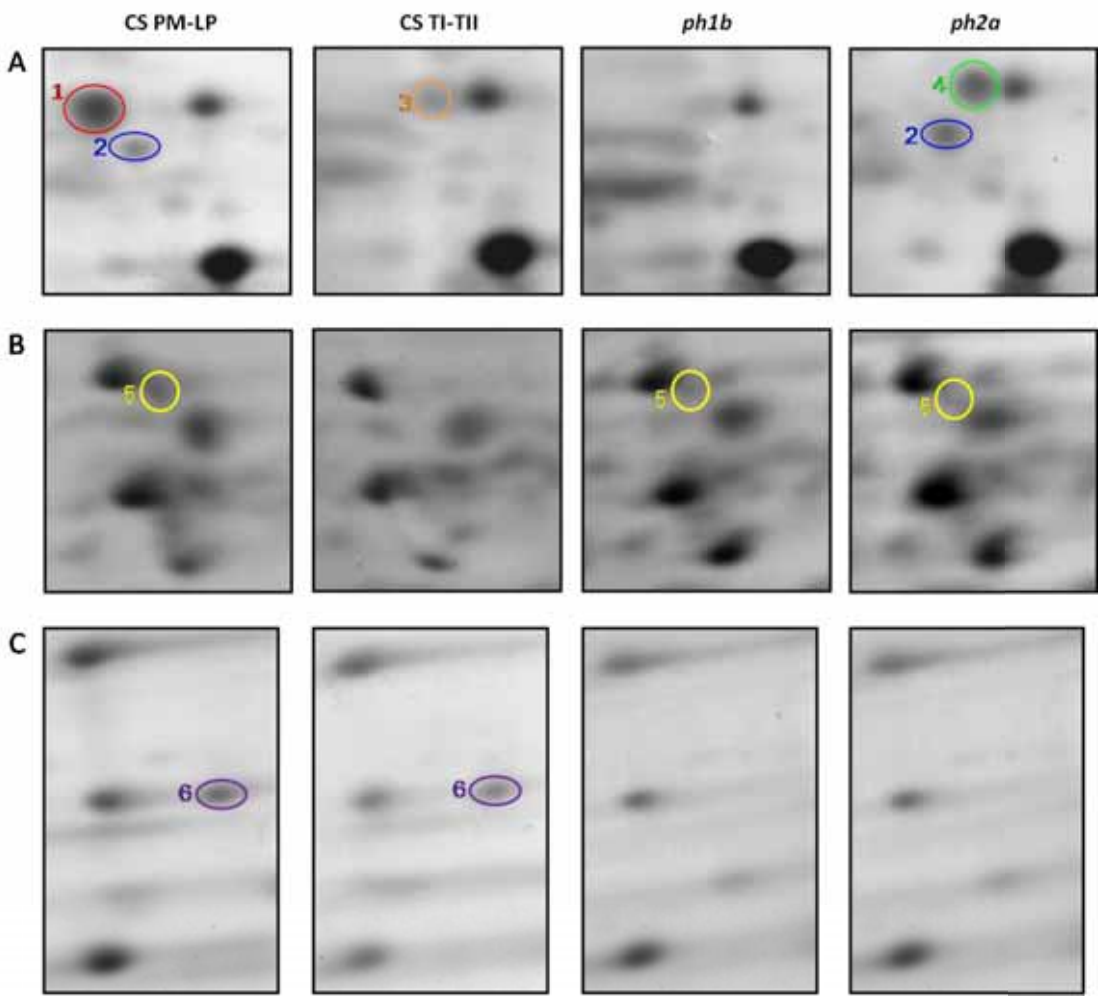


Figure 2.14 – Magnified view of the three regions with spot differences seen in Figure 2.13 (A – red square; B – blue rectangle, C – green rectangle). Candidate spots KK01 (red circle), KK02 (blue oval), KK03 (orange circle) and KK04 (green circle) were identified within the region denoted by the red square in Figure 2.13 while KK05 (yellow circle) and KK06 (purple oval) were identified within the blue and green rectangles in Figure 2.13 respectively. Spots were identified using ImageMaster 2D Elite software as well as by manual inspection.

Candidate spot KK01 was present only in the wild-type PM-LP sample while KK02 was present in both PM-LP samples of wild-type and the *ph2a* mutant (Figure 2.14 A, red oval and blue oval respectively). Analyses of KK03 and KK04 were complicated due to the fact that both spots appeared at almost the same coordinates within the gels *albeit* with KK04 having a slightly higher pI and spot intensity in the *ph2a* PM-LP samples than KK03 in the wild-type TI-TII samples (which possibly indicated that they were isoforms of the same protein

species). KK05 was present in all PM-LP samples but was absent in the later stages of meiosis (Figure 2.14 B, yellow circle) while KK06 was present in both pooled stages of the wild-type but absent in both PM-LP samples of the *ph* mutants (Figure 2.14 C, purple oval). Statistical analyses of the normalised spot volumes revealed all spots were present at different levels of abundance which were statistically significant ($p<0.05$) (Figure 2.15). The spots were then excised from the gels and sent for mass-peptide identification (refer to Chapter 3).

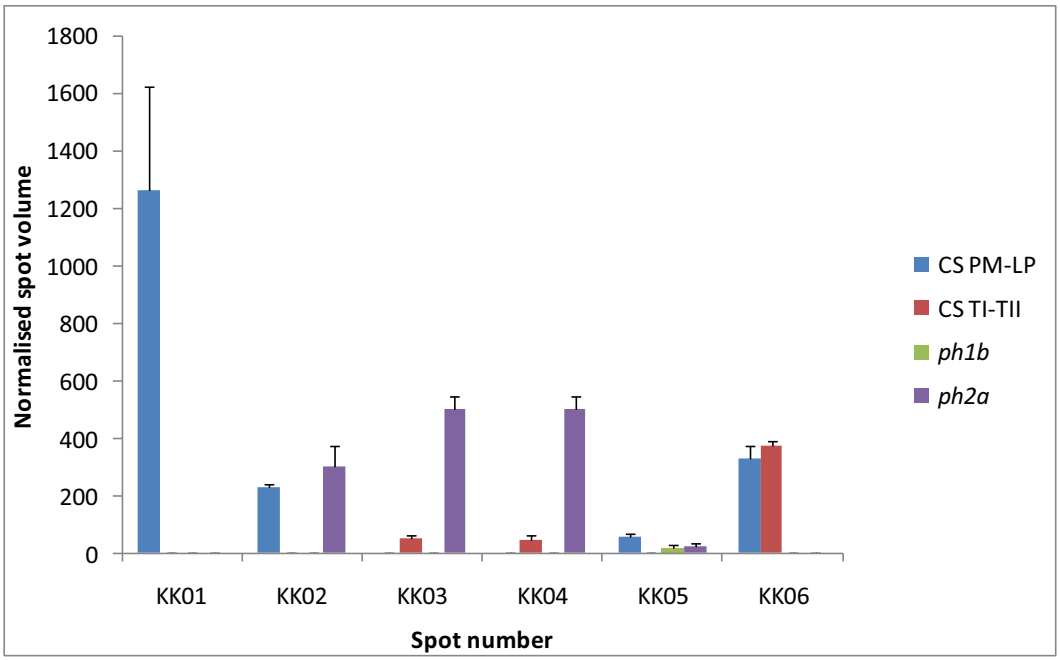


Figure 2.15 – Comparisons of normalised spot volumes of the six candidate spots identified. KK01 was present only in the wild-type PM-LP sample while KK02 was present in PM-LP samples of both wild-type and the ph2a mutant. Statistical analysis was conducted on the assumption that KK03 and KK04 were isoforms of the same protein species. However, if these were indeed separate protein spots, KK04 was present at significantly higher levels (based on LSD value) in the PM-LP ph2a samples than KK03 was in the TI-TII wild-type samples. KK05 was present in all PM-LP samples analysed albeit at very low levels while KK06 was only present in both pooled stages of the wild-type. LSD values ($p < 0.05$) for KK01, KK02, KK03, KK04, KK05 and KK06 were 558.7, 116.8, 70.5, 70.5, 32.62 and 74.0 respectively. Data shown are the mean ± SE of three biological replicates for each sample (n=12).

2.4 – Discussion

The aim of this study was to optimise a proteomics methodology that was suitable for investigating meiosis in bread wheat with a view to identifying proteins that have roles in meiosis. This was achieved by determining and optimising a suitable protein extraction technique from wheat anthers and meiocytes, systematically optimising parameters that affect 2DGE, investigating protein spot visualisation techniques of differing sensitivities, and refining the tissue type used for the analyses.

An essential factor in any proteomic analysis is sample quality which is directly affected by the protein extraction method utilised. The choice of extraction technique used is influenced by multiple factors such as the starting material (tissue type, availability and time required for harvesting the samples), the quantity and quality of the extracted sample required for downstream experiments as well as the subset of proteins targeted for analysis. Protein species extracted by any given protocol are determined by the biochemical properties of the proteins and their compatibility with the protocol used. Thus, no one single extraction is able to extract all cellular proteins; with differing protein extraction techniques having been shown to yield different subsets of protein species (Carpentier *et al.* 2005).

Of the three extraction techniques investigated in this study, the TCA-acetone extraction and precipitation technique proved the most successful with proteins of a broad molecular weight range (10 kD to 250 kD) being extracted. While the amount of protein yielded using this technique was lower than that of the other two investigated (data not shown), the quality of the proteins extracted was significantly better (as analysed by visual inspection of SDS-PAGE gels,

Figure 2.2) with tight bands indicating intact protein species were extracted. In comparison, both the urea-mercaptoethanol-NP40 and phenol extraction techniques yielded smears within the gel lanes, most likely indicative of protein degradation and possible contamination by interfering substances. Although either 1 or 5 µg of total protein (as quantified via the Bradford assay) were loaded into the gel lanes, no bands or smears were present in lanes loaded with extracts of extraction protocol 2 (using insoluble PVP). This was most likely due to contaminating substances (such as residual PVP debris from the extraction buffer, free amino acids, or plant pigment molecules) giving false spectrophotometric readings during the Bradford assay.

The use of the TCA-acetone extraction method had the added benefits of being a rapid, economical and efficient protocol with highly reproducible results (Figure 2.3). In addition, proteases and other modifying enzymes can be instantly inactivated thus preventing proteolysis and unwanted modifications to the proteins extracted (Wu and Wang 1984). To further prevent residual enzymatic activity and degradation, only pre-chilled solutions and apparatus were used.

A minor drawback of the TCA-acetone protocol is the high level of residual salts in the extracts that leads to high current conductivity during IEF resulting in heat build-up (Mao and Pawliszyn 1999) that can cause carbamylation of the protein samples (Isaacson *et al.* 2006). This consequently leads to incomplete sample focusing resulting in smearing of protein spots and possibly the presence of carbamylation trains in 2D gels. To avoid this, the protocol was optimised by the addition of a 90% acetone (v/v) in nanopure water wash step to reduce the salt concentration. A prolonged active rehydration step (30 V for 13 hr) and an initial low voltage step (200 V for 1 hr) were also used to draw

contaminating salt ions to the electrodes. In combination, these optimisation steps allowed complete isoelectric focusing to be achieved with samples reaching 8000 V and 50 µA per IEF strip. However, some residual salt contamination was still present and led to light spot smearing and background staining at the acidic end of the 2D gels (Figure 2.11 and 2.13).

Spot resolution for this study was increased by the use of an 11 cm IEF strip and gel format after initial tests using 7 cm strips and gels were completed. Optimisation of this larger format involved investigating the appropriate amount of protein to load and the IEF parameters required for correct focusing of the samples. While the manufacturer's protocol recommends 35000 Vhr, the results showed that spot resolution was increased and background noise decreased when the total IEF time was increased to 45000 Vhr (Figure 2.6, comparing A and C with B and D). These gel results further highlight the importance of optimising an IEF protocol that is specific to the samples used in each study.

Horizontal resolution was further increased by using narrow range pH IEF strips that allowed more targeted protein spot analyses. Final analyses of all samples was done using narrow range pH 5-8 strips as the majority of the protein spots appeared within this range (Figure 2.7).

Vertical resolution of the 2D gels was optimised by investigating the migration of reference protein spots in different percentage Bis-Tris gels. Experiments were conducted with Bis-Tris gels which had a significantly lower pH value (6.5) compared to that of a Tris-glycine system, which has a pH value of 8.5. Previous work done by Anfinsen and Haber (1961) has shown that the pH value at which a SDS-PAGE system operates influences the rate of protein migration through the gel. They showed that cysteine residues, required for

disulphide bond formation, have a pKa value of between 8 and 9. Disulphide bond formation therefore possibly increases at the pH of 8.5 used in a Tris-glycine SDS-PAGE system, This could have negative consequences on an SDS-PAGE conducted under denaturing conditions as refolding of proteins back to their tertiary configurations allows movement through the gel at different rates (Anfinsen and Haber 1961). This was avoided in this study by using the Bis-Tris system which has a lower pH value allowing denatured protein species to separate more efficiently and precisely according to their molecular weights. The use of fixed percentage gels was also preferred to that of gradient gels as fixed percentage gels allowed for more accurate molecular weight estimation. Twelve % gels were used in the final analyses as they gave superior vertical resolution compared to the 10% gels (refer to Figure 2.8).

Although the spot resolution achieved was deemed sufficient for this study, future work could investigate using even longer IEF strips and larger gel formats. This would increase resolution dramatically but would also require a larger amount of protein per gel. A possible way to overcome the requirement for increased amounts of protein is to use a more sensitive visualisation technique. Indeed, this study investigated three spot visualisation techniques in an effort to reduce the amount of protein needed as well as to generate the best possible gel data for analysis. The amount of protein loaded per strip is largely dependent on the sensitivity of the visualisation technique used downstream. When the nanogram-sensitive CBB R250 technique was used, 100 µg of protein was required while only 75 µg was needed for MS-compatible EBT silver staining. Impressively, only 33.33 µg of each protein extract was required with the super-sensitive DIGE technique. Moreover, DIGE also allows two different samples as

well as an internal standard (consisting of a mixture of both samples) to be run within a single gel due to the use of three Cy-dyes that fluoresce at different excitation wavelengths. This reduces gel-to-gel variation and allows for more accurate spot analyses. Substantial amounts of time and funds were invested to optimise this technique but multiple attempts were unsuccessful with unfocussed spots and smearing present even when ampholyte concentrations were reduced to 1% and varying lengths of total Vhr were used in the IEF protocol. Due to time and financial constraints, DIGE was deemed non-viable for this study and was replaced with the MS-compatible EBT silver stain technique. Although the EBT silver staining technique does not allow 100% accurate quantification of protein spots, this study successfully used it to identify six spots that were present at different levels within the cell.

Software analyses of 2D gels loaded with total cellular proteins extracted from wheat anthers that had been pooled into four stages proved difficult and time-consuming due to the large number of spots present (Figure 2.11). This was not completely unexpected as previously published literature has shown that spot detection software is not sophisticated enough to detect large numbers of spots, decipher complex spot patterns, nor can it accurately and reliably pair equivalent spots across multiple gels in large experiments without a large amount of manual assistance that can be subjective (reviewed by Clark and Gutstein 2008). Despite this, observation of the gels showed an increased number of protein spots as meiosis progressed, in agreement with mRNA transcript data previously published by Crismani and colleagues (2006). Manual inspection of the gels revealed that the spot patterns present on PM-LP and D-A gels were very similar with no discernable differences seen. Two possible explanations for this

observation are: 1) some protein species present during the PM-LP stage may also have roles (or persisting roles) during the D-A stage and hence are maintained within the cell; or 2) degradation of PM-LP proteins no longer required during the D-A stage is not yet complete (due to the short time-frame between the end of pachytene and the start of anaphase I). The T-IP gels however revealed a clear increase in higher molecular weight proteins and that the majority of the proteins resided within the 25 – 100 kD range. This was reflective of 2D gels loaded with protein extracts from rice anthers at the equivalent stage (Imin *et al.* 2001).

The change of tissue type used from whole anther to enriched meiocyte samples was made to reduce the complexity of the spot patterns obtained (Figure 2.12). The removal of the anther sac (which is made of mitotic cells) encasing the meiocytes helped reduce 'contamination' caused by proteins extracted from these cells. Conversely, enriching for meiocytes should result in the enrichment of proteins vital for meiotic cell division and development.

Both software and manual analyses of 2D gels loaded with total cellular proteins extracted from the PM-LP stage meiocyte-enriched samples of both wild-type and *ph* mutants as well as the D-A stage wild-type samples proved more manageable (Figure 2.11). Three regions, with a total of six spots which were present at differing levels either between stages of meiosis (PM-LP vs. D-A) or between genotypes (wild-type vs. *ph1b* vs. *ph2a*), were detected (Table 2.1, Figure 2.13, Figure 2.14 and Figure 2.15). The presence of KK01 and KK06 in wild-type samples but not in samples of the *ph* mutants suggests that these proteins may have roles in the mechanism by which homologous chromosome pairing is achieved or simply that they are protein products of genes present on the sections of the genome that are missing from each of the *ph* mutants.

Alternatively, they could also be protein products of genes located elsewhere in the genome but that are regulated by the respective *ph* loci absent in the mutants. KK06 is unique in that it appears to be present in both pooled stages of meiosis analysed but is absent in the PM-LP stage of both *ph* mutants. One plausible explanation is that the *KK06* gene is regulated by a combination of factors from both the *ph* loci where absence of either locus leads to the KK06 protein not being expressed. The presence of KK02 during early meiosis when the three tightly-linked processes of chromosome synapsis, homologous pairing and DNA recombination occur makes KK02 an interesting meiotic candidate. Its absence in the *ph1b* mutant further suggests that the *KK02* gene either resides within the *ph1* deletion region or is regulated by a *ph1* mechanism as has been seen in the case of another meiotic gene, *TaASY1* (Boden *et al.* 2009). Given that KK03 is only present in the TI-TII stage in wild-type samples, KK03 is likely to be a late meiosis protein and/or required for tetrad or pollen maturation. Interestingly, KK04 is present at almost the same coordinates within the PM-LP *ph2a* gels *albeit* at higher levels. Whether or not these two spots are the same, with KK04 representing abnormal regulation of the KK03 gene due to the loss of the *ph2a* locus, will be determined through mass-peptide identification of both proteins (Chapter 3). Similar to KK01 and KK02, KK05 appears to be required during early meiosis. However, the presence of KK05 in all three genotypes indicates that it neither resides within the *ph* loci nor is it regulated by them. Results of the mass-peptide identification of these six spots as well as the characterisation of the proteins and the genes that encode them are detailed in Chapter 3.

In conclusion, 2DGE is a viable but time-consuming and labour-intensive proteomics tool for investigating the wheat meiotic proteome. In addition, there

are limitations in its application especially when dealing with limited amounts of starting material and protein samples. Further optimisation to remove the residual salts and contaminants carried over from the extraction method used may enable the use of the DIGE procedure that would most likely overcome these limitations. In the future, the use of cutting-edge cell flow-cytometry techniques may enable the sorting of meiocytes into each stage (or even sub-stage) of meiosis thus reducing sample harvesting time and allowing more detailed analyses of the wheat meiotic proteome. However, even with the limitations imposed by this technique, six protein spots whose expression differed in a context likely to be biologically important were identified.

Chapter 3 – Characterisation of proteomics candidates

3.1 – Introduction

While plant meiosis research has been ongoing for decades, most studies have relied on various genetic-based approaches including mutant population screens, microarray analyses, cloning and transgenesis. However, protein-based approaches such as proteomics have been seldom reported, at least in investigating cellular processes like meiosis (Phillips *et al.* 2008, Sánchez-Morán *et al.* 2005).

One advantage of using a proteomics-based approach is that it provides a broad snap-shot of the dynamic protein profile present within a given cell type at any developmental stage. In addition, proteomics data is as important, if not more important, than transcript expression profiling data as it allows researchers to study the proteins, the work-horses of the cell, rather than the mRNA transcripts that may yet be post-transcriptionally regulated. Ideally, both transcriptional and proteomics data should bolster one another but multiple previous studies have proven otherwise (reviewed by Waters *et al.* 2006 and references therein). This highlights the need for proteomics as a research tool in our efforts to understand the cellular mechanisms of the living cell. Although 2DGE has been one of the most widely used proteomics approaches for studying various facets of cellular biology, 2DGE has only been utilised recently to study plant meiosis (Phillips *et al.* 2008, Sánchez-Morán *et al.* 2005).

The objective of the work presented in this chapter was to isolate and characterise candidate genes that encode the protein products responsible for the six spots (KK01-KK06) that were identified to be differentially expressed in

Chapter 2. Characterisation of selected candidates was achieved using several approaches including RACE PCR, Southern blot analyses, Q-PCR profiling, and *in vitro* DNA-binding assays. Through these combined approaches, two candidates with possible roles during the early stages of meiosis in bread wheat have been identified.

3.2 – Materials & Methods

3.2.1 – Mass-peptide identification of candidate spots

3.2.1.1 – Preparation of candidate protein spots

Gels were placed in a transparent glass tray over a light-box and the six spots were excised from their respective gels using 200 µL pipette tips that were snipped at the end to fit the size of the spots. Gel plugs obtained with these tips were washed twice for 20 min in 100 mM ammonium bicarbonate (Sigma) and then dried with a SpeedVac SVC100 (Selby Instruments, Notting Hill, Victoria, Australia). Samples were rehydrated in 100 mM ammonium bicarbonate and in-gel digestion was conducted with 40 ng of Trypsin Gold (Promega Corporation, New South Wales, Australia) at 37°C for 4 hr. The supernatants containing the digested proteins were recovered post-centrifugation at 13000 rpm and analysed by tandem mass spectrometry (MS/MS). This was conducted at the Flinders Proteomics Centre in conjunction with Dr Tim Chataway using Ion-Trap Liquid Chromatography-Electrospray Ionisation tandem mass spectrometry (MS/MS).

3.2.1.2 – Mass spectrometry (MS)

The digested peptides were analysed with a Thermo LTQ XL linear ion trap mass spectrometer fitted with a nanospray source (Thermo Electron Corp, San Jose, CA, USA). The samples were applied to a 300 μm i.d. x 5 mm C18 PepMap 100 precolumn and separated on a 75 μm x 150 mm C18 PepMap 100 column using a Dionex Ultimate 3000 HPLC (Dionex Corp, Sunnyvale, CA, USA) with a 55 min gradient from 2% acetonitrile to 45% acetonitrile containing 0.1% formic acid at a flow rate of 200 nl min^{-1} followed by a step to 77% acetonitrile for 9 min. The mass spectrometer was operated in positive ion mode with one full scan of mass/charge (m/z) 300-2000 followed by product ion scans of the three most intense ions with dynamic exclusion of 30 s and collision induced dissociation energy of 35%.

The MS spectra were searched with Bioworks 3.3 (Thermo Electron Corp, San Jose, CA) using the Sequest algorithm against the Rice Genome Annotation (Build 6.1) using Trypsin digestion as the protease and allowing for two missed cleavages, with the following filters: 1) the cross-correlation scores (Xcorr) of matches were greater than 1.5, 2.0 and 2.5 for charge state 1, 2 and 3 peptide ions respectively; and 2) peptide probability was greater than 0.01. The mass tolerance for peptide identification of precursor ions was 1 Da and 0.5 Da for product ions.

3.2.2 – Analyses of peptides identified by mass spectrometry

3.2.2.1 – Basic Logic Alignment Search Tool (BLAST) analyses of peptides obtained from mass spectrometry

Peptides obtained from the mass spectrometry were analysed using the online protein BLAST (BLASTp) website of the National Center for Biotechnology

Information (NCBI) (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins, NCBI-Genbank Flat File Release 173; accessed 27th October 2009). Searches were conducted using the default settings with the *Triticeae* and *Oryza* filters. BLASTp searches were also conducted using The Institute for Genomic Research (TIGR) Rice Genome Database (http://blast.jcvi.org/euk-blast/index.cgi?project=osa1, Rice Genome Annotation Build 6.1; accessed 27th October 2009). A 50% sequence identity cut-off point was used to reduce the non-specificity of the hits caused by the short peptide sequences used in the searches. In 2010, post-isolation of the candidate transcripts, the *Brachypodium distachyon* genome was completely sequenced. A second round of BLASTp searches was conducted against the completed Phytozome BLAST database (http://www.phytozome.net/search.php?show=blast, Version 5.0; accessed 2nd September 2010). Default settings were used with *Brachypodium distachyon* selected as the target genome.

3.2.3 – Isolation and amplification of candidate transcripts

3.2.3.1 – Primer design

Peptide sequences that passed the filtering and selection criteria in section 3.2.2.1 were back-translated to obtain their corresponding DNA sequences. 5′ and 3′ Rapid Amplification of cDNA (RACE) primers were designed from these DNA sequences according to the guidelines provided by the GeneRacer Kit (Invitrogen). Each RACE primer was evaluated by its annealing temperature, whereby primers with annealing temperatures lower than 52°C were rejected. Degenerate primers (with random nucleotides in the codon wobble positions) were also designed in some instances to increase the chances of amplifying a

product. Multiple rounds of RACE primers were designed based on sequencing results obtained after each RACE experiment (see Table 3.1). Primers were also designed based on the sequences of the RACE products (section 3.2.3.11) to amplify the full-length coding sequences of the candidate genes.

3.2.3.2 – Collection and staging of meiotic anthers

Staging and collection of Chinese Spring, *ph1b*, and *ph2a* wheat anther tissue from plants grown in a controlled-environment room programmed with a 16/8 hr photoperiod at approximately 23°C was performed as per section 2.2.1. Stages of anthers collected were Chinese Spring at PM-LP and TI-TII stages, and *ph1b* and *ph2a* at PM-LP stage.

3.2.3.3 – RNA extraction and quantification

Frozen anther tissue was vortex-ground using three 0.3 cm diameter metal ball-bearings in 2 mL microfuge tubes (Genomics Products Pty Ltd., Belconnen Business Centre, Australian Central Territory, Australia). Total RNA was extracted using TRI-REAGENT (Astral Scientific Pty Ltd., Caringbah, New South Wales, Australia) according to the manufacturer's instructions. The purified RNA samples were resuspended in 50 μL of 10 mM Tris[hydroxymethyl]aminomethane (TRIS) (pH7.5).

The quality of the extracted RNA was determined by gel electrophoresis. RNA samples (5 μL) were mixed with 3 μL of 6× Ficoll loading buffer and separated on a 1.5% agarose gel (w/v) using a gel electrophoresis tank (AdeLab Scientific Pty Ltd., Thebarton, South Australia, Australia). Visualisation of the RNA was conducted using ethidium bromide (Amresco) and illuminated using a

UVP First Light UV Illuminator (Ultra-Violet Products Ltd., Cambridge, United Kingdom).

Quantification of the RNA concentration was determined spectrophotometrically using a ND-1000 nanodrop machine (NanoDrop Technologies Inc., Wilmington, Delaware, USA). TRIS (10 mM; 1 μL) was used as a blank sample while 1 μL of each test/experimental sample was used to obtain the RNA concentrations and absorbance ratios ($A_{260/280}$).

Samples with low concentrations of RNA were concentrated to a volume of approximately 10 μL using a SpeedVac SVC100 (Selby Instruments) and quantified again as described above.

3.2.3.4 – Construction of cDNA and RACE libraries

5′ and 3′ RACE libraries were synthesised from 2.5 μg of total RNA per reaction using a GeneRacer Kit (Invitrogen) according to the manufacturer's instructions. The 5′ RACE library construction was performed over two days with RNA pellets frozen and stored overnight at −20°C in ethanol during the 'Ligation of RNA Oligo to Decapped mRNA' procedure. Total cDNA libraries for Chinese Spring at PM-LP and TI-TII stages, and *ph1b* and *ph2a* at PM-LP stage were also constructed using 1 μg of total RNA per IScript™ cDNA Synthesis Kit (BIO-RAD) reaction according to the manufacturer's protocol.

3.2.3.5 – 5′ and 3′ RACE Polymerase Chain Reaction (PCR)

Multiple rounds of 5′ and 3′ RACE were conducted according to the manufacturer's protocol with some modifications. Firstly, both primary and secondary reaction volumes were scaled down to 25 μL. Secondly, lowered

annealing temperatures were used in some instances to accommodate the lower annealing temperatures (T_m) of the primers (see Table 3.1 for primers and annealing temperatures used). Thermocycling was conducted in a BIO-RAD DNA Engine Dyad Cycler (BIO-RAD). A lowered annealing temperature of 55°C was also used in tandem with a doubled cDNA concentration for KK02, KK04 and KK05 primers when no products could be amplified with their ideal annealing temperatures.

3.2.3.6 – Agarose gel electrophoresis and gel purification of PCR products

Each PCR product (25 μL) obtained was mixed with 5 μL of 6× Ficoll dye and separated on a 1.0% agarose gel (w/v). The gels were ran at 110 V for 1 hr in 1× TAE buffer. HyperLadder I (Invitrogen) (5 μL) was used as the molecular weight ladder. Visualisation was done using ethidium bromide and UV illumination in a Firstlight® UV Illuminator (Ultra-Violet Products Ltd).

The PCR products were excised from the gel and extracted using a PureLink™ Quick Gel Extraction Kit (Invitrogen), as per the manufacturer's instructions. The PCR products were eluted with 50 μL of warm elution buffer to achieve approximately 95% recovery. The samples were then concentrated down to approximately 10 - 15 μL in a SpeedVac.

3.2.3.7 – Ligation of PCR products and bacterial transformation

The concentration of each PCR product insert was initially determined using a ND-1000 nanodrop machine. This was to ensure that the 3:1 ratio of PCR product to plasmid vector was used in each ligation reaction.

Ligation of PCR products was then completed using the pCR®8/GW/TOPO® TA Cloning® Kit (Invitrogen) according to the manufacturer's protocol. Ligation reactions were put into the BIO-RAD DNA Engine Dyad Cycler (BIO-RAD) for a 15 min incubation period at 23°C to allow the ligation reaction to occur. The ligated vector was then transformed into One Shot® Competent *E. coli* cells (Invitrogen) by heat shock treatment according to manufacturer's instructions.

Transformed cells were plated onto Luria Bertani (LB)-agar plates containing spectinomycin (100 μg mL^{-1}) (Sigma) as a selective agent and left to grow overnight at 37°C in an OM11 orbital mixer incubator (Ratek Instruments Pty Ltd.).

3.2.3.8 – Colony PCR screening: Identification of positive pCR®8/GW/TOPO® clones with candidate cDNA inserts

The transformed colonies were picked and screened for the presence of the inserts using a PCR-based assay. Each PCR reaction contained 3.2 μL of 10 mM dNTPs, 1.0 μL of 10 μM forward and reverse primers (refer to Table 3.1), 2.0 μL of 10× Immolase PCR buffer (Bioline Pty Ltd., Alexandria, New South Wales, Australia), 0.75 μL of 50 mM $MgCl_2$ (Bioline), 0.2 μL of Immolase DNA polymerase (Bioline), and 11.85 μL of autoclaved nanopure water.

Individual colonies were selected from plates using sterile 200 μL pipette tips, which were subsequently swirled in the PCR reaction mix. After selection of colonies and addition of bacterial cells to the PCR mix, the pipette tips were placed in 5 mL aliquots of LB media containing spectinomycin (100 μg mL^{-1}) and grown O/N as previously described (section 3.2.3.7). Colony PCR cycle

conditions used were as follows: initial denaturation step at 95°C for 5 min, then 35 cycles of 96°C for 30 sec, T_m°C for 30 sec, 72°C for 1 min 15 sec, followed by a final extension step at 72°C for 10 min. The T_m°C used differed between primer pairs (refer to Table 3.1).

Upon completion of the PCR reactions, each PCR product (20 μL) was separated by agarose gel electrophoresis and visualised as per section 3.2.3.6. After positive colonies were identified from the colony PCR screen, plasmids produced from the positive colonies were purified using a PureLink™ Quick Plasmid Miniprep Kit (Invitrogen) according to the manufacturer's instructions. All steps were performed, including the optional additional wash step. Elution was conducted using 50 μL of warm TE buffer. Glycerol stocks were made for all positive clones by adding 1 mL of bacterial culture to 1 mL of 50% glycerol. The stocks were snap-frozen using liquid nitrogen and stored at -80ºC.

3.2.3.9 – Sequencing reactions for positive pCR®8/GW/TOPO®-candidate insert clones

Sequencing reactions comprised of 2 μL of purified plasmid DNA template, 1.5 μL of Big Dye buffer (Institute of Medical and Veterinary Science – IMVS, Adelaide, South Australia, Australia), 0.5 μL of Big Dye (IMVS) and 1 μL of either 10 μM GW1 forward or GW2 reverse primer (see Table 3.1). Multiple positive colonies were selected for sequencing. Typically three reactions were conducted using the GW1 forward primer and two reactions with the GW2 reverse primer per clone to ensure sufficient coverage of the PCR insert and to allow for sequencing redundancy to identify and remove errors.

Sequencing cycle conditions were as follows; initial denaturation step at 96°C for 30 sec, then 25 cycles of 96°C for 10 sec, 50°C for 5 sec, 60°C for 4 min. This sequencing reaction method was optimised for sequencing with the ABI 3730 DNA Analyser (Applied Biosystems, Scoresby, Victoria, Australia).

3.2.3.10 – Clean-up of sequencing reactions

Sequencing reactions were allowed to equilibrate to room temperature following completion of the PCR cycling. Forty μL of fresh 75% isopropanol (v/v) (Chem-Supply) was added to each sequencing sample and left to incubate for 20 min at room temperature, during which time the DNA precipitated. Samples were then centrifuged for 20 min at maximum speed in a Heraeus Biofuge® Pico centrifuge (Kendro Laboratory Products Ltd., Hong Kong). Supernatant was then carefully discarded and the DNA pellets were washed with 200 μL of fresh 75% isopropanol (v/v).

Samples were centrifuged a second time at maximum speed for 10 min and the supernatant removed with a pipette; being careful not to dislodge the DNA pellets. The tubes were allowed to air-dry in the dark. Samples were submitted to IMVS for capillary separation using the ABI 3730 DNA Analyser (Applied Biosystems).

3.2.3.11 – Contig construction and sequence analysis of pCR®8/GW/TOPO®-candidate insert clones

The sequence data from individual clones were uploaded into the Contig Express program (Informax, VNTI Advance, Version 9.1, USA) and contigs representing the sequence of individual clones were generated. DNA sequence files were

generated using Vector NTI (Informax, VNTI Advance, Version 9.1, USA), and compared using AlignX (Informax). Using AlignX and Contig Express, sequence information from the candidate cDNA clones derived from Chinese Spring were obtained to form consensus sequences for each candidate. The sequences were then used in tBLASTX, BLASTp and BLASTn searches against online databases as per section 3.2.2 to ensure that the correct coding sequences were isolated.

3.2.4 – Isolation and amplification of full length candidate open reading frames (ORFs)

3.2.4.1 – High fidelity amplification and purification of candidate ORFs from meiotic cDNA

High fidelity amplification of the ORFs was conducted using Platinum Taq DNA Polymerase High Fidelity (Invitrogen). Each PCR mixture contained 2.5 µL of 10× High Fidelity Buffer, 3.2 µL of 10 mM dNTPs, 0.75 µL of 50 mM $MgSO_4$, 1 µL each of forward and reverse ORF primer, 0.5 µL of Chinese Spring meiotic cDNA (synthesised in section 3.2.3.4), 0.25 µL of Platinum Taq DNA Polymerase High Fidelity (Invitrogen) (5 units μL^{-1}) and 15.8 µL of sterile nanopure water. The PCR cycle conditions for each set of ORF-specific primers were as follows: initial denaturation step at 95°C for 5 min, then 35 cycles of 96°C for 30 sec, T_m°C for 30 sec, 72°C for 1 min 30 sec, followed by a final extension step at 72°C for 10 min. The T_m°C used differed between primer pairs (refer to Table 3.1). Agarose gel electrophoresis separation and extraction of the PCR products was carried out according to the protocol in section 3.2.3.6.

3.2.4.2 – Production, screening and sequencing of pCR®8/GW/TOPO®-candidate ORF clones

Refer to sections 3.2.3.7 - 3.2.3.10.

3.2.4.3 – Contig construction and sequence analysis of pCR®8/GW/TOPO®-candidate ORF clones

Contig construction and sequence analyses were conducted as per section 3.2.3.11. In addition, nucleotide sequences of all the clones were compared using AlignX to ensure that the ORF clones corresponded to the initial sequencing results from section 3.2.3.11. One positive ORF clone for each candidate was then selected for the construction of their respective heterologous protein expression plasmid (refer to section 3.2.7).

3.2.5 – Quantitative real-time PCR (Q-PCR)

3.2.5.1 – Q-PCR primer design

Primer design was conducted by analysing full-length transcripts obtained from section 3.2.4 using Primer3 (http://frodo.wi.mit.edu/primer3/) and NetPrimer (http://www.premierbiosoft.com/netprimer/index.html). Parameters used for Primer3 were as follows: product sizes – 150-300 bp; Max 3′ stability – 7.0; primer T_m was set to a minimum of 60°C, optimum of 65°C, and a maximum of 70°C; Max self-complementarity – 6.0; and, Max 3′ self-complementarity – 2.0. Primers designed can be viewed in Table 3.1.

3.2.5.2 – Testing of Q-PCR primers

A standard PCR protocol was performed with the Q-PCR primers as per section 3.2.4.1. The PCR cycle conditions for each set of primers were as follows: initial denaturation step at 95°C for 5 min, then 35 cycles of 96°C for 30 sec, 50°C for 30 sec, 72°C for 30 sec, followed by a final extension step at 72°C for 10 min. Agarose gel electrophoresis separation and extraction of the PCR products was carried out according to the protocol in section 3.2.3.6. Extracted PCR products were sent to IMVS for sequencing.

3.2.5.3 – Q-PCR

Q-PCR was conducted in triplicate according to Crismani *et al.* (2006). Amplification of products was completed using gene specific Q-PCR primers (see Table 3.1). The optimal acquisition temperatures for *KK01*, *KK03*, *KK04*, and *KK06* were 83°C, 79°C, 82°C and 85°C respectively. cDNA samples used were total cDNA libraries synthesised from wheat anther RNA in section 3.2.3.4.

Table 3.1 – Comprehensive list of primers used in this study to isolate and characterise candidates identified to be differentially expressed via 2DGE experiments in Chapter 2. The list includes primers required for RACE PCR, nested primers for determining nucleic acid sequences in large PCR products, degenerate primers for attempted amplification of candidates *KK02* and *KK05*, ORF-isolation primers as well as Q-PCR primers.

2DGE Candidate 3′ RACE primers

Primer name	**Sequence (5′ → 3′)**	**T_m (°C)**
KK01 F1	GCCGACGACGTGGGCTCC	59
KK01 F2	AAGCTTCGGAGCTGGTACCAGAATTTGGTA	62
KK01 F3	GCAAATACTCTTGCAACTGGGGAG	57
KK01 F6	GCGATGAATCAGAAGGAGACAAGAACGAGGT	63
KK02 F1	GAAATCGTGGCCGGCAACGAGATGCAA	63
KK02 F2 (degenerate primer)	GANATNGTNGCNGGNAANGANATNCANCGN	63
KK03 F1	ATGAATGTGTGGCGGTTTATCCACACAATCAATGAG	63
KK03 F2	GGCGGCAAGCAAGGGGACGCC	64
KK03 F3	ACCCCTCGTGTCGAGCCTGTG	60

KK04 F1	ATGGGGAGCATGCGGAGCACCTTGTTTCCTAACTTT	67
KK04 F3	AAAGCAATTCTCACCAAGAATTATGAAGTGTTA	57
KK04 F5	TGATGACGGTCTGGCTCCAGATGAAG	61
KK05 F1	CGAACGCTCTTGATTCGCTTCATCTCT	60
KK05 F3	GAAGAAAACAATCAGATCACCTGGACA	57
KK05 F4 (degenerate primer)	CGNACNCTNTTNATNCGNTTNATNTCN	57
KK05 F5 (degenerate primer)	GANGANAANAANCANATNACNTGNACN	56
KK06 F1.1	GCGGAGGGAGCTTTCAAG	52
KK06 F1.2	AAGATGGTGCAGGAGGCG	52
KK06 F1.3	GTGGGTGCAGAGGGTGCC	57
KK06 F4	GACCCACAGACTCAGAAGGAAATGAAGATGGTGCCTTA	67
KK06 F5	AACATAGAAGACGAACTGGTGCAGATGCTCCAGCTC	67
KK06 F7	GTTCAAAAGTTATTATATGGTGTCTTAATTACCGTC	59
KK06 F8 (nested primer)	GGGCTGCCATTCAGGGTGGCATCTTGA	64
KK06 F10	CTCCGTGGAGATGTCAAAGAGCTTCTTCTCCTTGA	68
KK06 F11	CAGCCCGGCGTGCTCATCCAGGTATATGA	69
GeneRacer™ 3′ Primer	GCTGTCAACGATACGCTACGTAACG	76
GeneRacer™ 3′ Nested Primer	CGCTACGTAACGGCATGACAGTG	72
2DGE Candidate 5′ RACE primers		
KK01 R1	GGAGCCCACGTCGTCGGC	59
KK01 R2	TACCAAATTCTGGTACCAGCTCCGAAGCTT	62
KK01 R3	CTCCCCAGTTGCAAGAGTATTTGC	57
KK01 R6	CGCCACGGTAGCTCGCACTGTTCATGTA	64
KK02 R1	TTGCATCTCGTTGCCGGCCACGATTTC	63
KK02 R2 (degenerate primer)	NCGNTGNATNTCNTTNCCNGCNACNATNTC	63
KK03 R1	CTCATTGATTGTGTGGATAAACCGCCACACATTCAT	63
KK03 R2	GGCGTCCCCTTGCTTGCCGCC	64
KK03 R3	CACAGGCTCGACACGAGGGGT	60
KK03 R4	CGCCTCCAACCTTTTCTTCTTCTTCTTCCATA	62
KK03 R5	CTGTCAAGTCATGTCTCTCATTGATTGTGTGGATAAA	62
KK03 R6	CAGCTTCAGTGTGGAACTGCGAGTTCTCA	63
KK04 R1	AAAGTTAGGAAACAAGGTGCTCCGCATGCTCCCCAT	67
KK04 R3	TAACACTTCATAATTCTTGGTGAGAATTGCTTT	57
KK04 R5	TGCTGCTACTTCTTCCCACACCACCCATTAT	63
KK05 R1	AGAGATGAAGCGAATCAAGAGCGTTCG	60
KK05 R3	TGTCCAGGTGATCTGATTGTTTTCTTC	57
KK05 R4 (degenerate primer)	NGANATNAANCGNATNAANAGNGTNCG	57
KK05 R5 (degenerate primer)	NGTNCANGTNATNTGNTTNTTNTCNTC	56

KK06 R1.1	CTTGAAAGCTCCCTCCGC	52
KK06 R1.2	CGCCTCCTGCACCATCTT	52
KK06 R1.3	GGCACCCTCTGCACCCAC	57
KK06 R4	TAAGGCACCATCTTCATTTCCTTCTGAGTCTGTGGGTC	67
KK06 R5	GAGCTGGAGCATCTGCACCAGTTCGTCTTCTATGTT	67
KK06 R7	GACGGTAATTAAGACACCATATAATAACTTTTGAAC	59
KK06 R8 (nested primer)	CCTGACCACTTCCGCAGGGACCTTGTCTTT	66
KK06 R9	CAGCGCGTTCTTGGCGTCCACCTTCTTC	66
KK06 R10	CAGCCATCTCCACCCAGGCATCACCATTT	64
GeneRacer™ 5′ Primer	CGACTGGAGCACGAGGACACTGA	74
GeneRacer™ 5′ Nested Primer	GGACACTGACATGGACTGAAGGAGTA	78
2DGE Candidate ORF isolation primers for *KK01*, *KK03*, and *KK06*		
KK01 ORF1 F1	ATGGACGACGACGCCGGC	60
KK01 ORF1 R1	TCAGATTCTTGGCCTCACCCTACG	
KK03 ORF1 F1	ATGGGCGACTCTGGCGGCTC	60
KK03 ORF1 R1	CTAAACCCTGAGATTGGCTCGCGTC	
KK06 ORF1 F1	ATGGCCATCGGATCTCTCCTCGC	60
KK06 ORF1 R1	CTACATCTTGGCCTCCTTGGGGTCCT	
2DGE Candidate Q-PCR primers for *KK01*, *KK03*, *KK04* and *KK06*		
KK01 QPCR F2	GGCAAGGGCAAGCACAAGGT	62
KK01 QPCR R2	AAGTCCGATGTCTCAAGGGCA	
KK03 QPCR F1	TGCGAGGAGGCTCTGCTGTA	63
KK03 QPCR R1	TGCGAGTTCTCACCGACAAATATCA	
KK04 QPCR F2	TGCCCCTGGAGGTCATCAC	59
KK04 QPCR R2	AGTCTTCCGTATCTGCTCTCTTTG	
KK06 QPCR F1	CGGACCCACAATCATCATCACG	63
KK06 QPCR R1	GCACCTTCAGCATTCTCAATCACC	
Plasmid sequencing primers		
GW1	GTTGCAACAAATTGATGAGCAATGC	55
GW2	GTTGCAACAAATTGATGAGCAATTA	55
T7 forward	TAATACGACTCACTATAGGG	55

3.2.6 – Chromosome location of candidates

3.2.6.1 – Southern blot analyses using nullisomic-tetrasomic membranes

3.2.6.1.1 – Pre-hybridisation of nullisomic-tetrasomic nylon membrane

Southern blot hybridisations were performed to determine the chromosome locations of each candidate within the three respective wheat genomes. The chromosome location of each gene was determined using nylon membranes containing nullisomic-tetrasomic Chinese Spring genomic DNA which had been digested with one of the following enzymes: *Bam*HI, *Dra*I, *Eco*RI, *Eco*RV or *Xba*I. All membranes were courtesy of Margie Pallotta from the Australian Centre for Plant Functional Genomics (Adelaide, South Australia).

Prior to pre-hybridisation, the hybridisation solution (which was also used as the pre-hybridisation solution) was made according to the following recipe: for 100 mL of hybridisation solution, a 5 mL volume of nanopure water was added to 30 mL of 5× HSB, 30 mL of Denhardt's III reagent, 30 mL of 25% dextran sulphate (v/v) and 5 mL of denatured salmon sperm (5 mg mL^{-1}). All components of the hybridisation solution were mixed together and heated at 65°C for 5 min.

Membranes were rinsed twice for 5 min with 2× SSC (0.3 M NaCl, 0.03 M tri-sodium citrate) and then rinsed briefly with reverse osmosis (RO) water. Each membrane was then placed in a hybridisation bottle, to which 10 mL of hybridisation buffer was added. The membranes were incubated overnight at 65°C with constant rotation in a hybridisation oven. The following day the solution was discarded and replaced with another 10 mL of hybridisation solution.

3.2.6.1.2 – Probe labelling and hybridisation

The Q-PCR fragment clone of each candidate from section 3.2.5 was used as template for the synthesis of labelled probes. In separate reactions, 3 µL (~150 ng) of each candidate probe with 1 µL (from a 10 µM stock) of each of its respective forward and reverse primers were combined with 3 µL of 10 µM random 9-mer primer mix and denatured at 100°C for 5 min, followed immediately by 5 min incubation on ice to maintain the DNA in single-stranded conformation. 12.5 µL of 2× Oligo-buffer was then added, in addition to 1 µL of Klenow polymerase (Roche) and 4 µL of radioactive ^{32}P (µ-dCTP, Amersham) (10 mCi mL^{-1}). The synthesis reaction was incubated at 37°C for approximately 60 min.

The probes were then made up to a total of 50 µL and purified using the illustra™ ProbeQuant™ G-50 Micro Column kit (GE Healthcare). The eluted probes were denatured at 100°C for 5 min, and subsequently incubated on ice for 5 min. The probes were then added to the hybridisation bottles which were incubated overnight at 65°C with constant rotation.

3.2.6.1.3 – Membrane washes

Excess labelled probe and hybridisation buffer were decanted from the hybridisation bottles. Membranes were washed at 65°C with 40 mL of the following solutions for 20 min until an appropriate level of radioactivity (1 to 2 counts per second) was achieved: wash 1 - 2× SSC, 0.1% SDS (w/v); wash 2 - 1× SSC, 0.1% SDS (w/v); wash 3 - 0.5× SSC, 0.1% SDS (w/v); wash 4 - 0.2× SSC, 0.1% SDS (w/v) and wash 5 - 0.1×SSC, 0.1% SDS (w/v).

3.2.6.1.4 – Autoradiography

The membranes were sealed in plastic sleeves and placed into X-ray cassettes along with X-ray film (Fuji HR-G-30 and Kodak Biomax were commonly used). The cassettes were stored at –80°C for 5 to 10 days depending on the intensity of the radioactive signal. The X-ray films were developed using an AGFA CP1000 Developer.

3.2.6.2 – Southern blot analyses using genotype membranes

3.2.6.2.1 – Genomic DNA extraction

To determine whether the KK06 candidate gene was located within the *ph1b* deletion region, Southern blots utilising genomic DNA extracted from Chinese Spring and the *ph1b* mutant were performed. 10 cm pieces of young healthy leaf tissue were collected and put into 2 mL tubes, snap-frozen and vortex-ground with ball-bearings. Six hundred µL of DNA Extraction Buffer (1% sarkosyl, 100 mM Tris-HCl, 100 mM NaCl, 10 mM EDTA, pH8.0), pre-warmed to 65°C, was added to each sample. Samples were left to incubate at 65°C for 30 min with occasional mixing before 600 µL of phenol/chloroform/iso-amyl alcohol (25:24:1) was added. Samples were mixed by continuously inverting the tube for 10 s then left to incubate on a rotary wheel for 10 min followed by 10 min of centrifugation at maximum speed in a micro-centrifuge. The upper aqueous phase was then transferred to a fresh tube and the phenol extraction, mixing, and centrifugation steps were repeated. Upon completion, the upper aqueous layer was transferred to a fresh tube. 60 µL of 3 M sodium acetate (pH4.8) and 600 µL of isopropanol were added and the samples were gently mixed by inversion. Samples were incubated at room temperature for 2 min then centrifuged for 10 min at maximum

speed in a microcentrifuge. The supernatant was removed and the DNA pellet washed with 1 mL of 70% ethanol. Samples were centrifuged for 5 min and the supernatant discarded before being dried at 37°C. DNA pellets were resuspended by first adding 100 µL of R40 buffer (TE buffer containing 40 µg/mL RNAse A) and incubating at room temperature for 10 min followed by a 5 min incubation at 37°C. Samples were then left to dissolve overnight at 4°C. The following day, samples were frozen and thawed twice. Concentrations of the samples were then determined using an ND-1000 nanodrop machine (NanoDrop Technologies Inc.).

Agarose gel electrophoresis was performed using 1 µg of DNA per sample as outlined in section 3.2.3.3 to ensure that the extracted genomic DNA was of sufficient quality.

3.2.6.2.2 – Enzymatic digestion of genomic DNA

Screening of the KK06 ORF with Vector NTI (informax) showed that it did not contain restriction sites for *Bam*HI and *Dra*I. 30 µg of each genomic DNA sample was digested with *Bam*HI (New England Biolabs, Genesearch Pty. Ltd., Queensland, Australia) and *Dra*I (NEB) according to the manufacturer's protocol.

3.2.6.2.3 – Capillary transfer of digested genomic DNA

Thirty µg of digested genomic DNA from each genotype was separated on a 1 % agarose gel (w/v). The DNA was then transferred onto Hybond-N^+ nylon membranes (Amersham) using a standard capillary transfer blotting technique with 0.4 M NaOH.

3.2.6.2.4 – Southern blot procedure

Pre-hybridisation, probe-labelling and hybridisation, membrane washes and autoradiography were performed as per section 3.2.6.1.

3.2.6.3 – *In silico* mapping of the *KK06* gene candidate

BLASTn searches of the TIGR Rice Genome Database (http://blast.jcvi.org/euk-blast/index.cgi?project=osa1, Rice Genome Annotation Build 6.1; accessed 4th April 2011) was used to match the *KK06* candidate coding sequence with its equivalent rice gene (Rice locus identifier: LOC_Os09g31486). This EST was then mapped in relation to the *Hyp6* gene (Rice locus identifier: LOC_Os09g30410), previously mapped by Griffiths and colleagues (2006) as the last gene definitively located within the *ph1b* deletion region.

3.2.7 – Preparation of candidate ORF clones for heterologous protein expression

3.2.7.1 – Construction of pDEST17-candidate ORF clones

Upon confirming that each candidate ORF sequence was correct and orientated appropriately within its respective pCR®8/GW/TOPO® vector (section 3.2.4.4), an LR recombination reaction was conducted to transfer the candidate ORF insert to the pDEST17 expression plasmid (Invitrogen). The pDEST17 plasmid vector encodes a 6× histidine (6×His) tag at the 5′ end of the insert entry site which enabled protein purification via nickel affinity chromatography in section 3.2.9.2.

The concentration of the purified miniprep-ed pCR®8/GW/TOPO®-candidate ORF plasmid samples selected in section 3.2.4.4 was determined using

a ND-1000 nanodrop machine (NanoDrop Technologies Inc.). The appropriate volume containing 150 ng of pCR®8/GW/TOPO®-candidate ORF clone plasmid DNA was then used in the LR recombination reaction. pDEST17 (1 μL at 150 ng μL^{-1}) was added to the sample with the appropriate volume of TE buffer (pH 8.0), to give the LR recombination reaction mix a final volume of 8 μL. 2 μL of LR Clonase™ II enzyme mix (Invitrogen) was then added to the reaction mix. The reaction mix was incubated in the BIO-RAD DNA Engine Dyad Cycler (BIO-RAD) for 1 hour at 25ºC before adding 1 μL of 2 μg μL^{-1} Proteinase K solution (Invitrogen). Samples were then incubated for a further 10 min at 37ºC to terminate the ligation reaction.

Transformation with One Shot® Competent *E. coli* was performed as per section 3.2.3.7, except that the selective antibiotic used for pDEST17 was ampicillin. Colony PCR screening and mini-prep procedures were conducted as per section 3.2.3.8. Directional sequencing of the positive pDEST17-candidate ORF clones was performed as outlined in section 3.2.3.9 except that the T7 forward primer was used. Chromatogram files containing the sequence data from individual clones were uploaded into the Contig Express program (Informax), and contigs representing the sequence of individual clones were generated. Using this information, pDEST17-candidate ORF clones with their respective candidate ORF inserts cloned in-frame and in the correct orientation were identified for heterologous protein expression.

3.2.8 – Protein production of candidates

3.2.8.1 – Transformation of BL21A1 *E. coli* cells with the pDEST17-candidate ORF clones

BL21-A1 *E. coli* cell lines containing each of the pDEST17-candidate ORF clones were produced by transforming the cells with one of each selected pDEST17-candidate ORF clones identified in section 3.2.7. The transformation protocol used was as per section 3.2.3.7; with the exception that ampicillin (100 $\mu g\ mL^{-1}$) was used as the selective antibiotic.

3.2.8.2 – Culturing of pDEST17-candidate ORF-transformed BL21-A1 cell lines

One transformed colony was picked from each of the plates made during the transformation procedure of section 3.2.8.1 and placed into a culture tube containing 10 mL of LB/carbenicillin (50 $\mu g\ mL^{-1}$). This starter culture was left to grow overnight at 37°C in an orbital mixer incubator (Ratek). Ten 25 mL secondary cultures were then set up using 24 mL of LB/carbenicillin (50 $\mu g\ mL^{-1}$) and 1 mL of each starter culture. The secondary cultures were grown overnight at 37°C in an orbital mixer incubator (Ratek) before each was used to inoculate 225 mL of LB/carbenicillin (50 $\mu g\ mL^{-1}$) in 1 L conical flasks. These 250 mL cultures were then grown at 37°C in an orbital mixer incubator (Ratek) until the cell density was OD_{600} = 0.4. The OD_{600} readings were obtained using a Metertech UV/Vis SP8001 spectrophotometer.

3.2.8.3 – Induction of protein production and subsequent cell harvesting

Upon reaching the required cell density, the cultures were cooled in ice baths to 16°C. Protein production was induced by the addition of L-(+)-arabinose (Sigma-Aldrich) and 0.1 M IPTG (Sigma-Aldrich) to a final concentration of 0.1% (w/v) and 1 mM respectively. For the non-induced control sample, glucose was added to a final concentration of 0.5% (w/v) to repress any leaky expression. The cultures were left to grow at 16°C in an orbital incubator for 6 hr prior to cell harvesting. Thereafter, the cell cultures were spun at 7000 rpm for 10 min at 16°C to pellet down the cells, with the supernatant discarded. The wet weight of the cell pellets were recorded and the pellets then snap-frozen using liquid nitrogen and stored at -80^0C until protein extraction and purification.

3.2.9 – Denatured extraction, purification and visualisation of proteins

3.2.9.1 – Cell lysis and protein extraction

Prior to starting this procedure, a 5 L stock solution containing 100 mM NaH_2PO_4 (AnalaR), 10 mM Tris-Cl (Sigma), and 8 M urea (Bio-Rad) was made. The stock solution was divided into five 1 L aliquots and adjusted to different pH values using either concentrated NaOH (AnalaR) or HCl (AnalaR) to make five solutions (Lysis buffer B, pH 8.0; Wash buffer C1, pH 7.1; Wash buffer C2, pH 6.3; Wash buffer D, pH 5.9; and Elution buffer E, pH 4.5).

The cell pellets from section 3.2.8.3 were thawed on ice for 15 min then resuspended in Lysis buffer B at a concentration of 1 g wet weight per 5 mL. The cells were lysed by rotation on a rotary wheel for 1 hr at 4°C with gentle shaking every 15 min. The cell lysate was then sonicated six times in 10 sec bursts using power output 4 (25 W) (Branson Sonicator B-12, Danbury, Connecticut, USA) to

ensure total cell lysis. The lysate was centrifuged at 8950 rpm for 15 min at room temperature to pellet down cell debris and clear the lysate solution. The cleared lysate was collected while the pelleted debris was discarded.

3.2.9.2 – Purification of 6×His-tagged candidate proteins via nickel-affinity chromatography

Prior to starting this procedure, a 20 μL aliquot of the cleared cell lysate from section 3.2.9.1 was taken as the Cell Lysate (CL) sample. Nickel-nitrilotriacetic acid (Ni-NTA) agarose beads (8 mL) (QIAGEN Pty Ltd., Clifton Hill, Victoria, Australia), which had been washed and equilibrated with 15 mL of Lysis buffer B, was added to the cleared cell lysate sample and allowed to bind the 6×His-tagged candidate protein. The cell lysate-Ni-NTA bead slurry was incubated on a rotary wheel in a 4°C cold-room for 2 hr to increase binding efficiency. The cell-lysate-Ni-NTA bead slurry was spun down at 2000 rpm for 2 min to pull down the beads. The supernatant was aspirated and kept as the Flow-Through (FT) sample.

For the washing procedure, 15 mL of Wash buffer C1 was added to the bead slurry, which was then vortexed briefly at low speed, and left to incubate on ice for 3 min. The beads were then spun down at 2000 rpm for 2 min and the supernatant aspirated and kept as Wash 1 (W1). The wash procedure was repeated using 15 mL of Wash buffer C2, and 15 mL of Wash Buffer D. The supernatant for each wash was aspirated and kept as Wash 2 (W2) and Wash 3 (W3) respectively. Elution of the 6×His-tagged candidate proteins was then conducted thrice, each time using 1 mL of Elution buffer E using the same procedure as the washing steps. The eluted candidate protein samples (E1 - E3) were snap-frozen, along with all other samples, and stored at -80°C until SDS-PAGE was conducted.

3.2.9.3 – SDS-PAGE for separation of proteins by molecular weight

A total of 6.5 μL of each sample collected in section 3.2.9.2 (CL, FT, W1, W2, W3, E1, E2, E3) was added with 2.5 μL of NuPAGE® LDS Sample Buffer (4×) (Invitrogen), and 1 μL of NuPAGE® Reducing Agent (10×) (Invitrogen). The samples were heated at 70°C for 10 min before being loaded into 15-well NuPAGE® Novex® 4-12% Bis(2-hydroxyethyl)-imino-tris(hydroxymethyl)-methane (Bis-Tris) mini gels (Invitrogen). Bio-Rad Precision Plus Dual Colour Protein Standard (6 μL) was also loaded into the gels as the molecular weight marker. Electrophoresis was performed according to the manufacturer's instructions. Gel fixing, sensitisation, and staining with only coomassie blue was conducted as outlined in section 2.2.4.

3.2.10 – Native extraction, purification and visualisation of proteins

3.2.10.1 – Cell lysis and protein extraction

Native protein extractions were done to isolate the candidate proteins in native form for functional analyses such as DNA-binding assays. Prior to starting this procedure, a 3 L stock solution containing 50 mM NaH_2PO_4 (AnalaR) and 300 mM NaCl (Sigma) was made and adjusted to pH 8.0 with concentrated NaOH (AnalaR). The stock solution was divided into three 1 L aliquots and each was added with differing amounts of imidazole (sigma) to make three solutions (Lysis buffer, 10 mM imidazole; Wash buffer, 20 mM imidazole; and Elution buffer, 250 mM imidazole).

The cell pellets from section 3.2.8.3 were thawed on ice for 15 min and resuspended in Lysis buffer at a concentration of 1 g wet weight per 5 mL. Lysozyme was added to a final concentration of 1 mg/mL and the cell suspensions left to incubate on ice for 30 min. The cell suspensions were then sonicated six times in 10 s bursts using power output 4 (25 W) (Branson Sonicator B-12) to ensure total cell lysis. The lysates were centrifuged at 8950 rpm for 15 min at 4°C to pellet down cell debris and clear the lysate solution. The cleared lysate was collected while the debris pellet was discarded.

3.2.10.2 – Purification of 6×His-tagged candidate proteins via nickel-affinity chromatography

Prior to starting this procedure, a 20 μL aliquot of the cleared cell lysate from section 3.2.10.1 was taken as the Cell Lysate (CL) sample. Protein purification was then conducted using Ni-NTA spin columns (QIAGEN) which had been equilibrated with 600 μL of Lysis buffer. The cleared cell lysate was spun through the Ni-NTA spin columns at 4°C and the flow-through (FT) kept for running on SDS-PAGE. The columns were washed with 9× column volumes of Wash buffer. The first, fourth and seventh washes were kept (W1, W2, and W3 respectively). Elution of the 6×His-tagged candidate proteins was then conducted thrice, each time using 100 μL of Elution buffer. The eluted candidate protein samples (E1 - E3) were snap-frozen, along with all the other samples, and kept at -80^{0}C until required.

3.2.10.3 – SDS-PAGE for separation of proteins by molecular weight

6.5 μL of each sample (CL, FT, W1, W2, W3, E1, E2, and E3) collected in section 3.2.10.2 was separated by SDS-PAGE as outlined in section 3.2.9.3.

3.2.11 – Functional analyses of candidates

3.2.11.1 – Competitive DNA-binding assay

Recombinant candidate proteins extracted under native conditions were quantified using the Bradford assay (Bradford 1976). The competitive DNA binding assay was conducted as described by Pezza *et al.* (2006) with modifications as per Khoo *et al.* (2008). DNA binding ability was tested with ΦX174 circular single-stranded DNA (ssDNA) (virion) (30 μM per nucleotide) (NEB) and ΦX174 linear double-stranded DNA (dsDNA) (RFI form *PstI*-digested) (15 μM per base pair) (NEB).

3.3 – Results

3.3.1 – Mass-peptide analyses of 2DGE protein spots

The six protein spots identified to be differentially expressed between stages and genotypes from Chapter 2 were analysed using Ion-Trap Liquid Chromatography-Electrospray Ionisation tandem mass spectrometry (MS/MS). The spots were named KK01 to KK06, with the numbers corresponding to the numbering given in Chapter 2 (refer to Figure 2.14). The peptides identified were matched against the complete rice proteome (obtained from the Rice Genome Annotation, Build 6.1) using the Bioworks 3.3 Sequest algorithm (refer to Appendix B1 for the summary reports of the MS/MS analyses). Searches of the rice proteome database

using the sequences obtained from the identified peptides using the BioWorks 3.3 program showed that none of the peptides obtained from each spot could be matched to the same rice database hit. Consequently, identified peptide sequences of each spot were used in BLASTp searches against the NCBI database (with *Triticeae* and *Oryza* filters selected), Rice Genome Database, and the *Brachypodium distachyon* Phytozome BLAST database to identify candidates within these databases that were matched by multiple peptides identified from the same spot. This strategy did not yield any successful results as no single database hit that was found matched more than one of the identified peptides from each spot. Thus a secondary round of BLAST analyses was carried out to identify the best possible hits (limited to *Triticeae* sequences, and where not available, to *Oryza sativa* or *Hordeum vulgare*) for each of the identified peptides (Table 3.2). In addition, the protein matches obtained from the MS/MS data of peptides from the KK03 and KK04 protein spots (that had mapped to coordinates very close to one another in the 2D gels) (Figure 2.14 A) were most likely the result of different protein species.

Table 3.2 – Identification of peptides obtained from 2DGE protein spots that were present at different levels isolated in Chapter 2. Although multiple peptides were obtained from each spot examined, mass peptide identification of the trypsin-digested peptides failed to identify any protein hits within the Rice Genome Database (Build 6.1) that matched at least two peptides obtained from any of the six spots examined. BLAST analysis was used to identify the best possible hit for each peptide. The molecular weight and pI of each hit was then predicted using the Expasy Compute pI/MW tool.

Spot	Identified peptide	Protein identified through Bioworks match in Rice Genome Database	Best NCBI database hit (Accession number)	Observed MW of spot vs. predicted MW of best database hit (kD)	Observed pI of spot vs. predicted pI of best database hit
KK01	KYKLKVVAADDVGSSSSQDGKG	retrotransposon protein, putative, Ty3-gypsy subclass	*Aegilops tauschii* putative Bowman Birk trypsin inhibitor (GenBank: AAS48158)	20-25 / 18.907	pH 6.0-6.5 / pH 4.82
	KLHWWYQNLV/ KLRSWYQNLV	Histone Acetyltransferase HAC5 protein	*Oryza sativa* hypothetical protein OsJ_01086 (GenBank: EEE54225)	20-25 / 142.919	pH 6.0-6.5 / pH 6.81
	AITLATGA/ ANTLATGA	NFD4 protein, putative	*Oryza sativa* hypothetical protein OsJ_34284 (GenBank: EEE52292)	20-25 / 64.599	pH 6.0-6.5 / pH 5.91
KK02	EIVAGNEMQR	Retrotransposon protein	*Oryza sativa* hypothetical protein (GenBank: AAP54312)	20-25 / 76.515	pH 6.0-6.5 / pH 5.69
KK03	MNVWRFIHTINE	Pollen-specific SF21 protein, putative	*Oryza sativa* putative SF21C1 protein (GenBank: BAD53753)	20-25 / 38.571	pH 6.0-6.5 / pH 5.85
	GGKQGDA	Retrotransposon protein	*Oryza sativa* hypothetical protein (GenBank: AAU10733)	20-25 / 54.799	pH 6.0-6.5 / pH 9.54
	TPRVEPV	Retrotransposon Ty3-gypsy subclass protein	*Oryza sativa* OSJNBa0059H15.13 retrotransposon gag protein (GenBank: CAD39362)	20-25 / 83.468	pH 6.0-6.5 / pH 9.54

KK04	ENKGSMRSTLFPNF/ EIMGSMRSTLFPNF/ EDKGSMRSTLFPNF	Hexose transporter family protein, putative	*Hordeum vulgare* hexose transporter, putative (GenBANK: CAD58958)	20-25 / 79.340	pH 6.0-6.5 / pH 4.86
	KAILTKNYEVL	Chromatin remodelling protein	*Oryza sativa* chromatin-remodeling factor CHD3 (GenBank: AAL47203)	20-25 / 128.861	pH 6.0-6.5 / pH 6.44
KK05	RTLLIRFIMSSR/ RTLLIRFISLSR	*Os*Man04-beta-endo-mannase	*Oryza sativa* hypothetical protein OsI_14157 (GenBank: EAY92423)	60-100 / 50.849	pH 7.0-7.5 / pH 5.79
	VAAKNTEP	Putative protein kinase	*Oryza sativa* hypothetical protein OsI_18275 (GenBank: EEC78436)	60-100 / 49.669	pH 7.0-7.5 / pH 8.59
	EENNQITWT	RNA polymerase IV largest subunit, putative	*Oryza sativa* hypothetical protein OsJ_04874 (GenBank: EEE56061)	60-100 / 169.939	pH 7.0-7.5 / pH 7.93
KK06	FAFTEILD	NIN protein, putative	*Oryza sativa* hypothetical protein OsJ_15319 (GenBank: EEE61255)	50-100 / 99.418	pH 7.5-8.0 / pH 5.69
	SVGAEGAFKMVQEAWTVLS	DnaJ/Hsp40 domain containing protein	*Oryza sativa* hypothetical protein OsI_11948 (GenBank: EEC75429)	50-100 / 83.388	pH 7.5-8.0 / pH 8.10
	DPQTQKEMKMVPYK	DnaK/Hsp70 family protein, putative	*Oryza sativa* hypothetical protein OsJ_29849 (GenBank: EEE69961)	50-100 / 72.798	pH 7.5-8.0 / pH 5.66
	NIEDELVQMLQLA	Inactive receptor kinase At2g26730 precursor, putative	*Oryza sativa* hypothetical protein OsI_11426 (GenBank: EEC75189)	50-100 / 54.237	pH 7.5-8.0 / 7.11
	VQKLLYGVLITV	Retrotransposon Ty3-gypsy subclass protein	*Oryza sativa* retrotransposon Ty3-gypsy subclass protein (GenBank: CAH66737)	50-100 / 85.441	pH 7.5-8.0 / 8.29

3.3.2 – Isolation and characterisation of the candidate gene transcripts

Based on the BLAST data obtained, 5′ and 3′ RACE primers were used to isolate and identify possible candidate genes that encode the six protein products from which these peptides were obtained.

Full-length candidate coding transcripts of *KK01* (1281 bp ORF), *KK03* (1047 bp ORF), and *KK06* (2055 bp ORF) were successfully isolated using multiple rounds of 5′ and 3′ RACE (Figure 3.1 A, B, and D respectively) (refer to Appendix B2 Figures 1, 2 and 4 for nucleotide sequences). In contrast, RACE reactions using primers for KK02 and KK05 failed to amplify specific PCR products. Degenerate primers (with random nucleotides inserted in the wobble position of each codon) were designed in an effort to overcome this problem but this approach did not yield any specific products either (Figure 3.2). No products were obtained even when RACE PCR reactions were performed with a lowered annealing temperature of 55°C and increased amounts of template cDNA. In contrast to KK02 and KK05, a portion of the *KK04* coding sequence, 1598 bp in length, was successfully isolated using the KK04 F1 primer (Figure 3.1 C, and Appendix B2 Figure 3). However subsequent primers designed to isolate the remainder of the coding sequence failed to amplify any products, even at lowered annealing temperatures. The isolation of two different candidate coding sequences for KK03 and KK04 further affirmed the MS/MS data that the KK03 and KK04 spots seen in the 2D gels were indeed from different protein species.

Matching the predicted proteins encoded by the candidate genes for KK01, KK03, and KK06 as well as the partial peptide encoded by the partial KK04 candidate showed that only one peptide identified by the MS/MS data could be found within each of their predicted amino acid sequences (Figure 3.3).

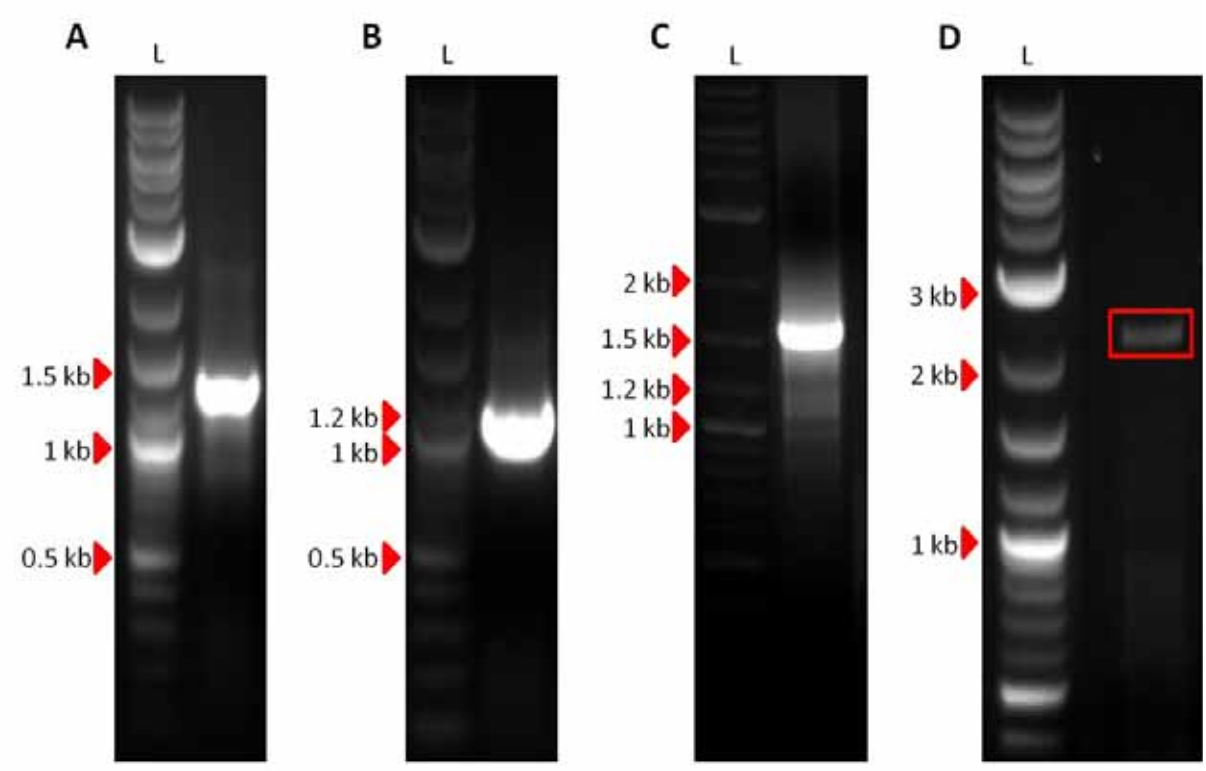


Figure 3.1 – Successful isolation of the full-length ORFs of *KK01, KK03, KK06* and a partial product for *KK04*. A) Isolation of the full-length 1281 bp *KK01* ORF. B) Isolation of the full-length 1047 bp *KK03* ORF. C) Isolation of a partial *KK04* coding sequence 1598 bp in length. D) Isolation of the full-length 2055 bp *KK06* ORF.

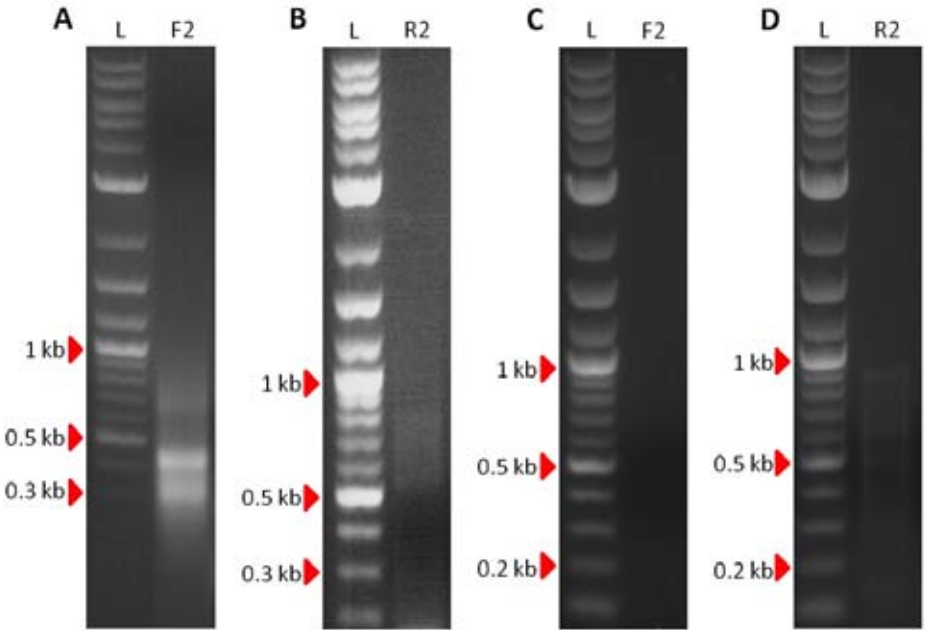


Figure 3.2 – Isolation of *KK02* and *KK05* candidates were unsuccessful even when degenerate primers were used. A) 3′ RACE using *KK02* F2 degenerate primer. B) 5′ RACE using *KK02* R2 degenerate primer. C) 3′ RACE using *KK05* F2 degenerate primer. D) 5′ RACE using *KK05* R2 degenerate primer.

A

MDDDAGDASPPPGPAPLAATGAPSLRDMAASPTSSRSVTETVNGSHRFVIQGYSLAKGMGVGKHIASETFTVGGYQWAIYFYPDGKNPEDNSAYVSVFIA
LASEGTDVRALFELTLQDQSGKGKHKVHSHFDRSLESGPYTLKYRGSMWGYKRFFRRTALETSDFLKDDCLRINCTVGVVVSTIDYSRPHSIQVPDSDIG
YHFGSLLDSNEGVDVILNVSGERFHAHKLVLAARSTVFRSKLFDDESEGDKNEVNESEDLKEIVLDDLEPKVFKAMLHFIYRDTLVDDYELDASSSMGSI
FDTLAAKLLAAADKYDLGRLRLLCESYLCKGISVASVANTLATGESHHAMELKAVCLKFAAENLSAVIRTDGFDYLKDNCPSLQSEILRTVAGCEEPCSS
GGKSQSVWGQLSDGGDTSGRRVRPRI*

B

MGDSGGSVVSIDVERISFGGKEHHIQTKHGPVSVAVYGDHDKHALITYPDVALNHMSCFQGLLFCPEAASLLLHNFCIYHISPPGHELGATPILPNSPVA
SVDELADQVAEVLDFFGLSSVMCLGVSAGAYILTLFATKYRERVLGLILVSPLCRTPSWTEWFYNKVMSNLLYYYGMCDVVKDCLLQRYFGKRVRGGSAV
PESDIMQACRSFLDQRQSMNVWRFIHTINERHDLTESLKQLQCRTLIFVGENSQFHTEAVHMTAKLDKRYSALVEVQACGSVVTEEQPHAMLIPMEYFLM
GYGLFRPSHVSSSPRSPLNPFCISPELLSPESMGVKLKPIKTRANLRV*

C

MGSMRSTLFPNFGSMFSVAEQQQAKADWDAESHRDDEDYASDHGADDIEDSLNSPLISRQATSVEGKEIAAPHGSIMGGVGRSSSMQGGEAVSSMGIGGG
WQLAWKWTEREGADGQKEGGFQRIYLHEEGVSGDRRGSILSMPGGDIPPGGEYIQAAALVSQPALYSKDLIEQQLAGPAMVHPSEAVAKGTKWAELFEPG
VKHALFVGIGLQILQQFAGINGVLYYTPQILEQAGVGVLLSNIGLSSSSASILISALTTLLMLPSIGIAMRLMDRSGRRFLLLSTIPVLIVALAVLVLVN
VLDVGTMVHAALSTISVIVYFCFFVMGFGPIPNILCAEIFPTSVRGICIAICALTFWIGDIIVTYTLPVMLNAIGLAGVFGIYAIVCILAFVFVYMKVPE
TKGMPLEVITEFFSVGAKQGKEATD*

D

MAIGSLLASRLARSGHALATRAIAQAPRTHNHHHATSPLLSRLGAVARAFSSRPAAADVIGIDLGTTNSCVSVMEGKTPRVIENAEGARTTPSIVATNSK
GEILIGITASRQAVTNAENTVRGSKRLIGRAFDDPQTQKEMKMVPYKIVRGTNGDAWVEMAGKSYSPSQIGAFVLTKMKETAEAYLGKSVSKAVITVPAY
FNDAQRQATKDAGRIAGLDVMRIINEPTAAALSYGMNNKEGLIAVFDLGGGTFDVSILEISNGVFEVKATNGDTFLGGEDFDATLLNYLVSEYKNSDNID
LSKDKLALQRLREAAEKAKVELSSTPQTEINLPFITADASGAKHFNITLTRSKFESLVGNLIERTRIPCTNCPKDAGVSAKEIDEVLLVGGMTRVPKVQD
IVSQIFGKSPSKGVNPDEAVAMGAAIQGGILRGDVKELLLLDVTPLSLGIETLGGIFTRLINRNTTIPTKKSQTFSTAADNQTQVGIKVLQGEREMATDN
KLLGEFQLEGIPPAPRGMPQIEVTFDIDANGIVKVSAKDKSTGKEQDITIKSSGGLSDSDIEKMVKEAELNSQRDQERKSLIDLRNSADTTIYSIEKSVS
EYKDKVPAEVVTEIQSAVSDLRAAMAGDDSDAIKQKLEAANKAVSKIGQHMQGGGGAAGGDSGSSGGGDQTPEAEYQDPKEAKM*

Figure 3.3 – Location of peptides identified through MS/MS within the predicted amino acid sequences of candidate transcripts for KK01, KK03, KK04 and KK06. In each case, only one peptide identified from the MS/MS data was able to be matched to the predicted proteins encoded by the candidate transcripts identified for KK01, KK03, KK04 and KK06. These peptides are represented as bold underlined text within the sequences. A) Predicted amino acid sequence of the *KK01* candidate transcript. B) Predicted amino acid sequence of the *KK03* candidate transcript. C) Predicted amino acid sequence of the partial *KK04* transcript. D) Predicted amino acid sequence of the *KK06* transcript.

3.3.3 – Characterisation of the candidate protein products

The full-length ORF of the *KK01* candidate encoded a protein with a predicted MW of 46.267 kD and a pI of 5.09. BLASTp analyses of this protein revealed that it shared a high level of sequence conservation (Identities = 90%, Positives = 95%, *E*-value = 0.0) with the rice *Bric-a-Brac, Tramtrack, Broad Complex (BTB) domain with Meprin and TRAF Homology (MATH) domain* speckle-type POZ protein (Rice locus identifier: LOC_Os07g07270.1). Consequently, KK01 possesses a complete MATH domain (spanning amino acids 44 to 177) and a partial BTB domain (spanning amino acids 203 to 283) (Figure 3.4 A).

Candidate *KK03* encoded a protein with a predicted MW of 38.860 kD and a pI of 7.04. BLASTp analyses of this protein revealed that it is very similar to the rice pollen-specific protein SF21 (Rice locus identifier: LOC_Os06g36740) (Identities = 88%, Positives = 95%, *E*-value = 9.0e-173). An esterase lipase domain spans amino acid positions 22 to 256 of the KK03 candidate protein (Figure 3.4 B).

Although only a portion of the *KK04* candidate coding sequence was successfully isolated, a partial ORF was identified within the *KK04* transcript. The resulting peptide was 425 aa in length and shared high levels of sequence similarities with a barley hexose transporter (GenBank accession: CAD58958) (Identities = 98%, Positives = 99%, *E*-value = 0.0) and a putative rice transporter family protein (Rice locus identifier: LOC_Os10g0539900) (Identities = 92%, Positives = 98%, *E*-value = 0.0). Based on comparisons with the barley and rice hexose transporter sequences which are both 2220 bp in length, 927 bp of sequence in the 5′ direction of the isolated *KK04* coding sequence is required to form the full-length *KK04* ORF. Even so, a Major Facilitator Superfamily (MFS) domain was detected within the isolated sequence and spans position 205 to 396 of the peptide (Figure 3.4 C).

The full-length ORF of *KK06* encoded a protein 683 aa in length with a predicted MW of 72.697 kD and a pI of 6.04. It shared high levels of sequence similarity with a rice Heat Shock Protein 70 hypothetical protein (GenBank accession: NP_001175918) (Identities = 91%, Positives = 95%, *E*-value = 0.0) and a maize Heat Shock Protein 70 (GenBank accession: NP_001152601) (Identities = 89%, Positives = 93%, *E*-value = 0.0). Inter-kingdom comparisons revealed that the *KK06* candidate shares relatively high levels of sequence similarity with the HSP70-2 variants found in mammalian species such as humans (GenBank accession: NP_068814.2) (Identities = 50%, Positives = 66%, *E*-value = 1e-159) and mice (GenBank accession: NP_001002012) (Identities = 50%, Positives = 66%, *E*-value = 1e-159). In addition, KK06 possesses a complete dnaK domain that spans amino acids position 69 to 683 (Figure 3.4 D).

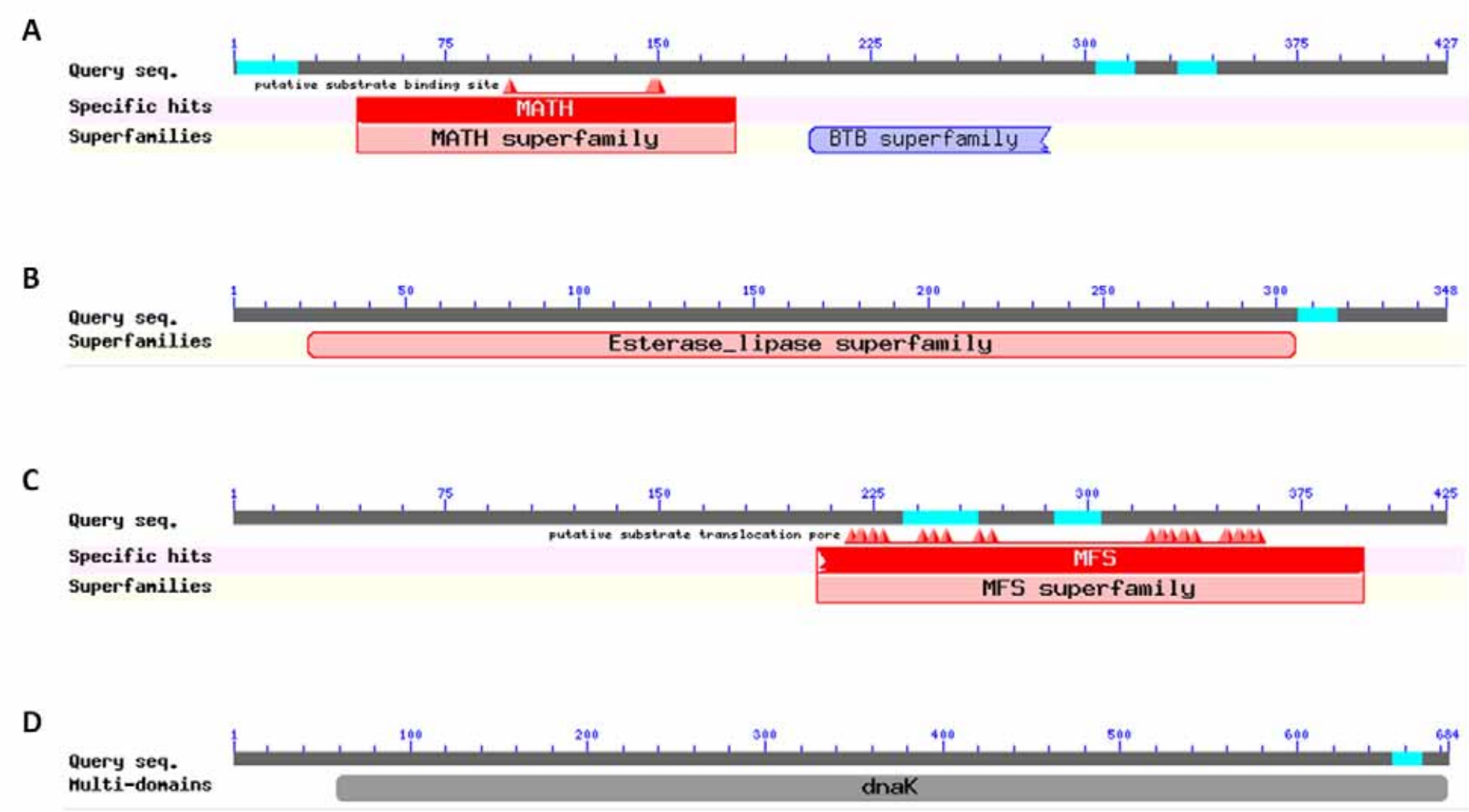


Figure 3.4 – Conserved domain analyses of the candidate proteins. A) KK01 full-length protein. B) KK03 full-length protein. C) KK04 partial peptide. D) KK06 full-length protein.

3.3.4 – Chromosome location of the candidate genes

Southern blot results indicated that the *KK01* candidate resides on chromosome group 2, with at least one copy on each of the A, B and D genomes (Figure 3.5 A). The absence of a band in the lane nullisomic for 5B-tetra-compensated with 5D was due to chromosome 2D also being deleted within that nulli5B-tetra5D mutant sample (post-communication with Ms Margie Pallotta). *KK03* is located on chromosome group 7 with copies on the A, B and D genomes (Figure 3.5 B), with the faint banding pattern being a result of the small probe size used. *KK04* and *KK06* were identified on chromosome groups 1 and 5 respectively, both with one copy on each of the A, B and D genomes (Figure 3.5 C and D respectively). The faint banding pattern observed in the lane nullisomic for 1D-tetra-compensated with 1B of the *KK06* Southern blot result was caused by uneven loading of the DNA on this membrane (post-communication with Ms Margie Pallotta).

Given that KK06 was identified on chromosome group 5, Southern analysis with membranes prepared from genomic DNA of wild-type and *ph1b* mutant plants was conducted to determine whether this candidate is located within the *ph1b* deletion region. Hybridisation results indicated that the *KK06* candidate resides outside of the *ph1b* region, with equivalent banding patterns present in both the wild-type and *ph1b* mutant lanes (Appendix B3, Figure 1). To substantiate this, *in silico* mapping of the *KK06* candidate was also conducted by mapping the genomic location of its rice homologue with regards to the wheat *Hyp6* gene, previously mapped as the last gene definitively located within the *ph1b* deletion region (Griffiths *et al.* 2006). In rice, the *Hyp6* gene and the rice equivalent of the *KK06* candidate are located roughly 477 kb apart while the start of the *ph1b* deletion breakpoint region has been mapped to a region between the *Hyp6* and *Elp1* gene that are separated by a distance of only 1 kb (Griffiths *et al.* 2006). The results of this approach

therefore suggest that the *KK06* candidate gene is located outside of the *ph1b* region, based on rice-wheat syntenic mapping (Figure 3.5 E).

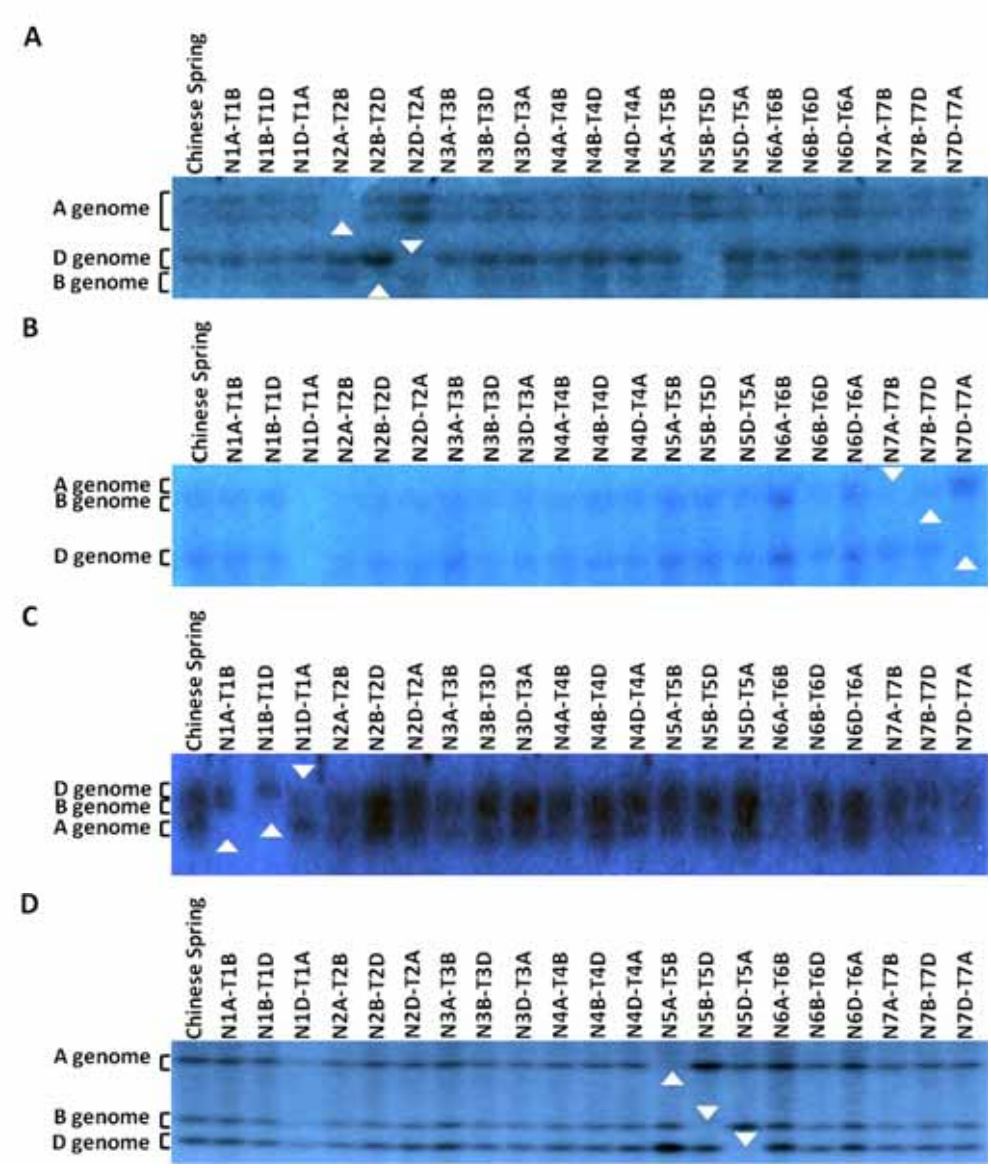
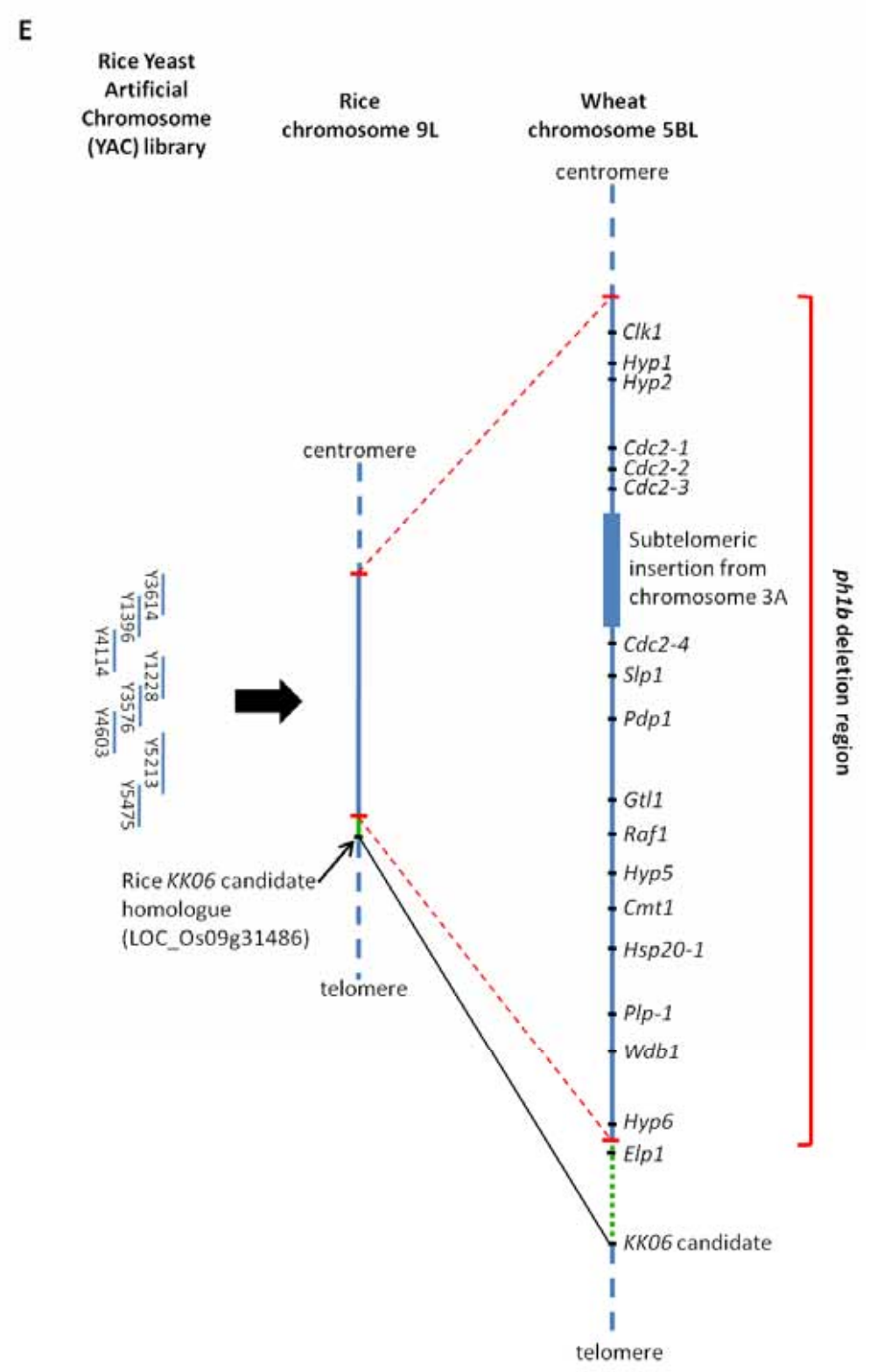


Figure 3.5 – Chromosomal locations of *KK01, KK03, KK04* and *KK06* candidate genes within the wheat genome. A) The *KK01* candidate resides on chromosome group 2. B) The *KK03* candidate resides on chromosome group 7. C) The *KK04* candidate resides on chromosome group 1. D) The *KK06* candidate resides on chromosome group 5. In all autoradiographs (A-D), white arrows indicate missing bands. E) A simplistic cartoon that highlights the predicted location of the *KK06* candidate gene in comparison to the wheat *ph1b* deletion region (see next page). 23 rice YACs were contigged to form the region of the rice chromosome 9L that corresponds to the wheat *ph1b* deletion region (only seven that completely span the region are shown here). The green line on rice chromosome 9 indicates the 477 kb distance between the rice homologue of the *KK06* candidate and the *ph1b* breakpoint located between the rice *Hyp6* and *Elp1* genes while the green line in the wheat chromosome 5BL denotes an undefined distance of the equivalent region in wheat. Marker allocations are as previously defined by Griffiths *et al.* (2006). Cartoon is not drawn to scale (see next page).


E
Rice Yeast Artificial Chromosome (YAC) library
Rice chromosome 9L
Wheat chromosome 5BL
centromere
Clk1
Hyp1
Hyp2
centromere
Cdc2-1
Cdc2-2
Cdc2-3
Subtelomeric insertion from chromosome 3A
Y3614
Y1396
Y4114
Y1228
Y3576
Y4603
Y5213
Y5475
Cdc2-4
Slp1
Pdp1
Gtl1
Raf1
Hyp5
Cmt1
Hsp20-1
Plp-1
Wdb1
Hyp6
Elp1
ph1b deletion region
Rice KK06 candidate homologue (LOC_Os09g31486)
telomere
KK06 candidate
telomere

3.3.5 – Q-PCR analysis of candidate genes

To determine whether the gene expression profiles of the candidate genes reflected their proteomic profiles, Q-PCR analysis was performed on *KK01, KK03, KK04* and *KK06* using total cDNA obtained from PM-LP staged anther tissue of the Chinese Spring wild-type, *ph1b* and *ph2a* mutants as well as TI-TII staged anther tissue of the wild-type (Figure 3.6). Comparisons between the three genotypes analysed showed that the *KK01* candidate transcript levels were significantly higher in the PM-LP stage in the wild-type compared to the two *ph* mutants (Figure 3.6 A). In contrast to the proteomics results of Chapter 2 where the KK01 protein spot was only present in the PM-LP wild-type gels, comparisons between the PM-LP and TI-TII stages in the wild-type showed that there was no significant difference in the mRNA transcript copy numbers of the *KK01* candidate.

Profiling of the *KK03* candidate revealed it was a relatively low-abundance transcript that was expressed in the wild-type TI-TII stage (Figure 3.6 B). This was in line with the proteomic profile of the KK03 protein spot that was only present in the TI-TII stage wild-type gels. However, *KK03* transcripts were also detected (albeit at significantly lower levels) in the PM-LP stage of wild-type, *ph1b* and *ph2a*.

The *KK04* transcript expression levels were the highest of all four candidates analysed. The highest expression level of *KK04* was seen in the TI-TII stage of the wild-type followed by the PM-LP stage of both the *ph2a* mutant and wild-type (Figure 3.6C). However, the differences between the Chinese Spring and *ph2a* PM-LP values were not significant. This is in contrast to the 2DGE data that had shown that the KK04 protein spot was only present in the PM-LP stage of the *ph2a* mutant.

Profiling of the *KK06* candidate revealed that it was a low abundance transcript (Figure 3.6 D). Within the PM-LP stage, *KK06* transcript levels were significantly higher in the wild-type compared to the *ph1b* and *ph2a* mutants. In the wild-type, *KK06* transcript

levels were present at a significantly higher level in the PM-LP stage than the TI-TII stage. This data corresponded somewhat with the proteomics data where the KK06 protein spot was present in the PM-LP and TI-TII stages of wild-type but absent in the *ph1b* and *ph2a* mutants at the equivalent stage.

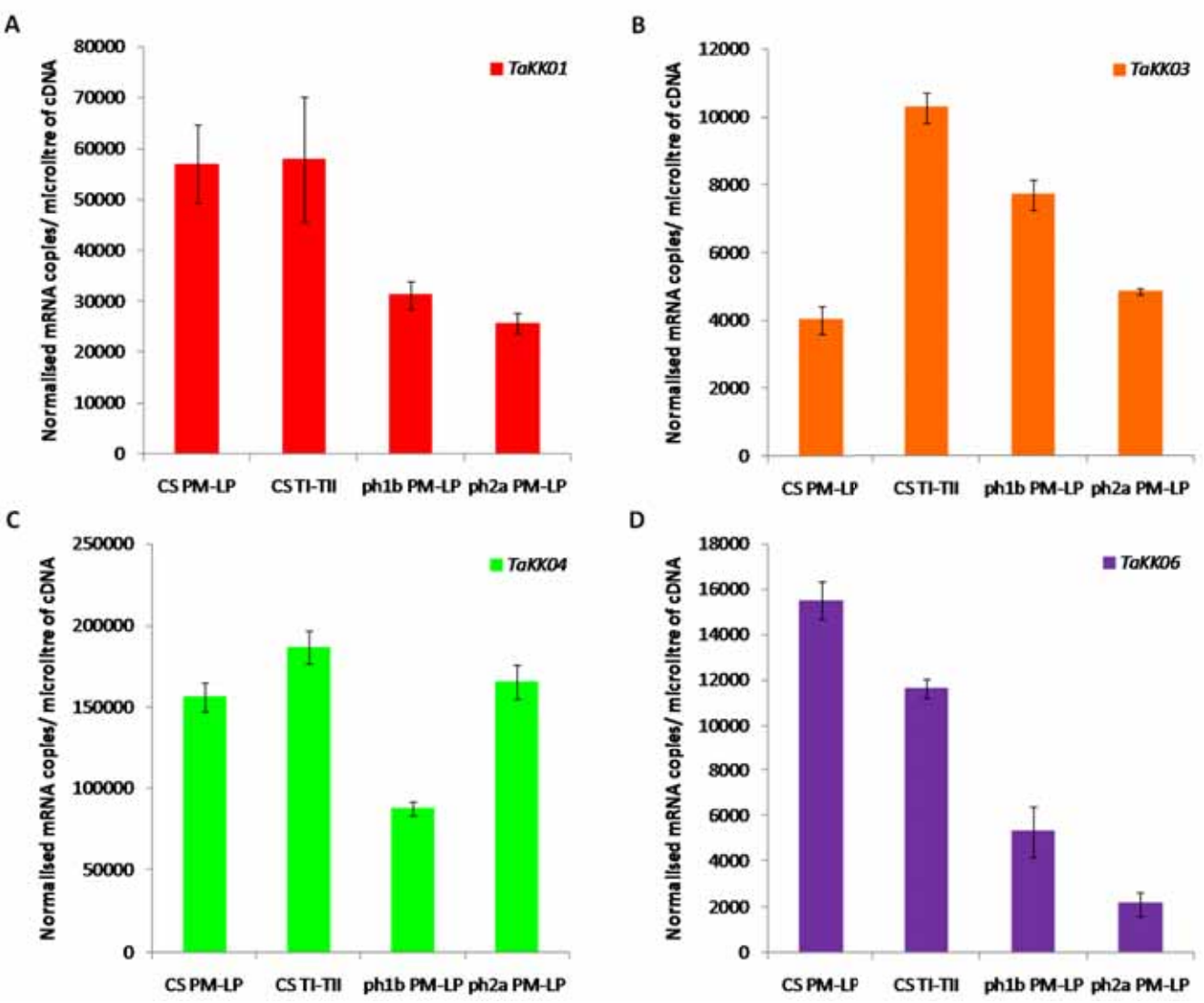


Figure 3.6 – Q-PCR analysis of the candidate genes. The expression profiles of the four candidates were analysed in the PM-LP stage of the Chinese Spring wild-type, *ph1b* and *ph2a* mutants, as well as the TI-TII stage of the wild-type. A) *KK01* showed significantly higher expression levels in both stages of the wild-type compared to the *ph* mutants during PM-LP. B) Expression of *KK03* was highest in the TI-TII stage of the wild-type compared to the wild-type and *ph* mutants during PM-LP. C) *KK04* appears to be a relatively high abundance transcript with no significant difference in its expression in the wild-type and *ph2a* mutant during PM-LP. D) *KK06* is a low abundance transcript with significantly higher levels of expression seen in the wild-type during PM-LP. Error bars represent the standard deviation of three replicate experiments.

3.3.6 – *In vitro* characterisation of proteins encoded by the candidate genes

3.3.6.1 – Extraction and isolation of protein products encoded by candidates transcripts

Construction of pDEST17 protein expression plasmids with the full-length candidate transcripts of *KK01*, *KK03*, and *KK06* allowed the proteins to be heterologously-expressed in *E. coli*. The 6×His-tagged proteins were then extracted under both denaturing and native conditions and purified using nickel affinity chromatography. All three protein products migrated close to their predicted molecular weights within denaturing SDS-PAGE gels (Figures 3.7, 3.8 and 3.9). Unlike KK01 and KK06 that were successfully isolated from their respective cell lysates, the KK03 protein product was completely lost in the washing steps. However, the production of the KK03 protein was still clearly seen as large bands (at its predicted molecular weight) in the cell lysate and flow-through lanes of the induced cell culture that are absent in the non-induced culture.

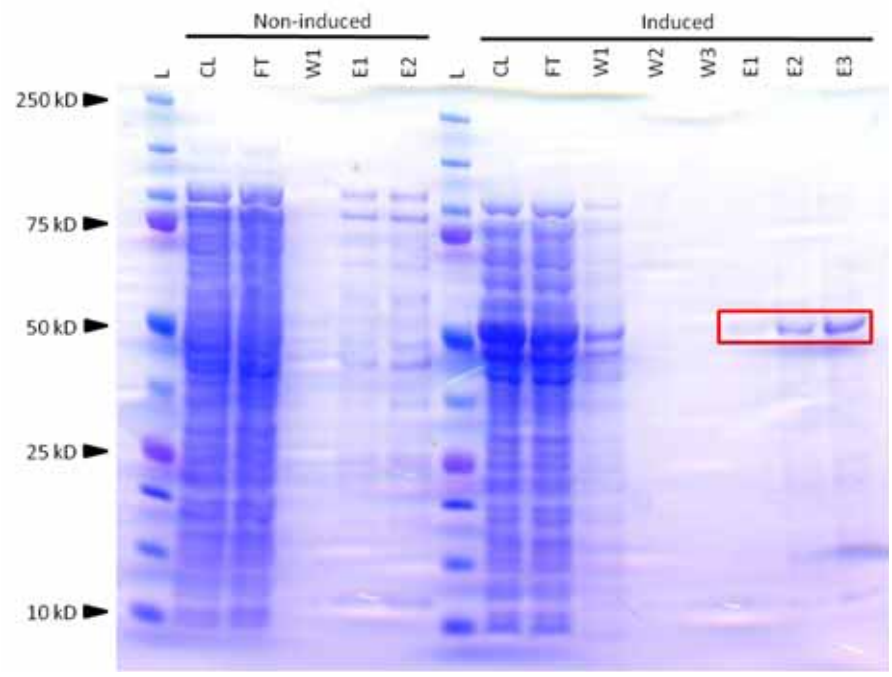


Figure 3.7 – SDS-PAGE analysis of the KK01 protein product. L – BIO-RAD Precision Plus molecular weight marker, Cell lysate (CL), Flow-through (FT), Wash 1 (W1), Wash 2 (W2), Wash 3 (W3), Elution 1 (E1), Elution 2 (E2) and Elution 3 (E3) samples are shown for the induced cell culture while W2, W3, and E3 were omitted from the non-induced cell culture as they were not required for the analysis of the result. The red rectangle highlights the 46.267 kD KK01 product.

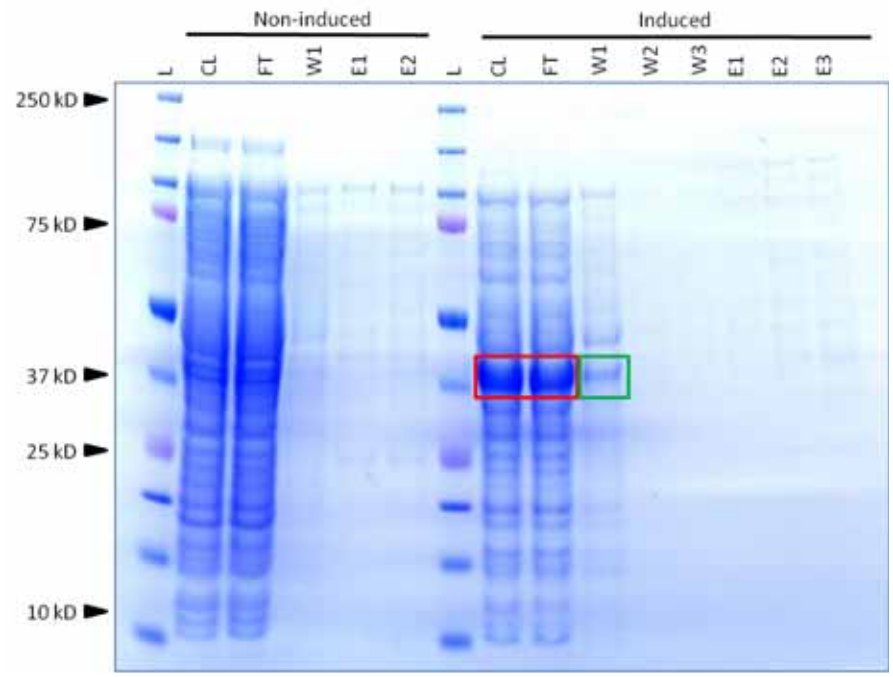


Figure 3.8 – SDS-PAGE analysis of the KK03 protein product. L – BIO-RAD Precision Plus molecular weight marker, Cell lysate (CL), Flow-through (FT), Wash 1 (W1), Wash 2 (W2), Wash 3 (W3), Elution 1 (E1), Elution 2 (E2) and Elution 3 (E3) samples are shown for the induced cell culture while W2, W3, and E3 were omitted from the non-induced cell culture as they were not required for the analysis of the result. The red rectangle highlights the 38.860 kD KK03 product that was produced but not captured within the column while the green rectangle highlights the residual 38.860 kD protein species that was lost in the first wash.

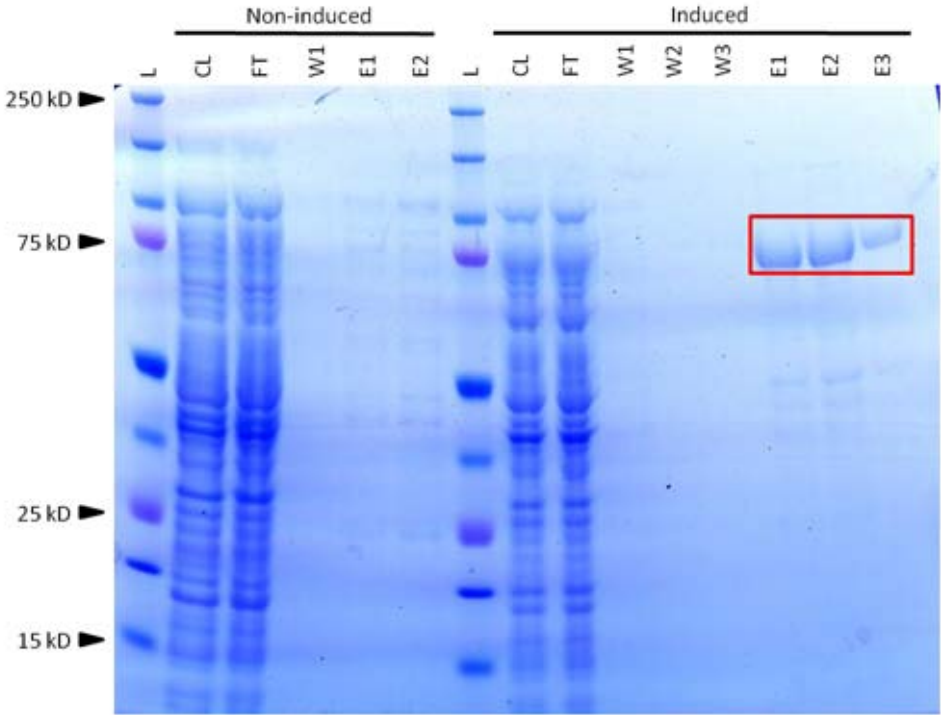


Figure 3.9 – SDS-PAGE analysis of the KK06 protein product. L – BIO-RAD Precision Plus molecular weight marker, Cell lysate (CL), Flow-through (FT), Wash 1 (W1), Wash 2 (W2), Wash 3 (W3), Elution 1 (E1), Elution 2 (E2) and Elution 3 (E3) samples are shown for the induced cell culture while W2, W3, and E3 were omitted from the non-induced cell culture as they were not required for the analysis of the result. The red rectangle highlights the 72.697 kD KK06 product.

3.3.6.2 – DNA-binding capabilities of full-length proteins KK01 and KK06

Purified samples of the KK01 and KK06 proteins extracted under native conditions were used in competitive DNA-binding assays to determine whether these proteins were able to interact with DNA *in vitro*. The competitive DNA-binding assays showed that both KK01 and KK06 were able to interact with single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA) *in vitro* (Figure 3.10). KK01 preferentially interacted with ssDNA when equivalent amounts of ss- and dsDNA were present as the retardation of the ssDNA species was more apparent at lower concentrations of the KK01 protein (Figure 3.10 A). However, a slight retardation of the dsDNA species was observed when the KK01 protein concentration was increased. In contrast, KK06 interacted with both DNA species indiscriminately as the rates of migration for both DNA species were equally retarded (Figure 3.10 B). In addition, the retardation of the dsDNA species in the KK06 samples was more apparent at each concentration compared to KK01. At 6 µM, the KK06-DNA-protein complexes had only just migrated into the gel matrix from the well (Figure 3.10 B).

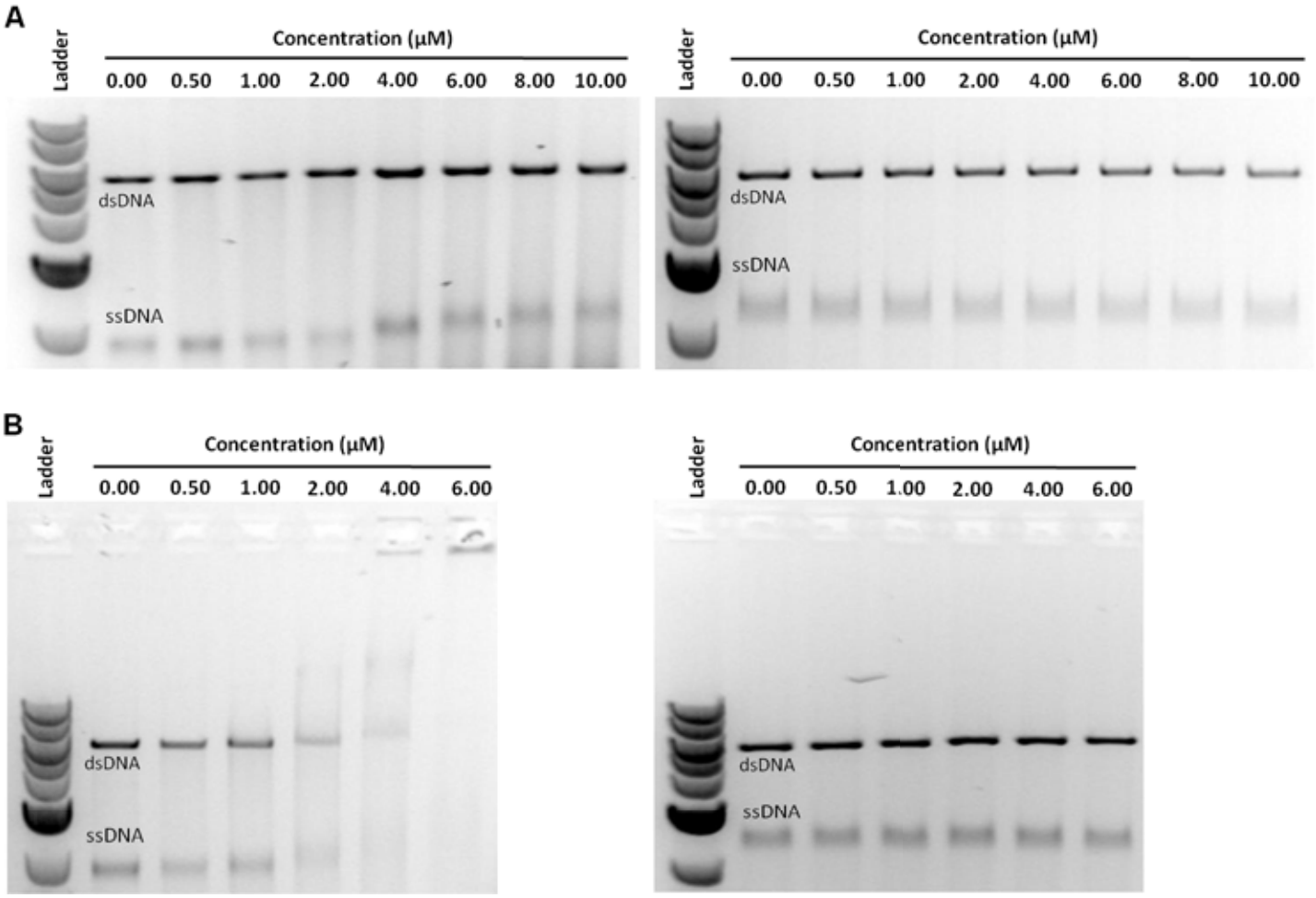


Figure 3.10 – Protein products for KK01 and KK06 interact with DNA *in vitro*. A) KK01 preferentially interacts with ssDNA but can interact with dsDNA when high concentrations of KK01 protein are present. B) KK06 interacts with both species of DNA equally and indiscriminately. At 6.00 µM of KK06 protein, the DNA has just entered the gel from the well. For both proteins, induced samples are shown on the left while the non-induced control samples are shown in the gels on the right.

3.4 – Discussion

The use of 2DGE paired with MS/MS to identify protein species has proven to be exceptionally useful, especially when the proteins to be identified come from an organism with a fully sequenced genome. Previous work conducted in Arabidopsis by Casasoli *et al.* (2008) and humans (Sharma *et al.* 2009) have shown the usefulness of working with a fully sequenced genome. Indeed, this approach has also been successfully used to varying degrees in the proteomic analyses of organisms without fully sequenced genomes as was reported by Sánchez-Morán *et al.* (2005) who conducted their study in *Brassica oleracea*. This was made possible by the high level of gene conservation (in excess of 85%) between *Brassica* and Arabidopsis (Cavell *et al.* 1998). Similarly, the aim of this study was to identify proteins that were present during the early stages of wheat meiosis by exploiting the gene conservation that exists between rice (*Oryza sativa*) and allohexaploid bread wheat (*Triticum aestivum*). In addition to having a fully sequenced genome, rice is also a closer relative to wheat compared to Arabidopsis, thus increasing the odds of gene conservation between the two species (Bolot *et al.* 2009).

This chapter reported on the isolation and characterisation (where possible) for the full-length coding sequences of genes responsible for encoding the proteins identified in Chapter 2. To achieve this, peptides of the six digested protein spots were matched against the rice proteome. In general, high level peptide matching accuracy is achieved when at least five peptides, which in combination cover at least 15% of an entire protein sequence, can be matched to a protein database hit (reviewed by Mann *et al.* 2001; and references therein). However, the results of the matching process in this study were limited. While the

MS/MS data reported multiple peptide species within each spot analysed, none of the peptides of each spot could be successfully matched to the same protein hit within the rice proteome database.

A few possible explanations exist as to why this was the case; firstly, the level of amino acid sequence conservation between the rice and wheat proteome may not be sufficiently high enough. This explanation is plausible based on the fact that the wheat genome is 40× larger than the rice genome with approximately 20-30% gene duplication (Akhunov *et al.* 2003, Arumuganathan and Earle 1991). In addition, comparative analysis of the fully-sequenced rice genome with the incomplete wheat genome have previously reported the presence of partial wheat ESTs that were mapped to regions of poor conservation within the rice genome (Sorrells *et al.* 2003). Akhunov and colleagues (2003) have previously hypothesised that genes within such regions may experience increased rates of recombination leading to formation of new genes due to the physical location of the genes on the chromosome arms. Therefore, in general, while a large proportion of rice and wheat proteins may share relatively good sequence conservation some wheat and rice proteins may only retain short regions of sequence similarities. As a consequence of the heavy reliance of this study on the amino acid sequence similarities between rice and wheat proteins, cross-species homology searches to identify such proteins would be highly difficult.

A second explanation is that the amino acid residues were misidentified due to the presence of post-translational modifications that were unaccounted for. To date, hundreds of post-translational modifications of amino acid residues have been identified (Wold 1981). However, only a few of these amino acid modifications have been fully characterised, with much of the work focused on

phosphorylation, glycosylation, SUMOylation and ubiquitination (reviewed by Ytterberg and Jensen 2010, and references therein). The physicochemical changes such modifications have on amino acid residues remain unknown and cannot be accounted for in MS/MS data. As a consequence, spectral data obtained from a highly-modified protein may be confounded by differences in molecular masses possibly leading to the misidentification of amino acids residues when analysed. Within this study, the Bioworks Version 3.3 software was configured to take into account three post-translational modifications (phosphorylation, glycosylations and SUMoylation) but did not take into account any other modifications; largely due to the lack of published data available. Furthermore, *de novo* sequencing was used to manually calculate the masses of the peptides identified in an effort to ensure that the amino acid sequences obtained were accurate.

However, the most likely explanation is that multiple protein species were present in each of the spots isolated thus resulting in a mix of peptides of the different protein species. Spectral data obtained from such a mixture would result in the identification of different protein hits within the database depending on the peptides used in the database search. The presence of multiple proteins within a single spot could have been caused by insufficient resolution by the 2D gel format used or by sample contamination. While the initial optimisation results reported in Chapter 2 showed sufficient resolution of protein spots using the 11 cm gel format, it is still possible that multiple proteins with the same (or very similar) MW and pI properties may not have been sufficiently resolved. Proteins with identical physicochemical properties migrate to the same coordinates within 2D gels and cannot be separated via this technique while proteins with highly similar, but not identical properties may not have been resolved far enough due to the size

of the 2D gels used. Furthermore, while the silver-staining protocol was very sensitive, it is possible that some of these protein spots (if they were present at low abundances) were not stained resulting in 'invisible' spots very close to the six identified spots analysed. Another possibility is that the samples may have been contaminated with keratin during the manual handling of the gels downstream of the 2DGE process. However this is less likely as peptides resulting from keratin contamination were automatically filtered and eliminated by the configured Bioworks Version 3.3 software. Furthermore, some proteins with different molecular weights and pI values could still have been held together in their protein complexes even under the denaturing conditions used in the protein extraction and 2DGE. Previous reports have shown that in some instances, strong protein-protein interactions lead to complexes being resistant not only to denaturation by SDS but also digestion by proteases (Haering *et al.* 2008, Kreisberg *et al.* 2002, Kubista *et al.* 2004, Manning and Colon 2004).

In an effort to overcome the fact that no two peptides identified from a single spot were able to be matched to any one protein within the rice database, primers were made based on the coding sequences of the codons for each peptide obtained from the MS/MS data, and were then used in RACE PCR experiments. To this end, full-length candidate coding sequences were successfully isolated for three of the spots (*KK01*, *KK03* and *KK06*) while a partial coding sequence was also isolated for *KK04*. Optimisation steps used in an effort to generate PCR products when the standard protocol yielded none included the use of degenerate primers, reduced annealing temperatures, and increased template cDNA concentrations. The use of degenerate primers and reduced annealing temperatures would have decreased the specificity of the primer binding

capacities thus increasing the chances of amplifying a PCR product while the increased cDNA concentrations eliminated the possibility that the candidate transcripts may have only been present at very low levels. New RACE libraries were also synthesised and used to eliminate the possibility that the RACE libraries used were degraded. However, these optimisation steps did not yield any specific PCR products for *KK02* and *KK05*. One likely explanation for the failure in isolating *KK02* and *KK05* is that the primers were made against misidentified peptide sequences.

Analyses of the four candidate coding sequences isolated (*KK01*, *KK03*, *KK04* and *KK06*) revealed that the RACE primer sequences used to amplify these sequences were present. In addition, these nucleotide sequences were in-frame with the rest of the ORFs and generated codons that corresponded to the amino acid sequences of their respective peptides obtained in the MS/MS data (Figure 3.3 and Table 3.2). This gave a reasonable level of confidence that the correct transcripts were isolated. Furthermore, BLASTp results of the predicted amino acid sequences of these four candidates against the rice database showed that *KK03*, *KK04* and *KK06* corresponded to search results yielded from the Bioworks software comparisons of the MS/MS peptide data against the rice proteome. Although *KK01* was the only transcript not to encode a protein that was identified by the Bioworks software, its predicted amino acid sequence did contain a stretch of amino acids that were identical to one of the peptides identified in the MS/MS data of protein spot KK01. Comparisons of the four predicted amino acid sequences against their rice homologues revealed that they shared relatively high levels of sequence similarities. This suggests that the other peptides isolated from the MS/MS data of the protein spots were most likely caused by protein

contamination or by changes to the masses of the amino acids by any of the hundreds of post-translational modifications that could lead to the misidentification of the amino acid residues and hence peptides.

Analyses of the predicted protein sequence of KK01 revealed that it shared high levels of sequence similarity to a rice speckle-type POZ protein and possesses a complete N-terminal MATH domain and a partial C-terminal BTB domain (Figure 3.4 A). Proteins which possess both these domains have previously been characterised and been shown to have the same domain order as as that seen in KK01 (Stogios *et al.* 2005). Furthermore, previous work has shown that the MATH and BTB domains are found in dimerising proteins that are capable of interacting with DNA (thus regulating the transcriptional expression of target genes through the recruitment of nuclear corepressor proteins by the BTB domain) (Sunnerhagen *et al.* 1997, Wolffe 1997). Such data suggested that KK01 could interact with DNA and play a role in gene expression regulation or chromatin architecture modification.

The predicted amino acid sequence of KK03 showed high sequence similarity with the rice pollen-specific *SF21C1* gene and possesses an esterase lipase superfamily domain (Figure 3.4 B). While not much is known about the *SF21* gene in rice, previous work conducted in *Helianthus annuus* (sunflower) has shown that the *SF21* gene family consists of multiple members that are differentially expressed in a tissue-dependant manner with different splice variants present within different cell types (Krauter-Canham *et al.* 1997, Lazarescu *et al.* 2006). Based on BLAST and homology search results against publicly-available databases, *KK03* is similar to the sunflower *SF21C1* gene. In sunflower, this gene encodes at least three splice variants whose protein products

vary in length, the longest being 355 amino acid residues while the other two variants are truncated products of 90 and 138 amino acids respectively (Lazarescu *et al.* 2006). These splice variants are detected in various organs with one of the truncated variants only found in pollinated florets (Lazarescu *et al.* 2006). The *SF21* gene family has previously been shown to be evolutionarily linked to the *Negative differentiation regulator* (*Ndr*) gene family (Krauter-Canham *et al.* 2001). However, the precise function(s) of both these gene families are not yet known although they have been hypothesised to be required for cell differentiation (Okuda and Kondoh 1999) and appear to have an enzymatic function due to their sequence similarities to hydrolase enzymes and the ligand-binding region of the inositol 1,4,5, triphosphate receptor (Hotelier *et al.* 2004, Krauter-Canham *et al.* 1997). In poppy, inositol 1,4,5, triphosphate has previously been shown to be part of the calcium signalling pathway required for pollen tube growth (Franklin-Tong *et al.* 1996). Previous work conducted by Nas and colleagues (2005) indicated that the rice *SF21* gene they identified was temperature-sensitive and controls male sterility; while the SF21 protein in tobacco is localised to the tips of pollen tubes (Lazarescu *et al.* 2006). In light of the fact that the SF21 protein shares moderate sequence similarity with the inositol 1,4,5 triphosphate receptor and that it localises to the tips of pollen tubes in tobacco, SF21 protein may be required for pollen tube growth (Allen *et al.* 2010).

Although isolation of the full-length *KK04* coding sequence proved unsuccessful, analysis of its predicted amino acid sequence showed it shares sequence homology with two putative barley hexose transporter-like proteins. These putative proteins are hypothesised to transport hexose sugars to the mitotic

endosperm cells during the early stages of grain tissue development (Weschke *et al.* 2003).

In contrast to *KK04*, the full-length coding sequence of the *KK06* candidate was successfully isolated. The high levels of sequence similarity seen between *KK06* and the HSP70-2 proteins found in both humans and mice may provide clues as to its role and explain why it was so abundant in the Chinese Spring PM-LP gels in Chapter 2. Previous studies have shown that the HSP70-2 protein of both mammalian species form part of the synaptonemal complex during spermatogenesis and are vital for disassembly of the synaptonemal complex (Allen *et al.* 1996, Dix *et al.* 1996, Dix *et al.* 1997, Zakeri *et al.* 1988). Thus the KK06 candidate isolated may possibly play a similar role in bread wheat.

Nullisomic-tetrasomic Southern blot analyses revealed that *KK01*, *KK03* and *KK04* were not located within the *ph1b* or *ph2a* deletion regions. Although *KK06* was present on chromosome group 5 with a copy on all three genomes, genotype Southern blots (as well as *in silico* rice-wheat synteny mapping) revealed that this candidate is located outside of the *ph1b* deletion region. When taken into account with the proteomics results of Chapter 2, these results can be used to decipher the proteomics data. For the case of *KK01* that was only present in the PM-LP sample of the wild-type, this indicates that *KK01* is either directly or indirectly regulated by a combination of elements encoded by both the *Ph* loci or that the KK01 protein levels are severely affected in mutants with homoeologous pairing phenotypes. Similarly, when the *KK06* mapping and proteomics results are considered together, the data suggests that *KK06* may also be directly or indirectly regulated by a combination of elements from both *Ph* loci where the absence of either locus leads to reduced levels of the KK06 protein. The

data from this component of work also indicated that *KK04*, that resides on chromosome group 1 and is only present in the PM-LP *ph2a* gels, may possibly be mis-regulated in the absence of the *Ph2* locus leading to ectopic expression in the PM-LP stage of the *ph2a* mutant.

Q-PCR expression analysis of the four candidates was performed to determine whether their transcript profiles could be correlated to the presence of their respective spots in the proteomics data of Chapter 2. While the presence of the KK01 spot being present in the PM-LP Chinese Spring gels is explained by the high levels of *KK01* expression during PM-LP, the absence of the KK01 spot in the TI-TII stage gels was at odds with the Q-PCR data that detected *KK01* transcripts at levels similar to the PM-LP stage (Figure 3.6 A). This could be due to post-transcriptional regulation of the mRNA transcripts resulting in no KK01 protein being produced in the TI-TII sample. Indeed, such differences in gene expression and protein levels have previously been reported and are not uncommon (reviewed by Waters *et al.* 2006, and references therein). *KK01* expression was also detected in the PM-LP stage of both *ph* mutants albeit at significantly lower levels (Figure 3.6 A) which may have resulted in the KK01 protein being produced at levels that were too low to detect in Chapter 2. This may explain the absence of the KK01 spot in the PM-LP stage gels of both *ph* mutants as well as strengthen the hypothesis that the KK01 candidate is directly/indirectly regulated by a combination of elements from both these loci and that KK01 protein levels are severely affected in these mutants. The *KK03* candidate expression profile fits its putative role as a *SF21C1* pollen-specific gene as it showed increased transcript levels as meiosis progressed from PM-LP to the TI-TII stage in wild-type (Figure 3.6 B). This data was in agreement with the

proteomics data that showed that the KK03 protein was only present in the TI-TII gels. In contrast to the proteomics data where the KK04 spot was only present in the PM-LP stage *ph2a* gels, *KK04* expression levels were relatively similar between both stages of the wild-type and the PM-LP stage of the *ph2a* mutant but was significantly lower in the *ph1b* mutant (Figure 3.6 C). This conflicting gene expression and proteomics data may be explained by post-transcriptional regulation of the *KK04* transcript (leading the KK04 transcript to only be translated in the *ph2a* mutant). Both the proteomics data and the Q-PCR expression profile of *KK06* fit its putative function as a HSP70-2 variant that may form part of the synaptonemal complex and regulate chromosome desynapsis during early meiosis as it is highly expressed during the PM-LP stage of wild-type and reduces significantly during the TI-TII stage (Figure 3.6 D). The significantly reduced levels of *KK06* transcript seen in both *ph* mutants again reaffirms the hypothesis that this gene may be directly/indirectly regulated or affected by the actions of a combination of elements encoded by both *Ph* loci. Both the similarities and differences in the proteomics and Q-PCR expression datasets of all four candidates highlights the need to study cellular processes at the protein level and that much more research must be conducted before we can truly grasp the complexities of gene transcription and translation.

While heterologous expression and purification of the KK01 and KK06 full-length candidates were straightforward, KK03 could not be purified even though induction of protein production was successful. The SDS-PAGE results showed that the KK03 candidate protein did not bind to the Ni-NTA column and that any residual KK03 protein was washed out in the first wash step. The inability of the 6×His-tag to bind to the Ni-NTA column could have been caused

by the 6×His-Tag interacting with a region of negatively-charged amino acid residues within the KK03 protein (thus becoming unavailable for ligand interaction with the Ni-NTA resin). An alternative approach using a glutathione-S-transferase (GST) tag instead of the 6×His-tag could have overcome this problem. However, due to time constraints, this alternative approach was not investigated.

The results of the DNA-binding assays confirmed the putative functions of KK01 and KK06 *in vitro*. Similar to some MATH-BTB domain proteins, KK01 interacts with DNA (Figure 3.10 A). Although KK01 preferentially bound ssDNA in the competitive DNA-binding assay, it is possible that KK01 is able to interact much more strongly with dsDNA *in planta* through protein-protein interactions with partner proteins that were not present in the assay. The slight shift of the dsDNA species at high concentrations of KK01 protein indicates that it does weakly interact with dsDNA when present alone. In contrast, the KK06 protein product interacts with both species of DNA indiscriminately regardless of the protein concentration used *in vitro* (Figure 3.10 B). The ability to interact with both species of DNA suggests that it could be a component of the synaptonemal complex and that it may interact with chromatin during early meiosis. Now that the basic characterisation of the KK01 and KK06 candidate proteins have been completed and their putative functions of DNA-binding confirmed *in vitro*, future work can focus on characterising the functions of both these proteins *in planta*.

Chapter 4 – Poor homologous synapsis 1 (PHS1) interacts with chromatin but does not co-localise with ASYnapsis 1 (ASY1) during early meiosis in bread wheat

BMC Plant Biology 2011, *submitted* 27th May 2011.

A signed declaration by all authors allowing the inclusion of this manuscript is presented in Appendix C1.

Poor Homologous Synapsis 1 (PHS1) interacts with chromatin but does not co-localise with ASYnapsis 1 (ASY1) during early meiosis in bread wheat

Kelvin H.P. Khoo, Amanda J. Able & Jason A. Able*

School of Agriculture, Food & Wine, The University of Adelaide, Waite Campus, PMB1, Glen Osmond, South Australia, 5064

* Correspondence:
School of Agriculture, Food & Wine, The University of Adelaide, Waite Campus, PMB1, Glen Osmond, South Australia, 5064.
Email: jason.able@adelaide.edu.au
Phone: +61 8 8303 7075
Fax: +61 8 8303 7109

E-mail:
kelvin.khoo@adelaide.edu.au
amanda.able@adelaide.edu.au

Abstract

Background

Chromosome pairing, synapsis and DNA recombination are three key processes that occur during early meiosis. A previous study of *Poor Homologous Synapsis 1* (*PHS1*) in maize suggested that PHS1 has a role in coordinating these three processes.

Results

Here we report the isolation of wheat (*Triticum aestivum*) *PHS1* (*TaPHS1*), and its expression profile during, and post-meiosis. While the *Ta*PHS1 protein has sequence similarity to other plant PHS1/PHS1-like proteins, it also possesses a unique region of oligopeptide repeat units. In addition, we show that *Ta*PHS1 interacts with both single- and double-stranded DNA *in vitro*, and provide evidence of the protein region that imparts the DNA-binding ability. Immunolocalisation data from assays conducted using an antibody raised against *Ta*PHS1 clearly show that *Ta*PHS1 associates with chromatin during early meiosis, with the signal persisting beyond chromosome synapsis. Furthermore, *Ta*PHS1 does not appear to co-localise with the asynapsis protein – *Ta*ASY1 – possibly suggesting that these proteins are independently coordinated.

Conclusions

Significantly, the data from the DNA-binding assays and 3-dimensional immunolocalisation of *Ta*PHS1 during early meiosis indicates that *Ta*PHS1 interacts with DNA – a function not previously observed in either the Arabidopsis or maize PHS1 homologues. As such, these results provide new insight into the function of PHS1 during early meiosis in bread wheat.

Background

For the majority of sexually reproducing organisms, meiosis is a cellular process required for gamete formation, and is composed of one round of DNA replication, followed by two rounds of chromosome division. During meiosis I, a reductional division event leads to the segregation of homologous chromosome pairs while an equational division during meiosis II leads to the segregation of the sister chromatids.

For the successful juxtaposition of homologous chromosomes, three key processes occur during prophase I, namely pairing, synapsis and recombination. Previous studies investigating the molecular mechanisms of homologous chromosome pairing have revealed complex interplay between these three tightly-linked processes (Armstrong *et al.* 2001, Chen *et al.* 2004, Higgins *et al.* 2005, Kerzendorfer *et al.* 2006, Martínez-Pérez *et al.* 2003).

In allopolyploid organisms such as bread wheat (*Triticum aestivum*), correct alignment and pairing of homologous chromosomes is complicated by the presence of genetically similar genomes, known as homoeologous genomes. Although bread wheat possesses three homoeologous genomes (termed A, B and D), meiosis proceeds as if the organism is a diploid, in that pairing only occurs between homologous chromosomes from the same genome (Able *et al.* 2009, Able and Langridge 2006, Able *et al.* 2007, Moore and Shaw 2009, and references within). This strict pairing interaction between homologous chromosomes has previously been shown to be controlled by *Pairing homoeologous* (*Ph*) loci (Riley and Chapman 1958, Sears 1977). The most extensively studied of these loci is the *Ph1* locus located on the long arm of chromosome 5B. While the molecular mechanism by which the *Ph1* locus operates is still subject to intensive research, *Ph1* appears to indirectly promote

homologous chromosome pairing by suppressing homoeologous chromosome interactions through regulation of the specificity of chromosome interactions at centromeric and telomeric regions (Martinez-Pérez *et al.* 2001, Prieto *et al.* 2004).

In *ph1* deletion mutants, the chromatin is prematurely and asynchronously remodelled, leading to abnormal chromosome conformation that results in increased interactions between homoeologous chromosomes (Colas *et al.* 2008, Prieto *et al.* 2004). These mutants also display other meiotic defects such as the arrest of synapsis during zygotene and the presence of uncorrected multiple axial element associations, which in wild-type are normally corrected prior to entry into pachytene (Holm 1988, Holm and Wang 1988). While the deletion region in the *ph1b* mutant is extensive, the *Ph1* locus has recently been refined to an area that contains, among other genes, seven *Cyclin-dependent kinase*-like (*Cdk*-like) genes (Al-Kaff *et al.* 2008, Griffiths *et al.* 2006).

Our current knowledge of other meiotic genes mostly comes from research on model species such as yeast and Arabidopsis. However, putative homologues of many of these genes have also been identified in the cereals (Bovill *et al.* 2009). Some of the early meiotic genes characterised in various plant species include *ASY1* (*ASYnapsis 1*) (Boden *et al.* 2009, Boden *et al.* 2007, Caryl *et al.* 2000, Nonomura *et al.* 2004, Ross *et al.* 1997), *RAD51* (*RADiation sensitive 51*) (Bleuyard *et al.* 2005, Li *et al.* 2007) and *PHS1* (*Poor Homologous Synapsis 1*) (Pawlowski *et al.* 2004, Ronceret *et al.* 2009). In wheat, *ASY1* (*TaASY1*) is involved in chromosome synapsis and promotes homologous chromosome pairing during meiosis I (Boden *et al.* 2009, Boden *et al.* 2007). Interestingly, *Taasy1* knock-down mutants have been reported to display defective chromosome characteristics similar to that of the *ph1b* mutant (Boden *et al.* 2009). In the

absence of *Ph1*, the expression of *TaASY1* is approximately 20-fold higher compared to wild-type while the localisation of its protein product is also affected. This indicates that *TaASY1* is intimately involved in the *Ph1*-dependent control of chromosome pairing in wheat (Boden *et al.* 2009, Moore and Shaw 2009).

PHS1 was first identified in a *Mutator* transposon-mutagenised maize population, with no known homologues in yeast or other non-plant species (Pawlowski *et al.* 2004). While phenotypic analysis of the *phs1-0* mutant by Pawlowski *et al.* (Pawlowski *et al.* 2004) revealed no vegetative defects, meiosis was disrupted resulting in male and female sterility. Transmission electron microscopy of meiotic spreads from *Zmphs1-0* meiocytes revealed significantly reduced levels of synapsis during zygotene and improper alignment of the chromosomes in the synapsed regions observed. Additionally, the chromosomes showed synapsis with multiple partners. Coupled together with results of their fluorescent *in situ* hybridisation (FISH), the data indicated that non-homologous chromosome synapsis was present in the *Zmphs1-0* mutant.

FISH results from recent work by Ronceret *et al.* (Ronceret *et al.* 2009) on the Arabidopsis homologue of PHS1 (*At*PHS1) showed that PHS1 appears to function in a similar manner in different species independent of genome size and complexity. Chromosome axis formation and installation of the synaptonemal complex components in both wild-type and *phs1* mutant cells of Arabidopsis and maize appeared similar albeit with ZYP1 loading being delayed in some instances (Ronceret *et al.* 2009). Immunolocalisation in Arabidopsis and maize revealed that PHS1 is located within the cytoplasm during the early stages of meiosis with some foci clustered along the nuclear envelope during zygotene and present in the nucleus during pachytene. With a reduced number of RAD50 foci observed in the

nuclei of the *phs1* mutants, Ronceret and colleagues (Ronceret *et al.* 2009) concluded that PHS1 regulates meiotic recombination and chromosome pairing by controlling the transport of RAD50; a protein which is required during double-strand break processing.

Our analysis of the PHS1 protein in bread wheat, provides evidence that *Ta*PHS1 possesses DNA-binding capabilities even though no known DNA-binding domains were identified *in silico*. Our data also show that *PHS1* is up-regulated in the *ph1b* bread wheat mutant when compared to wild-type, and that *Ta*PHS1 is associated with chromatin and is present on the nucleolar periphery during the early stages of meiosis as indicated through immunolocalisation analysis using an antibody that was raised against the full-length wheat PHS1 protein.

Results

PHS1 is highly similar across plant species and in wheat it encodes a predicted protein product with a unique oligopeptide repeat sequence

PCR amplification from whole meiotic spike cDNA using the primers listed in Table S1 (Additional Information File 1) resulted in the isolation of *TaPHS1* which has a 1071 bp ORF. This encodes a 357 amino acid protein with a predicted molecular weight of 38.958 kD and a pI of 5.23. Despite no nuclear localisation signal (NLS) peptides being detected within the *Ta*PHS1 amino acid sequence using SignalP 3.0, WoLF PSORT analysis predicted that *Ta*PHS1 is most likely to be located within the cell nucleus. Using AlignX, comparative amino acid sequence analysis of full-length annotated PHS1 or PHS1-like

proteins obtained through database searches (refer to Methods) showed that *Ta*PHS1 shared relatively high levels of sequence identity with its homologues in other species (*Sorghum bicolor* [*Sb*PHS1] – 53.6%; *Zea mays* [*Zm*PHS1] – 51.2%; *Oryza sativa* [*Os*PHS1] – 41.4%; and *Arabidopsis thaliana* [*At*PHS1-like] – 21.5%). We propose that there are four prominent regions within the PHS1 amino acid sequence that are well-conserved (Figure 1A). While a portion of Region 2 (corresponding to amino acid positions 99 – 145 of *Ta*PHS1) was previously identified to contain two conserved domains (Pawlowski *et al.* 2004), inter-species comparisons made in the current study suggest that this conserved region can now be extended by 11 amino acid residues toward the N-termini of PHS1 proteins in monocot species (Figure 1A, dashed line). In addition there is a short region of oligopeptide repeats from position 242 to 265 [YSGFPEGYSGFPEGYSGFPEGYSG] unique to *Ta*PHS1 (Figure 1A, boxed feature).

To assess the phylogenetic relationships between the five homologues shown in Figure 1A, a neighbour-joining tree was constructed using the full-length amino acid sequences (Figure 1B). As expected, Arabidopsis is the most divergent, while sorghum and maize share a higher degree of similarity with one another. Although wheat and rice fall within the same cluster, the internal branch length difference between the two species suggests that a significant level of sequence divergence has occurred. To determine whether other PHS1 sequences could be identified in the public databases, the more sensitive Hidden Markov Model (HMM) and MaxHom functions of the PredictProtein program were used. Three additional sequences were identified that were similar to *Ta*PHS1, and all from dicot species; namely poplar (*Populus trichocarpa*) (*E*-value: $7E^{-97}$), grape

(*Vitis vinifera*) (*E*-value: $7E^{-91}$) and castor oil (*Ricinus communis*) (*E*-value: $2E^{-97}$). The addition of these three sequences to the phylogenetic analysis shows that they cluster with Arabidopsis, the only other dicot species (Figure 1C).

Southern blot analysis showed that *TaPHS1* is located on chromosome group 7, with a copy on each of the three genomes (Figure 2). *In silico* mapping revealed that *TaPHS1* is likely to reside on the short-arm of this chromosome group (Bin 7AS8-0.45-0.59, data not shown). To determine this, rice genetic markers that are located close to *OsPHS1* (on rice chromosome 6) were used to screen wheat deletion bins. One marker previously bin-mapped to wheat chromosome group 7 [GenBank: BE404111.1] was identified to be approximately 35 kb from *OsPHS1*.

*Ta*PHS1 interacts with DNA and is expressed during meiosis

Previously reported homology searches using the maize PHS1 protein revealed low levels of sequence similarity with two families of fungal helicases, possibly indicating that PHS1 may interact with DNA (Pawlowski *et al.* 2004). However, extensive *in silico* amino acid analysis of *Ta*PHS1 revealed that it has no known conserved DNA-binding domains. To determine whether *Ta*PHS1 interacts with DNA, a competitive DNA-binding assay using recombinant *Ta*PHS1 extracted under native conditions was conducted (Figure 3A). Interactions occurred with both single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA). Interestingly there appears to be a higher affinity for ssDNA with a dsDNA shift only occurring at higher concentrations of *Ta*PHS1. Competitive DNA binding assays using partial *Ta*PHS1 peptides corresponding to the four prominent conserved regions identified in this study revealed that only Region 1 possesses

DNA-binding capabilities (Figure 3B-E). Region 1, like the full-length *Ta*PHS1 protein, appears to have a higher affinity for ssDNA compared to dsDNA. Sequence analysis of Region 1 shows that it contains two S/TPXX motifs (TPPP – amino acid positions 46 to 49; and SPAA – amino acid positions 71 to 74), which have previously been reported to bind DNA (Suzuki 1989).

Quantitative real-time PCR (Q-PCR) profiling of *TaPHS1* in wild-type Chinese Spring shows that it has relatively low transcript abundance during meiosis (Figure 4). Although *TaPHS1* is expressed in wheat anther tissue throughout all stages of meiosis examined and beyond, peak expression occurs during pre-meiotic interphase and immature pollen. Between the pooled stages of leptotene-pachytene and diplotene-anaphase I, there is no statistically significant difference in *TaPHS1* expression. Given that Boden *et al.* (Boden *et al.* 2009) demonstrated that the *TaASY1* transcript was significantly up-regulated in a *ph1b* background when compared to wild-type (approximately 20-fold), we also investigated transcription levels of *TaPHS1* in the *ph1b* mutant. While not as dramatic as that reported in Boden *et al.* (Boden *et al.* 2009), *TaPHS1* was also up-regulated in the *ph1b* mutant when compared to wild-type but by between 1.5-fold (pre-meiosis) and 2-fold (leptotene-pachytene) (Figure 4).

*Ta*PHS1 is localised in the nucleus and associates with chromatin during early meiosis

3D-immunolocalisation of *Ta*PHS1 in wild-type wheat meiocytes shows that it associates with chromatin during early meiosis (Figure 5A-F). While the signals of both *Ta*PHS1 and *Ta*ASY1 were located within close proximity of each other, the two proteins do not appear to co-localise (e.g. merged panel of Figure 5F). In addition to its association with chromatin, the *Ta*PHS1 signal was also observed at

the nucleolus (Figure 5B-E). This labelling of the nucleolus appears to be on the surface, with a greater signal intensity seen at the nucleolar periphery. From the images presented, this signal appeared most intense during early to late zygotene (Figure 5C-E). In general, the *Ta*PHS1 signal appeared either as diffuse tracts and/or punctated foci that follow the chromatin, unlike the distinct continuous tracts of *Ta*ASY1. The *Ta*PHS1 signal was observed from the telomere bouquet stage and persisted on the chromatin until late pachytene where it faded. Although *Ta*PHS1 was not detected on the chromatin in diplotene cells, detection of a weak signal was still observed in the cytoplasm in what appeared to be randomly distributed foci (Figure 5G).

Discussion

This study has reported the isolation and characterisation of *PHS1* from hexaploid wheat, with the amino acid sequences of *Ta*PHS1 being relatively well-conserved when compared with homologues in other plant species. *In silico* analysis of the *Ta*PHS1 amino acid sequence suggests that it does not contain any known nuclear localisation signal (NLS) peptide motif. However, predictions using WoLF PSORT show that *Ta*PHS1 fits the profile of a nuclear protein. In addition, the immunolocalisation results (Figure 5) show that *Ta*PHS1 localises to the nuclei of wheat meiocytes *in vivo*. These results together indicate that *Ta*PHS1, and its homologues, might possess an uncharacterised NLS motif. Alternatively, the PHS1 protein may be transported into the nucleus by a yet unknown process and/or protein. Given that Region 4, referred to as the CR2 region in Ronceret *et al.* (Ronceret *et al.* 2009), has been identified as a putative SUMOylation site; there may be no requirement for an NLS motif. Previously, post-translational

modifications such as SUMOylation have been shown to enable transport of proteins from the cytoplasm into the nucleus (de Carvalho and Colaiacovo 2006).

Sequence alignments of *Ta*PHS1 with PHS1 and PHS1-like proteins of four other species obtained from BLAST searches revealed that PHS1 is more similar between the cereal plants than Arabidopsis (Figure 1A and 1B). This was not unexpected as the cereals are monocots while Arabidopsis is a dicot. In addition, the Arabidopsis PHS1-like sequence contained an additional sixty-one amino acid residues on the C-terminal end that was not present in the four monocot species. With the addition of three more dicot PHS1/PHS1-like sequences, individual monocot and dicot branches were still maintained. However, a high level of divergence between the wheat and rice PHS1 sequences was evident (Figure 1C). A possible explanation for this is that rice is a diploid organism while wheat is a hexaploid. PHS1 proteins in the two species may have evolved to function differently, as is suggested by the presence of the *Ta*PHS1-specific oligopeptide repeat units from position 242 to 265 (Figure 1A, boxed feature). Even when compared with maize and sorghum, both of which are tetraploid (*albeit* with maize being a cryptic tetraploid), the rice PHS1 sequence is significantly different. PHS1 may have evolved to function differently according to the ploidy number, and consequently genome complexity, of each species. A less likely explanation for this sequence divergence is that the rice PHS1 sequence, which is putatively annotated as a PHS1 protein, is not a true PHS1 homologue but instead a PHS1-like protein. However, this seems improbable for a number of reasons. Firstly, the rice genome has been sequenced and exhaustive BLAST searches identified *Os*PHS1 as the most significantly similar match to both *Ta*PHS1 and *Zm*PHS1 at the nucleic and amino acid level. Secondly, the *in*

silico mapping identified a rice marker on rice chromosome 6 approximately 35 kb away from *OsPHS1* that is syntenous to a marker that has been bin-mapped to the short-arm of wheat chromosome 7. Finally, it has been documented across different organisms that there can be significant variance between the amino acid sequences of protein equivalents involved in meiosis, even though the function may be conserved (see Able *et al.* 2009; and references within).

The importance of the aforementioned oligopeptide repeat units unique to *Ta*PHS1 remains to be determined. These repeats could be either three tandem hepta-peptide units [YSGFPEG] that span positions 242 to 262, or a series of alternating tri-peptide [YSG] and tetra-peptide [FPEG] units that span positions 242 to 265. Comprehensive *in silico* database searches using amino acids 242 to 265 resulted in no significantly similar matches with any repeats reported to date. As single oligopeptide units; the tri-peptides, tetra-peptides, and hepta-peptides are relatively short and may therefore not form any independent structural units. However, when arranged successively, these oligopeptide units may form a regular repeating structure within *Ta*PHS1, as has previously been reported in other proteins (Katti *et al.* 2000). Based on previously reported amino acid nomenclature by Yoder *et al.* (Yoder *et al.* 1993), the glutamic acid (E) and glycine (G) residues on the end of the hepta-peptide unit [YSGFPEG] could represent turn-residues that link the hepta-peptides together, allowing the units to stack on top of each other. This series of oligopeptides may therefore impart a slightly different structure, and possibly function, for *Ta*PHS1 when compared to the rest of the PHS1/PHS1-like proteins that lack these oligopeptide repeat units.

Although two conserved domains have been previously described for PHS1 (Pawlowski *et al.* 2004), we have not only shown these to be part of one

conserved region across species but that this region is only one of four conserved regions within the amino acid sequences investigated. The four conserved regions reported here, termed Region 1 to 4, are in contrast to the two conserved regions, termed CR1 and CR2, reported by Ronceret *et al.* (Ronceret *et al.* 2009). Regions 2 and 4 identified in this study approximately correspond with CR1 and CR2. Discrepancies in the lengths of the conserved regions identified in the two studies are most likely artefacts of the differing alignment algorithms used; in addition to the fact that only full-length annotated transcripts of PHS1 were used in this study. *In silico* searches of protein databases revealed that *Ta*PHS1 does not have any known conserved DNA-binding domain that has been reported to date. Irrespective, the results of the DNA-binding assays in this study (Figure 3A and B) and the immunolocalisation of *Ta*PHS1 to chromatin (Figure 5) clearly show that *Ta*PHS1 does have DNA-binding capabilities. In the presence of equivalent amounts of ssDNA and dsDNA, *Ta*PHS1 appears to preferentially bind ssDNA *in vitro* but will also bind dsDNA when the protein is present at higher concentrations. Furthermore, we have shown that Region 1 appears to be responsible for the DNA-binding ability (Figure 3B). Although the retardation of the DNA is less significant, this can be attributed to the reduced size of the Region 1 partial peptide in comparison to that of the full-length *Ta*PHS1 protein (5.854 kD vs. 38.958 kD). Other regions within *Ta*PHS1 may also be required to further enhance the DNA-binding capabilities of Region 1. Although bioinformatic analysis of Region 1, and indeed the full-length *Ta*PHS1, failed to identify any conserved DNA-binding domains, two S/TPXX DNA-binding motifs previously identified by Suzuki (Suzuki 1989) are located within Region 1 (TPPP – amino acid positions 46 to 49; SPAA – amino acid positions 71 to 74). Region

1 could therefore possibly represent a novel DNA-interaction domain. Our immunolocalisation data suggests that *Ta*PHS1 is closely associated with chromatin (and therefore dsDNA) *in vivo* during early meiosis. *Ta*PHS1 may require a currently unknown protein partner (one that is not present in the *in vitro* DNA binding assay) to facilitate its ability to interact with dsDNA *in vivo*.

Although the Q-PCR profiling suggests *TaPHS1* is a transcript of relatively low abundance in the cell, some significant differences were detected during the stages examined, as well as between the wild-type Chinese Spring and *ph1b* mutant. While the same general trend of expression is observed in both wild-type and the *ph1b* mutant, *TaPHS1* is up-regulated in the mutant background by approximately 1.5-fold in pre-meiosis, 2-fold in both leptotene-pachytene and diplotene-anaphase I, and 1.5-fold in immature pollen. This 2-fold increase in expression during early meiosis may suggest that the *Ph1* locus could directly or indirectly affect *TaPHS1*.

The presence of substantial *Ta*PHS1 signal within the nucleoli of early-stage meiocytes may indicate that *Ta*PHS1 is sequestered to the nucleolus either for degradation or storage until required as has been shown for other proteins (reviewed by Carmo-Fonseca *et al.* 2000, Olson *et al.* 2000). Hypothetically, should PHS1 act as a direct physical shuttling protein (as indicated but not favoured by Ronceret and colleagues (Ronceret *et al.* 2009)) that transports specific meiotic proteins into the nucleus, it is likely that it may then be sequestered to the nucleolus for degradation or to undergo further post-translational modifications to mark it for return to the cytoplasm so that PHS1 molecules can be reused. *Ta*PHS1 may also have a role within the nucleolus given its intense signal. Such a role has been shown for the yeast meiosis-specific

pachytene checkpoint (PCH2), which is sequestered to the nucleolus and prevents HOP1 (the yeast homologue of *Ta*ASY1) from entering the nucleolus (San-Segundo and Roeder 1999). Given the lack of *Ta*ASY1 signal and the intensity of *Ta*PHS1 in the nucleolus, a similar role to PCH2 may be played by *Ta*PHS1. Clearly, *Ta*PHS1 and *Ta*ASY1 do not co-localise with one another, even though our immunolocalisation data shows that both these proteins are loaded and associated with the chromatin at the same time during early meiosis in bread wheat. The association between these two proteins appears to be particularly pronounced during late-zygotene to pachytene (Figure 5E, F). Could it be that *Ta*PHS1 is involved in a pachytene check-point mechanism to ensure that only homologous chromosomes have paired and recombined? Another intriguing result is that the *Ta*PHS1 signal profile appears as faint tracts with punctate foci along regions of the chromatin. Do these punctate foci denote possible recombination sites where *Ta*PHS1 may be loading the recombination machinery? This is plausible as previous reports of recombination proteins including RAD51 (Franklin *et al.* 1999), RAD50 (Ronceret *et al.* 2009) and MLH3 (Jackson *et al.* 2006) localise to chromatin as foci. A second question is whether the diffuse tracts of *Ta*PHS1 provide clues to suggest a direct role for *Ta*PHS1 in homology searching (in wheat at least) as previously suggested by both Pawlowski *et al.* (Pawlowski and Cande 2005) and Ronceret *et al.* (Ronceret *et al.* 2009).

Conclusions

In concluding, the data that we have presented clearly demonstrate that *Ta*PHS1 has an important and possible novel role during the early stages of wheat meiosis. Data from the DNA-binding assays as well as 3-dimensional immunolocalisation

of *Ta*PHS1 during early meiosis in wild-type cells indicate that *Ta*PHS1 interacts with DNA – a function not previously observed in the Arabidopsis and maize PHS1 homologues. The localisation signal profile of *Ta*PHS1 may indicate that it is a direct transporter of other meiotic proteins into the nucleus and that it could have a role in homology searching. While the role(s) of this protein are yet to be fully understood, we are currently in the process of generating *Taphs1* knock-down and *TaPHS1* over-expression mutants to further elucidate the meiotic function in bread wheat.

Methods

Plant material

Hexaploid wheat (*Triticum aestivum* L.) cv. Chinese Spring plants and a Chinese Spring mutant lacking the *Ph1* locus (*ph1b*) were grown in a glasshouse with a 16/8 h photoperiod at 23°C. Harvesting and staging of meiotic anthers from both wild-type and mutant plants, for quantitative real-time PCR (Q-PCR) and fluorescence immunolocalisation, were conducted as per Boden *et al.* (2009). Whole meiotic spike tissue was collected for the isolation and amplification of the gDNA and cDNA *Triticum aestivum PHS1* (*TaPHS1*) sequences.

RNA isolation and cDNA synthesis

Collected tissues for RNA isolation were initially ground in liquid nitrogen. Total RNA was extracted using Trizol reagent (Gibco-BRL, Carlsbad, CA, USA) according to the manufacturer's instructions. RNA concentration was determined using a Nanodrop (ND-1000) (Nanodrop, Wilmington, DE, USA). cDNA was

synthesised from 2 µg of total RNA using the iScript cDNA synthesis kit (Bio-Rad, Hercules, CA, USA) according to the manufacturer's instructions.

cDNA amplification and sequencing of the PHS1 coding sequence

Primers (see Table S1) for isolating the *TaPHS1* ORF were designed using the *OsPHS1* sequence (LOC_Os06g27860, MSU Rice Genome Annotation database - http://rice.plantbiology.msu.edu) identified through a TIGR rice expressed sequence tag (EST) BLAST search (accessed 21st October 2008).

Each PCR contained 100 ng cDNA, 0.2 mM dNTPs, 0.2 µM primers, 1U FastStart high fidelity *Taq* polymerase (Roche Applied Science, Mannheim, Germany) in 25 µL of 1× high fidelity buffer supplemented with 1× GC-RICH solution (Roche). PCR products were cloned into pCR8/GW/TOPO (Invitrogen) for DNA sequencing (15× coverage). Sequencing PCR and capillary separation was conducted using the same methods as described earlier except that GW1 and GW2 primers were used (see Table S1). Secondary sets of primers were designed on the sequenced products to specifically amplify the *TaPHS1* ORF. Amplification and sequencing was repeated as above. PCR cycling parameters were denaturation at 95°C for 5 min, followed by 35 cycles of 96°C for 30 s, T_m°C for 30 s, 72°C for 75 s, with a final elongation step at 72°C for 10 min (see Table S1 for T_m of primer sets). The assigned NCBI accession number for *TaPHS1* is GQ851928.

Bioinformatics analysis

DNA sequence alignments and comparisons were conducted with AlignX and Contig Express (Informax, VNTI Advance, Version 11, Frederick, MC, USA)

software programs. VNTI software was also used to predict the molecular weight and pI of the protein. To predict the cellular localisation of *Ta*PHS1, SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP/) (Bendtsen *et al.* 2004) and WoLF PSORT (http://wolfpsort.org/) (Horton *et al.* 2007) were used; while detection for conserved domains was performed using the NCBI Conserved Domain Search Tool (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi), InterPro Scan (http://www.ebi.ac.uk/Tools/InterProScan/) and Pfam 23.0 (http://pfam.janelia.org/). Amino acid alignments and comparisons of full-length PHS1 sequences (obtained from various BLAST searches using the NCBI, TIGR, and PredictProtein [http://www.predictprotein.org/; (Rost *et al.* 2004)] databases), and subsequent construction of the phylogenetic tree (neighbour-joining method) (Saitou and Nei 1987) was completed using Molecular Evolutionary Genetics Analysis (MEGA) software (version 4.0) (Tamura *et al.* 2007). Default parameters were used except for the following: the pair-wise deletion option was used, the internal branch test bootstrap value was set at 10,000 re-samplings, and the model setting was amino acid: Poisson correction with predicted gamma parameters set at 2.0. Accession numbers of the sequences used were: *Ta* [GenBank: GQ851928]; *Sb* [TIGR EST assemblies: TA33290_4558]; *Zm* [GenBank: NP_001141750]; *Os* [MSU Rice Genome Annotation: LOC_Os06g27860]; *At* [GenBank: NP_172541]; *Pt* [UniProtKB: B9HTU7_POPTR]; *Vv* [UniProtKB: A7QY03_VITVI]; and *Rc* [UniProtKB: B9SPJ5_RICCO].

Southern blot hybridisation

A 371 bp fragment of the *TaPHS1* gene was used as the template for the synthesis of an α-^{32}P dCTP labelled probe that was hybridised to a Chinese Spring nullisomic-tetrasomic membrane as per Lloyd *et al.* (Lloyd *et al.* 2007). Autoradiography films were developed using an AGFA CP1000 Developer (AGFA, Nunawading, VIC, Australia). For *in silico* mapping experimental procedures refer to Additional Information File 1.

Q-PCR

Q-PCR was conducted in triplicate according to Crismani *et al.* (Crismani *et al.* 2006). Amplification of products was completed using gene specific Q-PCR primers (see Table S1 in Additional Information File 1). The optimal acquisition temperature for *TaPHS1* was 80°C.

Protein analysis

The *TaPHS1* insert within the pCR8/GW/TOPO vector was cloned into a pDEST17 expression plasmid (Invitrogen) according to the manufacturer's LR clonase protocol. BL-21 A1 *E. coli* were transformed with the pDEST17-*TaPHS1* ORF vector, and protein production was induced with 0.4% L-(+)-arabinose (w/v) (Sigma-Aldrich, St Louis, MO, USA). Production of four partial *Ta*PHS1 peptides corresponding to the four conserved regions identified in this study were also performed as described above using DNA inserts encoding these regions. Protein isolation and purification was performed using nickel-nitrilotriacetic acid (Ni-NTA) beads (Qiagen, Clifton, VIC, Australia) according to the manufacturer's extraction protocols. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE) was performed using NuPAGE Novex 4–12% Bis-Tris 7 cm mini-

gels (Invitrogen) according to the manufacturer's protocol. Staining and destaining of gels were performed as previously reported (Wang *et al.* 2007).

The identity of the recombinant *Ta*PHS1 protein was confirmed by ion trap liquid chromatography-electrospray ionisation tandem mass spectrometry (LC-MS/MS). Gel slices containing the recombinant *Ta*PHS1 protein were washed with 100 mM ammonium bicarbonate, dried, rehydrated with 100mM ammonium carbonate and subjected to in-gel tryptic digestion. LC-MS/MS of the digested peptides was then conducted as reported by March *et al.* (March *et al.* 2007).

Polyclonal antibody production

Full-length recombinant *Ta*PHS1 protein was dissolved in 1× PBS (10 µg μL^{-1}), added with an equivalent amount of Freund's complete adjuvant (Sigma-Aldrich) and used for primary immunisation of two rats via subcutaneous injection. Three subsequent immunisations were administered in three-week intervals, with Freund's incomplete adjuvant (Sigma-Aldrich) added to the dissolved antigen in 1× PBS. All immunisation doses contained 200 µg of *Ta*PHS1 antigen. Immune sera was collected 10.5 weeks after the first injection.

Competitive DNA binding assay

Recombinant full length *Ta*PHS1 and the four partial peptides extracted under native conditions were quantified using the Bradford assay (Bradford 1976). Competitive DNA binding assays were conducted as described by Pezza *et al.* (Pezza *et al.* 2006) with modifications as per Khoo *et al.* (Khoo *et al.* 2008). The DNA binding abilities of *Ta*PHS1 and its partial peptides were tested with ΦX174

circular single-stranded DNA (ssDNA) (virion) (30 µM per nucleotide) (New England Biolabs, Beverly, MA, USA) and ΦX174 linear double-stranded DNA (dsDNA) (RFI form *Pst1*-digested) (15 µM per base pair) (New England Biolabs).

Fluorescence immunolocalisation

Fluorescence immunolocalisation of *Ta*ASY1 and *Ta*PHS1 was performed as per Franklin *et al.* (Franklin *et al.* 1999) and Boden *et al.* (Boden *et al.* 2009) with the following changes: anthers were fixed with 2% paraformaldehyde and cells permeabilised for 3 h. For detecting the localisation pattern of *Ta*PHS1, a rat anti-*Ta*PHS1 antibody (1:100) and an AlexaFluor® 488 conjugated donkey anti-rat antibody (1:50; Molecular Probes, Invitrogen) was used. Optical sections (90–120 per nucleus) of meiocytes were collected using a Leica TCS SP5 Spectral Scanning Confocal Microscope (Leica Microsystems, http://www.leica-microsystems.com/) equipped with an oil immersion HCX Plan Apochromat 63 × /1.4 lens, a 405 nm pulsed laser and an Argon laser using an excitation wavelength of 468 nm. Images were processed using Leica Application Suite Advanced Fluorescence (LAS-AF; version 1.8.2, build 1465, Leica Microsystems) software to generate maximum intensity projections of each nucleus.

Authors' contributions

KK performed all experimental procedures and drafted the manuscript. KK, JA and AA participated in the design of the study. JA and AA drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This research was supported in part by the Australian government under the

Australia-India Strategic Research Fund (AISRF) and the School of Agriculture,

Food & Wine, The University of Adelaide.

References

1. Armstrong SJ, Franklin FCH, Jones GH: **Nucleolus-associated telomere clustering and pairing precede meiotic chromosome synapsis in *Arabidopsis thaliana*.** *Journal of Cell Science* 2001, **114:**4207-4217.
2. Chen YK, Leng CH, Olivares H, Lee MH, Chang YC, Kung WM, Ti SC, Lo YH, Wang AHJ, Chang CS, *et al*: **Heterodimeric complexes of Hop2 and Mnd1 function with Dmc1 to promote meiotic homolog juxtaposition and strand assimilation.** *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101:**10572-10577.
3. Higgins JD, Sanchez-Moran E, Armstrong SJ, Jones GH, Franklin FCH: **The Arabidopsis synaptonemal complex protein ZYP1 is required for chromosome synapsis and normal fidelity of crossing over.** *Genes & Development* 2005, **19:**2488-2500.
4. Kerzendorfer C, Vignard J, Pedrosa-Harand A, Siwiec T, Akimcheva S, Jolivet S, Sablowski R, Armstrong S, Schweizer D, Mercier R, Schlogelhofer P: **The *Arabidopsis thaliana* MND1 homologue plays a key role in meiotic homologous pairing, synapsis and recombination.** *Journal of Cell Science* 2006, **119:**2486-2496.

5. Martínez-Pérez E, Shaw P, Aragon-Alcaide L, Moore G: **Chromosomes form into seven groups in hexaploid and tetraploid wheat as a prelude to meiosis.** *Plant Journal* 2003, **36:**21-29.
6. Able JA, Crismani W, Boden SA: **Understanding meiosis and the implications for crop improvement.** *Functional Plant Biology* 2009, **36:**575-588.
7. Able JA, Langridge P: **Wild sex in the grasses.** *Trends in Plant Science* 2006, **11:**261-263.
8. Able JA, Langridge P, Milligan AS: **Capturing diversity in the cereals: many options but little promiscuity.** *Trends in Plant Science* 2007, **12:**71-79.
9. Moore G, Shaw P: **Improving the chances of finding the right partner.** *Current Opinion in Genetics & Development* 2009, **19:**99-104.
10. Riley R, Chapman V: **Genetic control of the cytologically diploid behaviour of hexaploid wheat.** *Nature* 1958, **182:**713-715.
11. Sears ER: **Induced mutant with homoeologous pairing in common wheat.** *Canadian Journal of Genetics and Cytology* 1977, **19:**585-593.
12. Martinez-Pérez E, Shaw P, Moore G: **The *Ph1* locus is needed to ensure specific somatic and meiotic centromere association.** *Nature* 2001, **411:**204-207.
13. Prieto P, Shaw P, Moore G: **Homologue recognition during meiosis is associated with a change in chromatin conformation.** *Nature Cell Biology* 2004, **6:**906-908.
14. Colas I, Shaw P, Prieto P, Wanous M, Spielmeyer W, Mago R, Moore G: **Effective chromosome pairing requires chromatin remodeling at the onset of meiosis.** *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105:**6075-6080.
15. Holm PB: **Chromosome-pairing and synaptonemal complex-formation in hexaploid wheat, nullisomic for chromosome-5B.** *Carlsberg Research Communications* 1988, **53:**91-110.
16. Holm PB, Wang XZ: **The effect of chromosome-5B on synapsis and chiasma formation in wheat, *Triticum aestivum* cv Chinese Spring.** *Carlsberg Research Communications* 1988, **53:**191-208.

17. Al-Kaff N, Knight E, Bertin I, Foote T, Hart N, Griffiths S, Moore G: **Detailed dissection of the chromosomal region containing the *Ph1* locus in wheat *Triticum aestivum*: With deletion mutants and expression profiling.** *Annals of Botany* 2008, **101:**863-872.

18. Griffiths S, Sharp R, Foote TN, Bertin I, Wanous M, Reader S, Colas I, Moore G: **Molecular characterization of *Ph1* as a major chromosome pairing locus in polyploid wheat.** *Nature* 2006, **439:**749-752.

19. Bovill WD, Priyanka D, Sanjay K, Able JA: **Whole genome approaches to identify early meiotic gene candidates in cereals.** *Functional & Integrative Genomics* 2009, **9:**219-229.

20. Boden SA, Langridge P, Spangenberg G, Able JA: ***Ta*ASY1 promotes homologous chromosome interactions and is affected by deletion of *Ph1*.** *Plant Journal* 2009, **57:**487-497.

21. Boden SA, Shadiac N, Tucker EJ, Langridge P, Able JA: **Expression and functional analysis of *Ta*ASY1 during meiosis of bread wheat (*Triticum aestivum*).** *BMC Molecular Biology* 2007, **8:**Article No.: 65.

22. Caryl AP, Armstrong SJ, Jones GH, Franklin FCH: **A homologue of the yeast *HOP1* gene is inactivated in the Arabidopsis meiotic mutant *asy1*.** *Chromosoma* 2000, **109:**62-71.

23. Nonomura KI, Nakano M, Murata K, Miyoshi K, Eiguchi M, Miyao A, Hirochika H, Kurata N: **An insertional mutation in the rice *PAIR2* gene, the ortholog of Arabidopsis *ASY1*, results in a defect in homologous chromosome pairing during meiosis.** *Molecular Genetics and Genomics* 2004, **271:**121-129.

24. Ross KJ, Fransz P, Armstrong SJ, Vizir I, Mulligan B, Franklin FCH, Jones GH: **Cytological characterization of four meiotic mutants of Arabidopsis isolated from T-DNA-transformed lines.** *Chromosome Research* 1997, **5:**551-559.

25. Bleuyard JY, Gallego ME, Savigny F, White CI: **Differing requirements for the Arabidopsis Rad51 paralogs in meiosis and DNA repair.** *Plant Journal* 2005, **41:**533-545.

26. Li J, Harper LC, Golubovskaya I, Wang CR, Weber D, Meeley RB, McElver J, Bowen B, Cande WZ, Schnable PS: **Functional analysis of**

maize RAD51 in meiosis and double-strand break repair. *Genetics* 2007, **176:**1469-1482.

27. Pawlowski WP, Golubovskaya IN, Timofejeva L, Meeley RB, Sheridan WF, Cande WZ: **Coordination of meiotic recombination, pairing, and synapsis by PHS1.** *Science* 2004, **303:**89-92.
28. Ronceret A, Doutriaux MP, Golubovskaya IN, Pawlowski WP: **PHS1 regulates meiotic recombination and homologous chromosome pairing by controlling the transport of RAD50 to the nucleus.** *Proceedings of the National Academy of Sciences of the United States of America* 2009, **106:**20121-20126.
29. Suzuki M: **SPXX, a frequent sequence motif in gene regulatory proteins.** *Journal of Molecular Biology* 1989, **207:**61-84.
30. de Carvalho CE, Colaiacovo MP: **SUMO-mediated regulation of synaptonemal complex formation during meiosis.** *Genes & Development* 2006, **20:**1986-1992.
31. Katti MV, Sami-Subbu R, Ranjekar PK, Gupta VS: **Amino acid repeat patterns in protein sequences: Their diversity and structural-functional implications.** *Protein Science* 2000, **9:**1203-1209.
32. Yoder MD, Lietzke SE, Jurnak F: **Unusual structural features in the parallel beta-helix in pectate lyases.** *Structure* 1993, **1:**241-251.
33. Carmo-Fonseca M, Mendes-Soares L, Campos I: **To be or not to be in the nucleolus.** *Nature Cell Biology* 2000, **2:**E107-E112.
34. Olson MOJ, Dundr M, Szebeni A: **The nucleolus: an old factory with unexpected capabilities.** *Trends in Cell Biology* 2000, **10:**189-196.
35. San-Segundo PA, Roeder GS: **Pch2 links chromatin silencing to meiotic checkpoint control.** *Cell* 1999, **97:**313-324.
36. Franklin AE, McElver J, Sunjevaric I, Rothstein R, Bowen B, Cande WZ: **Three-dimensional microscopy of the Rad51 recombination protein during meiotic prophase.** *Plant Cell* 1999, **11:**809-824.
37. Jackson N, Sanchez-Moran E, Buckling E, Armstrong SJ, Jones GH, Franklin FCH: **Reduced meiotic crossovers and delayed prophase I progression in *At*MLH3-deficient *Arabidopsis*.** *Embo Journal* 2006, **25:**1315-1323.

38. Pawlowski WP, Cande WZ: **Coordinating the events of the meiotic prophase.** *Trends in Cell Biology* 2005, **15:**674-681.

39. Rost B, Yachdav G, Liu JF: **The PredictProtein server.** *Nucleic Acids Research* 2004, **32:**W321-W326.

40. Saitou N, Nei M: **The neighbor-joining method - a new method for reconstructing phylogenetic trees.** *Molecular Biology and Evolution* 1987, **4:**406-425.

41. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0.** *Molecular Biology and Evolution* 2007, **24:**1596-1599.

42. Lloyd AH, Milligan AS, Langridge P, Able JA: ***TaMSH7*: a cereal mismatch repair gene that affects fertility in transgenic barley (*Hordeum vulgare L.*).** *BMC Plant Biology* 2007, **7:**(20 December 2007).

43. Crismani W, Baumann U, Sutton T, Shirley N, Webster T, Spangenberg G, Langridge P, Able JA: **Microarray expression analysis of meiosis and microsporogenesis in hexaploid bread wheat.** *Bmc Genomics* 2006, **7**.

44. Wang XC, Li XF, Li YX: **A modified Coomassie Brilliant Blue staining method at nanogram sensitivity compatible with proteomic analysis.** *Biotechnology Letters* 2007, **29:**1599-1603.

45. March TJ, Able JA, Schultz CJ, Able AJ: **A novel late embryogenesis abundant protein and peroxidase associated with black point in barley grains.** *Proteomics* 2007, **7:**3800-3808.

46. Bradford MM: **Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding.** *Analytical Biochemistry* 1976, **72:**248-254.

47. Pezza RJ, Petukhova GV, Ghirlando R, Camerini-Otero RD: **Molecular activities of meiosis-specific proteins Hop2, Mnd1, and the Hop2-Mnd1 complex.** *Journal of Biological Chemistry* 2006, **281:**18426-18434.

48. Khoo KHP, Jolly HR, Able JA: **The *RAD51* gene family in bread wheat is highly conserved across eukaryotes, with *RAD51A* upregulated during early meiosis.** *Functional Plant Biology* 2008, **35:**1267-1277.

49. Felsenstein J: **Confidence-limits on phylogenies - An approach using the bootstrap.** *Evolution* 1985, **39:**783-791.

50. Zuckerkandl E, Pauling L: **Evolutionary divergence and convergence in proteins.** In *Evolving genes and proteins.* Edited by Bryson V, Vogel HJ. New York: Academic Press; 1965: 97-166.

Figures

Figure 1 - The PHS1 amino acid sequence is well-conserved across plant species.

(A) Alignment of *Ta*PHS1 with four homologues shows high levels of sequence conservation. Amino acid positions that are conserved across at least three species, using the *Ta*PHS1 amino acid sequence as the reference, are denoted by a '+' above. Four conserved regions were identified within the PHS1 protein and may represent functional domains: Region 1 (unbroken line), Region 2 (dashed line), Region 3 (dotted line) and Region 4 (dashed-dotted line). Oligopeptide repeat units (denoted by box) unique to *Ta*PHS1 are also highlighted. *Ta* – *Ta*PHS1, *Sb* – *Sb*PHS1, *Zm* – *Zm*PHS1, *Os* – *Os*PHS1 and *At* – *At*PHS1-like (see Methods for accession numbers). (B) The evolutionary history of *Ta*PHS1 was inferred using the Neighbor-Joining method (Saitou and Nei 1987). (C) Three additional PHS1/PHS1-like sequences were obtained through Hidden Markov Model and MaxHom searches and were also assessed using phylogenetics. *Vv* – *Vv*PHS1, *Rc* – *Rc*PHS1, *Pt* – *Pt*PHS1. The reliability of the internal branches of the trees (B, C) were assessed with 10,000 bootstrap re-samplings (Felsenstein 1985), with the confidence probabilities shown next to the branches. The trees are drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling 1965) and are in the units of the number of amino acid substitutions per site. There were

a total of 442 positions in the final datasets. Phylogenetic analyses were conducted in MEGA4 (Tamura *et al.* 2007).

Figure 2 - *TaPHS1* is located on chromosome group 7 of wheat.
Membranes prepared with DNA from nullisomic (N) – tetrasomic (T) wheat lines of Chinese Spring (CS) were hybridised with a *TaPHS1*-specific probe showing that there is a copy on the A, B and D genomes (indicated by black arrowheads).

Figure 3 - *Ta*PHS1 interacts with DNA *in vitro.*
Although *in silico* analysis of the *Ta*PHS1 amino acid sequence revealed no known DNA-binding domains, *Ta*PHS1 has DNA-binding ability indicating it possesses a novel/uncharacterised DNA-binding domain within Region 1. Using competitive DNA-binding assays with equivalent amounts of single- and double-stranded DNA, it appears that *Ta*PHS1 preferentially binds single-stranded DNA (ssDNA). At higher concentrations, it also interacts with double-stranded DNA (dsDNA). (A) Full-length *Ta*PHS1, (B) Region 1 peptide, (C) Region 2 peptide, (D) Region 3 peptide, (E) Region 4 peptide.

Figure 4 - Q-PCR profiling of *TaPHS1* shows that it is expressed during meiosis.
While the amount of mRNA transcript is relatively low, it has higher levels of expression during pre-meiosis when compared to the other stages of meiosis examined in Chinese Spring wild-type (black bar). In the *ph1b* mutant (open bar), *TaPHS1* is up-regulated between 1.5- and 2-fold across the time-points analysed. Data represent the means ±SE of three replicates. Units on the y-axis represent normalised mRNA transcript μL^{-1}.

Figure 5 - *Ta*PHS1 localisation during early meiosis in wild-type bread wheat.

(A) Telomere bouquet stage, (B) leptotene, (C) early-zygotene, (D) mid-zygotene, (E) late-zygotene, (F) pachytene, (G) diplotene. *Ta*PHS1 (green; left panel) localises to 4´-6-diamidino-2-phenylindole (DAPI)-stained chromatin (blue) as diffuse tracts and/or punctated foci. Middle panels show the *Ta*ASY1 signal (red), while the panels on the right show merged *Ta*PHS1, *Ta*ASY1 and DAPI. Arrowheads (white) represent the nucleolus. Scale bars, 7.5 µm.

Additional files

Additional Information File 1

MS Word document containing methods for *in silico* mapping of *TaPHS1* and a detailed list of primers used in this study.

Table S1 - Detailed list of primers used in this study.

Genome walking primers were used for amplification of the 5´ region of *TaPHS1* while open reading frame (ORF)-targeting primers were used to isolate the ORF sequence for protein production. Primers based on the 5´ genomic *TaPHS1* sequence were required to isolate the probe used for the Southern blot hybridisation. Q-PCR primers were used to determine the expression profile of *TaPHS1* in meiotic tissues of both wild-type and *ph1b* mutant plants. Abbreviations: F1, forward 1; R1, reverse 1; Reg1, region 1; Reg2, region 2; Reg3, region 3, Reg4, region 4; QF1, quantitative forward 1; QR1, quantitative reverse 1; SF1, Southern forward 1; SR1, Southern reverse 1.

Figure 1


(a)
Ta 1 MAGAGGRSRERLTSRAEEAAGGK----RRRQRWEVEFARYFAKPRRAPSTPPPPGLRYIS 56
Sb 1 MADASDNSMALVHARLAVSAVAPPRMLRQRQKWEVEYARYFGTPRRDPSAPPPPGLRHII 60
Zm 1 MADAADSSMALVHSSLADSVLTSPRTLRQGQKWEVEYARYFGTPRRDPTAAPPSGLRYIM 60
Os 1 MADADVRSGALLPARPTPQ--------RRRQKWAVEFARFFRTPRRDPSKPPPPGLRLVA 52
At 1 MAGSLTASNRRRNAEDSSE----------IYRWTIGFARFVHYP---SSPSPHPVLKPLG 47
Ta 57 RGKQLH--QGTWLLAASPAALCISRPTHSFAARVLTVSICDVVYEEHYVSILNFSWPQVA 114
Sb 61 RGVHRH--QGTWIPASCPASLCVSHPSLPSAVPVLTISIGDVVFEEHFVSILNFSWPQIT 118
Zm 61 RGVHRH--QGTWIPASCPASLCVCHPSLPSAVPVLTISIGDVVFEEHFVSILNFSWPQVT 118
Os 53 RGKLRH--HGTWLPAASPAALSISCPSQSFAVPVLTVSIGDVVFVRTDPAPLPR---GAF 107
At 48 KREQYHSPHGTWLSASSSTVSLHIVDELNRSDVILSVKLGQKVLEEHYISKLNFTWPQMS 107
Ta 115 CVTECPVRGSRVVFVSFCDRSKQIQKFAVRFPRLSDAESFLNSVIVKELSS-NTMDIMPS 173
Sb 119 CVMQCPIRGSRVVFVSFCDKSKQIQKFAVRFPQLCDAELFLSCVKECSC---ETMDMIPS 175
Zm 119 CVTQCPIRGSRVVFVSFCDKFKQIQKFAVRFPQPCDAESFLSCV-ECSCGSSGTMDIIPF 177
Os 108 CVHSQFFVASGYMCDTMPNKWEQSG---------VSLNAFLYGLFTKECST-ETMDIRPS 157
At 108 CVSGFPSRGSRAIFVTYMD3ANQIQKFALRFSTCDAALEFVEALKEKIKGL-KEASTQNQ 166
Ta 174 GSDYMCELEDSSSSEYIPSNGLQYRPDE----------AVSFEEPTSD--HRTDAPAVG 220
Sb 176 GSDYVCE--DSSTSEYIAYNGLHHRPDD----------ASGFEEQSSD--HTIDEPTMS 220
Zm 178 GSDYVCE--DSSASEYIVSNGLHHRLDD----------ASNLEEQCFD--HTIDEPPMN 222
Os 158 CSDYLCE--DSSASEYIASSGIHQSFEEPDQPVHRTETPALCYHAEPDEPIHRTEAPALS 215
At 167 KNKTRCD--VSFQSDYNPSDAIIPRATQK----------EPNMVRPLNSYVPEMLPRIVY 214
Ta 221 --------YHMEPDQPVLQSPIATNINSIYSGFPEGYSGFPEGYSGFPEGYSGSVKIERD 272
Sb 221 --------YHEEQDQPVLEPLSASNTSNSYSAFPPSFNQMLT---------NCSIDYDQE 263
Zm 223 --------YHEETDQHVLEPLSASNTSNN-SAFPPSFNQMLK---------SCSIDYDQE 264
Os 216 QPETPSLRHHEAPEEPLLQPLLATNIDTVFSGFPPSFTDMLT-------QFSCKTEKDAE 268
At 215 -------EAQYQKSETRSEVSFQSDYNPSIEIFPRATEEEPNMVRFFDSSVPEVLPRPEY 267
Ta 273 GGPFPATITDHAPEKAYILDTRIDAAGGNSVADKGKGAGKEIDVSDVTRDILAGIETYGG 332
Sb 264 E-PCPLAASNHALQEVYALDTSHDVANEETTAGKGLDAGEGVDTSILTYDIMARIKTYMA 322
Zm 265 E-PCPLAASNHVLQEVYVLDTSHDVANEERTAGKGMDAAEGVDASILTYDLMARIKTYMA 323
Os 269 E-PYPVTATDHAPQEVSMLDTSHNVAISTTSAN-------EIDVNRETSDIMTRIKTYIS 320
At 268 EAGQALYPSQSTLNQIPSLPPSFTTLLSGCFPDSTLDAG--QTTVKQNPDLKSQILKYME 325
Ta 333 DDSFHDMLSKLDKAIDELGGDLSL---------- 356
Sb 323 DESFNDMLFKLDKVIDELGGDMSL---------- 346
Zm 324 DESFNDMLLKLDKAIDELGGDMSL---------- 347
Os 321 DGAFHDMLFKLERVIDELGGDLSL---------- 344
At 326 DSSFQDMLQKVERIIDEIGDRCVGSTIGRSSKIYIPLCFHNNCIHPFKDDCWCCLAAGTK 385
Ta ----------
Sb ----------
Zm ----------
Os ----------
At 386 KDWCWLEKDFPDAKELCMKTCTRKI 410
(b)
87
Ta
Os
Sb
87
Zm
At
0.2
(c)
94
Ta
Os
Sb
91
Zm
Vv
At
99
95
Rc
65
Pt
0.2

Figure 2

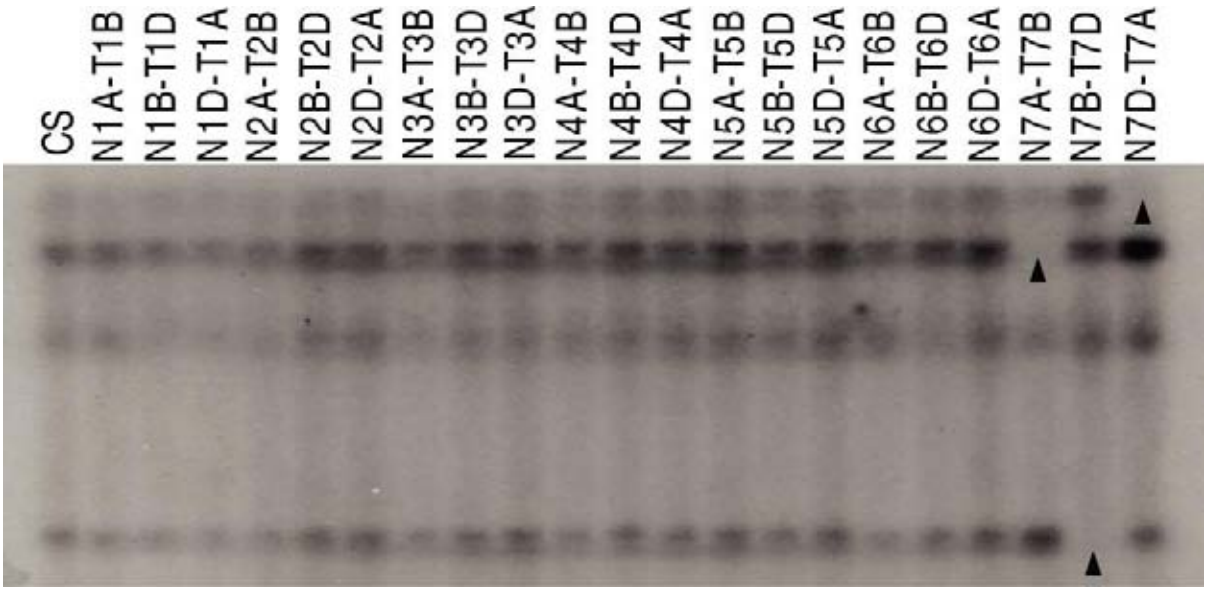

CS
N1A-T1B
N1B-T1D
N1D-T1A
N2A-T2B
N2B-T2D
N2D-T2A
N3A-T3B
N3B-T3D
N3D-T3A
N4A-T4B
N4B-T4D
N4D-T4A
N5A-T5B
N5B-T5D
N5D-T5A
N6A-T6B
N6B-T6D
N6D-T6A
N7A-T7B
N7B-T7D
N7D-T7A

Figure 3

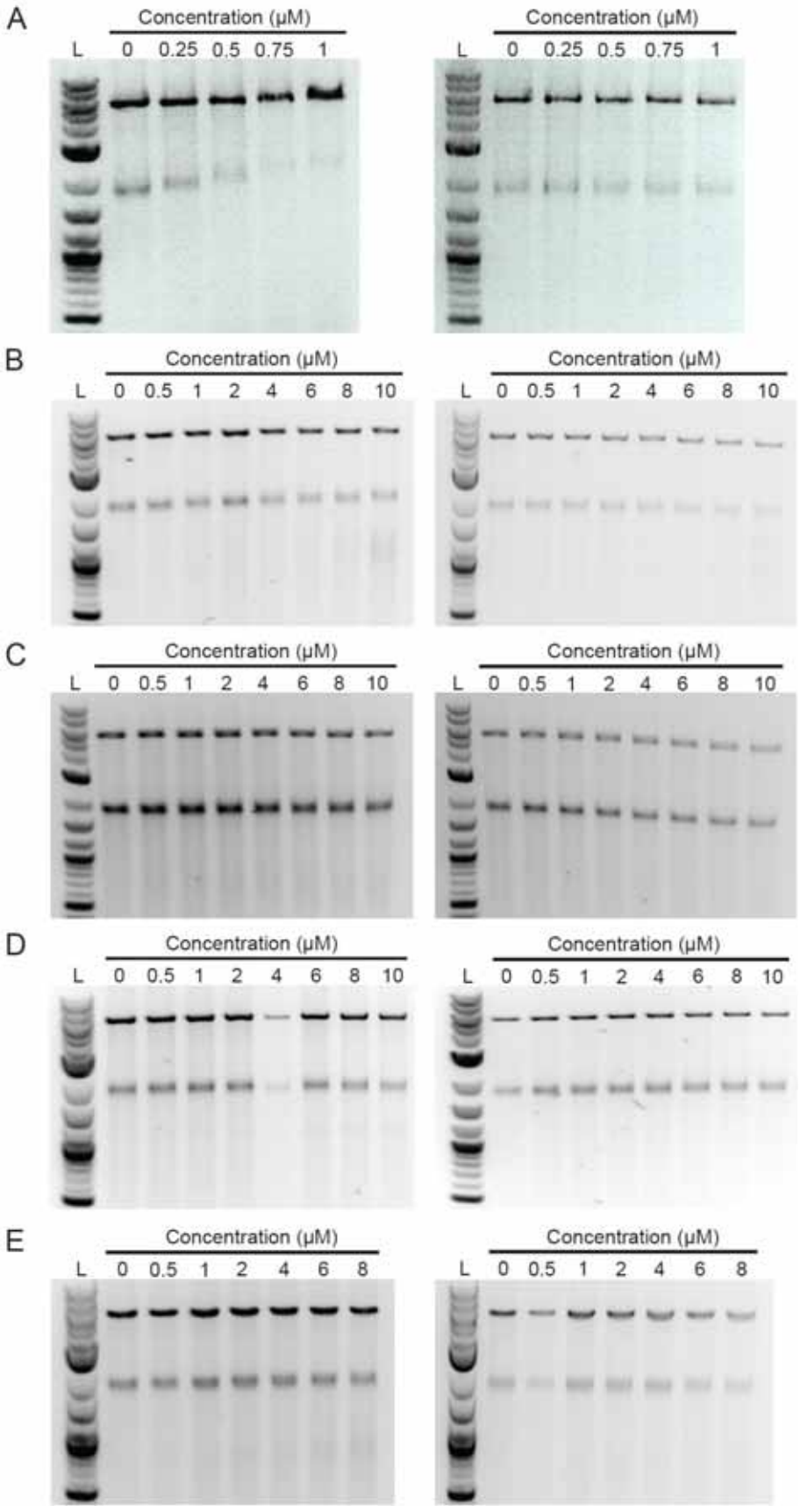

A
Concentration (μM)
L 0 0.25 0.5 0.75 1
Concentration (μM)
L 0 0.25 0.5 0.75 1
B
Concentration (μM)
L 0 0.5 1 2 4 6 8 10
Concentration (μM)
L 0 0.5 1 2 4 6 8 10
C
Concentration (μM)
L 0 0.5 1 2 4 6 8 10
Concentration (μM)
L 0 0.5 1 2 4 6 8 10
D
Concentration (μM)
L 0 0.5 1 2 4 6 8 10
Concentration (μM)
L 0 0.5 1 2 4 6 8 10
E
Concentration (μM)
L 0 0.5 1 2 4 6 8
Concentration (μM)
L 0 0.5 1 2 4 6 8

Figure 4

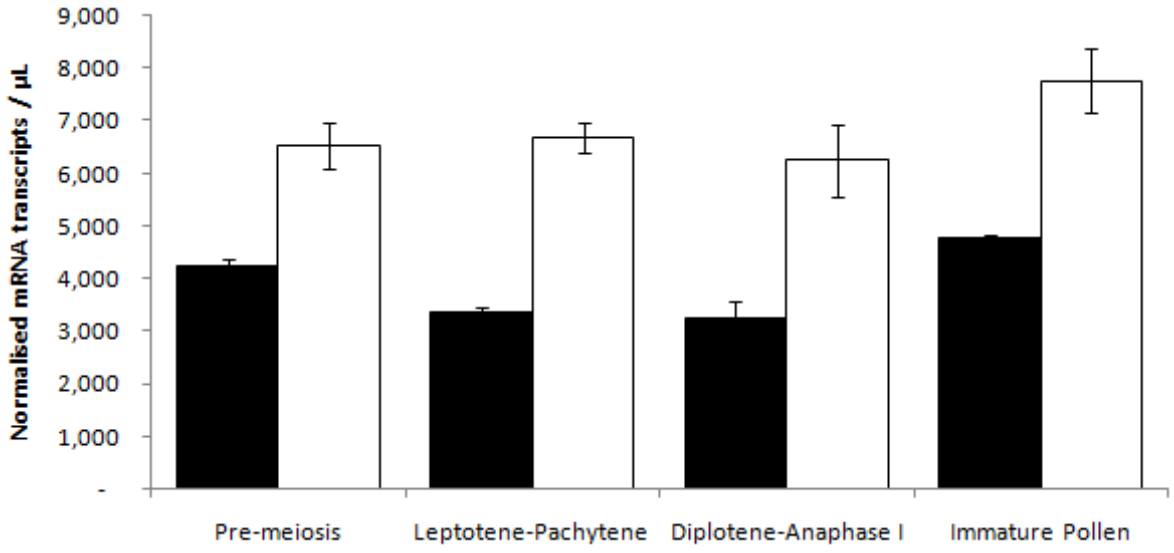
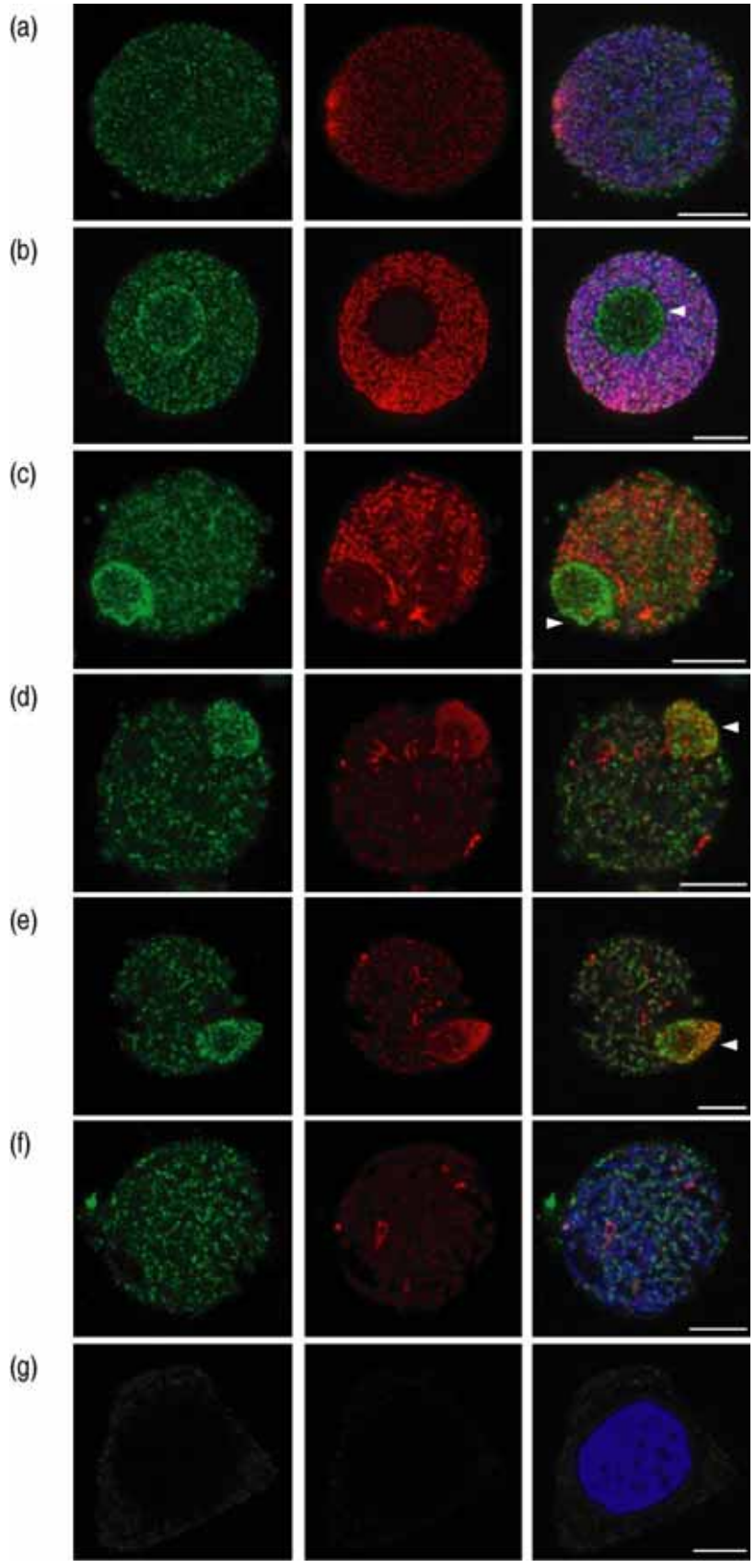

Normalised mRNA transcripts / µL
9,000
8,000
7,000
6,000
5,000
4,000
3,000
2,000
1,000
-
Pre-meiosis
Leptotene-Pachytene
Diplotene-Anaphase I
Immature Pollen

Figure 5



Additional Information File 1

Kelvin H.P. Khoo, Amanda J Able, & Jason A. Able (2011).

Poor homologous synapsis 1 (PHS1) interacts with chromatin but does not co-localise with *Ta*ASY1 during early meiosis in bread wheat.

Methods

In silico mapping

In silico mapping of *TaPHS1* was conducted by first identifying the bacterial artificial chromosome (BAC) [GenBank accession: AP003528] containing the *OsPHS1* genomic sequence [MSU Rice Genome Annotation: LOC_Os06g27860.1] using the Gramene rice database (http://www.gramene.org/Oryza_sativa_japonica/Location/Chromosome?r=6:1-31246789; Gramene Release 29; accessed 20th April 2009). A single-long-sequence (SLS) BLASTN search was performed using the rice BAC sequence as the query. A total of 116 wheat ESTs were identified in the SLS search. Using the rice deletion bin database (http://wheat.pw.usda.gov/wEST/binmaps/), wheat chromosome group 7 shares a high degree of similarity to rice chromosome 6. Each of the 116 identified wheat ESTs was then compared to the available markers on rice chromosome 6 to determine whether the location of *TaPHS1* on wheat chromosome group 7 could be refined.

Table S1 - Detailed list of primers used in this study. Genome walking primers were used for amplification of the 5´ region of *TaPHS1* while open reading frame (ORF)-targeting primers were used to isolate the ORF sequence for protein production. Primers based on the 5´ genomic *TaPHS1* sequence were required to

isolate the probe used for the Southern blot hybridisation. Q-PCR primers were used to determine the expression profile of *TaPHS1* in meiotic tissues of both wild-type and *ph1b* mutant plants. Abbreviations: F1, forward 1; R1, reverse 1; Reg1, region 1; Reg2, region 2; Reg3, region 3, Reg4, region 4; QF1, quantitative forward 1; QR1, quantitative reverse 1; SF1, Southern forward 1; SR1, Southern reverse 1.

Primer name	**Primer sequence (5´ → 3´)**	**T_m (°C)**
***TaPHS1* gene isolation primers**		
*TaPHS1*_F1	CATTTTCGGCGTCATCGTCGTCG	65
*TaPHS1*_R1	CTACAGTGACAAGTCGCCACCCAGTTCA	
***TaPHS1* gene expression primers (to amplify ORF for protein production)**		
*TaPHS1*_F2	ATGGCGGGCGCCGGC	62
*TaPHS1*_R1	CTACAGTGACAAGTCGCCACCCAGTTCA	
***TaPHS1* peptide primers (to amplify conserved regions for protein production)**		
*TaPHS1*_Reg1_F1	CGGCGGAGGCAGAGGTGG	65
*TaPHS1*_Reg1_R1	GAAGGAGTGGGTGGGGCGAGA	
*TaPHS1*_Reg2_F1	GTCTACGAGGAGCACTATGTATCTATCCTCAACTT	66
*TaPHS1*_Reg2_R1	AGGGAAACGTACAGCAAACTTCTGGAT	
*TaPHS1*_Reg3_F1	AAGGAACTCTCAAGCAACACCAT	59
*TaPHS1*_Reg3_R1	CGCTTCATCTGGCCTGTATTG	
*TaPHS1*_Reg4_F1	GGCGGGGACGACTCTTTTCAT	62
*TaPHS1*_Reg4_R1	Primer used was *TaPHS1*_R1	
Plasmid vector sequencing primers		
T7 (pGEM®-T Easy)	TAATACGACTCACTATAGGG	50
SP6 (pGEM®-T Easy)	ATTTAGGTGACACTATAG	
GW1 (pCR®8/GW/TOPO®)	GTTGCAACAAATTGATGAGCAATGC	50
GW2 (pCR®8/GW/TOPO®)	GTTGCAACAAATTGATGAGCAATTA	
Quantitative real-time PCR (Q-PCR) primers		
*TaPHS1*_QF1	CACTCGGATTGATGCTGCTG	55
*TaPHS1*_QR1	TGACAAGTCGCCACCCAGTT	
Southern blot probe primers		
*TaPHS1*_SF1	TGAACTGAATTGTGTCGGATGAA	54
*TaPHS1*_SR1	AAAGCTCACGAACACCACCCT	

Chapter 4 – Addendum

Research conducted but not presented in Khoo *et al.* (2011).

4.1 – Introduction

Based on the limited literature available on the PHS1 protein (Pawlowski *et al.* 2004, Ronceret *et al.* 2009, Khoo *et al.* 2011), further research was conducted in an effort to add to the current knowledge base. Among the key findings reported to date are that disruption of PHS1 function leads to impaired transport of RAD50 from the cytoplasm to the nucleus (Ronceret *et al.* 2009) and that there are two regions of sequence conservation (Conserved Regions 1 and 2 – CR1 and CR2) in the amino acid sequences of PHS1 proteins amongst various species (Pawlowski *et al.* 2004, Ronceret *et al.* 2009).

The results of immunolocalisation assays performed in this study (Khoo *et al*. 2011, *submitted*) show that *Ta*PHS1 is present within the nucleus where it associates with chromatin and accumulates within the nucleolus during early meiosis. This localisation pattern is similar to that reported by Ronceret *et al.* (2009) where the PHS1 proteins in both maize and Arabidopsis are found within the cytoplasm and within the nucleus albeit as just a few foci. Immunolocalisation of the RAD50 protein in both maize and Arabidopsis *phs1* mutants (*phs1-0* and *Atphs1-1*) showed dramatically reduced numbers of RAD50 foci within the nuclei leading Ronceret and colleagues to conclude that PHS1 must somehow, directly or indirectly, regulate transport of RAD50.

Work reported in this addendum was aimed at uncovering the nature of the PHS1 interaction with RAD50. To this end, a yeast-two-hybrid (Y2H) experiment was attempted.

4.2 – Materials & Methods

4.2.1 – Primer design

4.2.1.1 – *PHS1* primer design

In preparation for ligation into the pGADT7 prey plasmid (MatchMaker™ Gold Yeast-Two-Hybrid kit, Clontech Inc.), a set of primers were designed to attach an *Eco*R1 cut-site to the 5′end and a *Cla*1 cut-site to the 3′end of the *TaPHS1* sequence (Table 4.1).

4.2.1.2 – *RAD50* primer design

Forward and reverse primers for the amplification of *RAD50* were designed based on the *OsRAD50* sequence (Rice Locus Identifier – LOC_Os02g29464) (Table 4.1). In preparation for ligation into the pGBKT7 bait plasmid (MatchMaker™ Gold Yeast-Two-Hybrid kit, Clontech Inc., Mountain View, California, United States of America), two secondary sets of primers were designed to add enzyme cut-sites to the *TaRAD50* clone. The first set added an *Eco*R1 cut-site to the 5′end of the *TaRAD50* sequence and a *Pst*1 cut-site to the 3′end of the *TaRAD50* sequence while the second set added an *Nde*1 site to the 5′end and an *Eco*R1 site to the 3′end (Table 4.1). In addition, multiple sets of nested primers were designed to sequence the *RAD50* sequence isolated (Table 4.1).

Table 4.1 – Complete list of primers used in this study.

Primer name	Primer sequence (5′ → 3′)	T_m (°C)
pGADT7-*TaPHS1* plasmid construction primers		
TaPHS1-F1-*Eco*R1	TGAATTCATGGCGGGCGCCGGC	62
TaPHS1-R1-*Pst*1	TATCGATCTACAGTGACAAGTCGCCACCCAGTTCA	
***TaRAD50* isolation primers**		
TaRAD50-F1	ATGAGCACGGTGGACAAGATGCTGATC	64
TaRAD50-R1	GTCAAATATCTCTTGGGCTTCTATTTTGCTGT	
***TaRAD50* nested primers**		
TaRAD50-nF1	TTAGCACCAAAGCAACTGCTAGAAGTACATA	59
TaRAD50-nR1	ATGAGTAAGTCGATCCACTTCAGCCTTT	58
TaRAD50-nF2	CAAGATCAAGAGAAGTCAGACGCCTTAA	58
TaRAD50-nR2	GATTTCCACTCCATCCCTGTCCAT	57
pGBKT7-*TaRAD50* plasmid construction primers		
TaRAD50-F1-*Eco*R1	AGAATTCATGAGCACGGTGGACAAGATGCT	62
TaRAD50-R1-*Pst*1	ACTGCAGTCAGTCAAATATCTCTTGGGCTTCTATTTTG	
TaRAD50-F1-*Nde*1	AAACATATGATGAGCACGGTGGACAAGATGCT	60
TaRAD50-R1-*Eco*R1	AAAGAATTCCTAGTCAAATATCTCTTGGGCTTCTATTTTG	

4.2.2 – Preparation of meiotic cDNA

4.2.2.1 – cDNA synthesis from purified meiotic RNA

Collection and staging of meiotic anthers was done as per section 2.2.1. RNA extraction and quantification from Chinese Spring tissue was performed as per section 3.2.3.3. Wild-type cDNA was synthesised from 1 μg of total RNA using an iScript Kit (BIO-RAD) according to the manufacturer's instructions.

4.2.3 – Preparation of *RAD50* and *PHS1* clones

4.2.3.1 – High fidelity amplification and purification of *RAD50* and *PHS1* from meiotic cDNA

All PCR reactions contained 0.2 μL of template meiotic cDNA, 3.2 μL of 5 mM dNTPs, 1 μL of 10 μM *RAD50* forward and reverse primers (refer to Table 4.1),

2.5 μL of 10× high fidelity PCR buffer +$MgCl_2$ (Roche), 5 μL of GC-rich solution (Roche), 0.2 μL of Fast Start *Taq* polymerase (Roche), and 11.9 μL of autoclaved nanopure water. PCR cycle conditions for *RAD50* were as follows; initial denaturation step at 95°C for 5 min, then 35 cycles of 96°C for 30 sec, 62°C for 30 sec, 72°C for 4 min, followed by a final extension step at 72°C for 10 min. The same protocol was used for *TaPHS1* with the following modifications: 1) *PHS1* primers were used; and 2) Elongation time for the PCR protocol was 1 min 15 sec. Agarose gel electrophoresis and gel purification of the separated PCR products was conducted as per section 3.2.3.6.

4.2.3.2 – Ligation of *RAD50* and *PHS1* PCR products and bacterial transformation

Ligation of the *RAD50* and *PHS1* PCR products were performed with pGEM®T-Easy (Promega Corporation, Sydney, New South Wales, Australia) as per the manufacturer's instructions. Transformation of *E. coli* was performed as per section 3.2.3.7 except that transformants were plated onto LB agar plates containing ampicillin (100 μg mL^{-1}). In addition, 5-bromo-4-chloro-3-indolyl P-D-galactopyranoside (X-Gal, 40 μL of 20 mg mL^{-1}) and isopropyl P-D-1-thiogalactopyranoside (IPTG, 100 μL of 100 μM) were added to each plate for selection of positive colonies based on blue/white colour selection.

4.2.3.3 – Colony PCR: identification of positive clones

Colonies were screened for presence of the insert using a PCR-based assay. Each PCR reaction contained 0.2 μL of template meiotic cDNA, 3.2 μL of 5 mM dNTPs, 1 μL of 10 μM forward and reverse primers (refer to Table 4.1), 2.5 μL of

10× high fidelity PCR buffer +$MgCl_2$ (Roche), 5 μL of GC-rich solution (Roche), 0.2 μL of Fast Start *Taq* polymerase (Roche), and 11.9 μL of autoclaved nanopure water.

Individual colonies were selected from plates using sterile 200 μL pipette tips, which were subsequently swirled in the PCR reaction mix. Colony PCR cycle conditions were the same as per section 4.2.3.1. Agarose gel electrophoresis and gel purification of the separated PCR products was conducted as per section 3.2.3.6.

Culturing of cells and isolation of the plasmids using a PureLink™ Quick Plasmid Miniprep Kit (Invitrogen) were performed as per section 3.2.3.8. Glycerol stocks were made for all positive clones by adding 1 mL of bacterial culture to 1 mL of 50% glycerol. The stocks were snap-frozen using liquid nitrogen and kept in the -80ºC freezer.

4.2.4 – Sequencing of *RAD50* and *PHS1* clones

4.2.4.1 – Sequencing PCR and clean-up

Sequencing PCR reactions and subsequent clean-up was done as per section 3.2.3.9 and 3.2.3.10. Samples were then submitted to AGRF for capillary separation using the ABI 3730 DNA Analyser (Applied Biosystems).

4.2.4.2 – Contig construction and sequence analysis of clones

Contig construction was performed as per section 3.2.3.11. Using both AlignX and Contig Express programs, sequence information from the *RAD50* clones derived from Chinese Spring were used to form a consensus sequence for

TaRAD50. The *TaRAD50* sequence was then compared to the four other annotated *RAD50* homologue sequences from various organisms to ensure that the correct complete coding sequence was isolated (Table 4.2). ClustalW alignments and phylogenetic analyses were conducted using MEGA4. Phylogenetic parameters used were: bootstrap value of 10000, pairwise-deletion, and amino acid p-distance. Conserved domain analyses of the predicted *Ta*RAD50 protein was conducted using the online NCBI Conserved Domain Architecture Retrieval Tool (CDART) (http://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi). The *TaRAD50* sequence was also checked to ensure that the enzyme cut-sites added using the secondary sets of the primers were present.

The *PHS1* clone sequence was compared to that of the *TaPHS1* CDS (GenBank - GQ851928) to ensure that the sequence was correct and that the enzyme cut-sites added were present.

4.2.5 – Construction of Y2H constructs

4.2.5.1 – Construction of pGADT7-*TaPHS1* activation domain prey plasmid

After quantification with a ND-1000 nanodrop machine (NanoDrop Technologies Inc.), 1 µg of both pGEM®T-Easy-*TaPHS1* plasmid and pGADT7 plasmids were double-digested with *Eco*R1 (NEB) and *Cla*1 (NEB) according to the manufacturer's protocol except that the digestion was performed for 5 hr. Agarose gel electrophoresis and gel purification of the DNA fragments were performed as per section 3.2.3.6.

The liberated *TaPHS1* insert was then ligated into the pGADT7 plasmid with T4 DNA ligase (NEB) according to the manufacturer's protocol at the recommended 3:1 insert-to-vector ratio. Bacterial transformation with the

pGADT7-*TaPHS1* plasmid, colony PCR screening, and plasmid mini-preps were performed as per section 4.2.3.2 and 4.2.3.3. Sequencing and downstream analyses of the pGADT7-*TaPHS1* plasmids were performed as per section 4.2.4.

4.2.5.2 – Construction of pGBKT7-*TaRAD50* binding domain bait plasmid

After quantification with a ND-1000 nanodrop machine (NanoDrop Technologies Inc.), 1 µg of both pGEM®T-Easy-*TaRAD50* plasmid and pGBKT7 plasmids were sequentially digested with *Eco*R1 (NEB) followed by either *Pst*1 (NEB) or *Nde*1 (NEB) according to the manufacturer's protocol except that the digestion was performed for 5 hr. Agarose gel electrophoresis and gel purification of the DNA fragments were performed as per section 3.2.3.6.

The liberated *TaRAD50* insert was then ligated into the pGADT7 plasmid with T4 DNA ligase (NEB) according to the manufacturer's protocol. Multiple insert-to-vector ratios (5:1, 4:1, 3:1, 2:1, 1:1) were used to facilitate ligation of the *TaRAD50* insert into pGBKT7. Additional optimisation included ligation of the *TaRAD50* insert liberated from pGEM®T-Easy using only *Eco*R1. In this case, pGBKT7 plasmid was also digested with only *Eco*R1 and the cut vector was then dephosphorylated with shrimp alkaline phosphatase (Invitrogen) according to the manufacturer's protocol to prevent recircularisation. In addition, ligation reactions with increased concentrations of T4 DNA ligase were also attempted.

4.3 – Results

4.3.1 – Construction of pGADT7-*TaPHS1* prey plasmid

Utilising the coding sequence of *TaPHS1* reported by Khoo *et al.* (2011) (refer to Appendix C2, Figure 1 for full-length *TaPHS1* ORF sequence), primers designed to add an *Eco*R1 and a *Cla*1 restriction enzyme cut-site 5´ and 3´ of the *TaPHS1* coding sequence respectively were successfully used to reamplify *TaPHS1* (Figure 4.1 A). Ligation of the *TaPHS1* product with the newly-added restriction sites into pGEM®T-Easy and subsequent bacterial plasmid propagation enabled high concentrations of the purified plasmid to be obtained. The *TaPHS1* insert was then liberated from the pGEM®T-Easy plasmid via double-digestions with *EcoR1* and *Cla1* restriction enzymes (Figure 4.1 B). This allowed the *TaPHS1* digested product to be directionally ligated into the pGADT7 plasmid that had been double-digested with the same restriction enzymes. Screening of bacterial colonies transformed with the pGADT7-*TaPHS1* plasmid via colony PCR identified one colony positive for the *TaPHS1* insert (Figure 4.1 C). This was further confirmed using Big Dye chemistry sequencing of the plasmid.

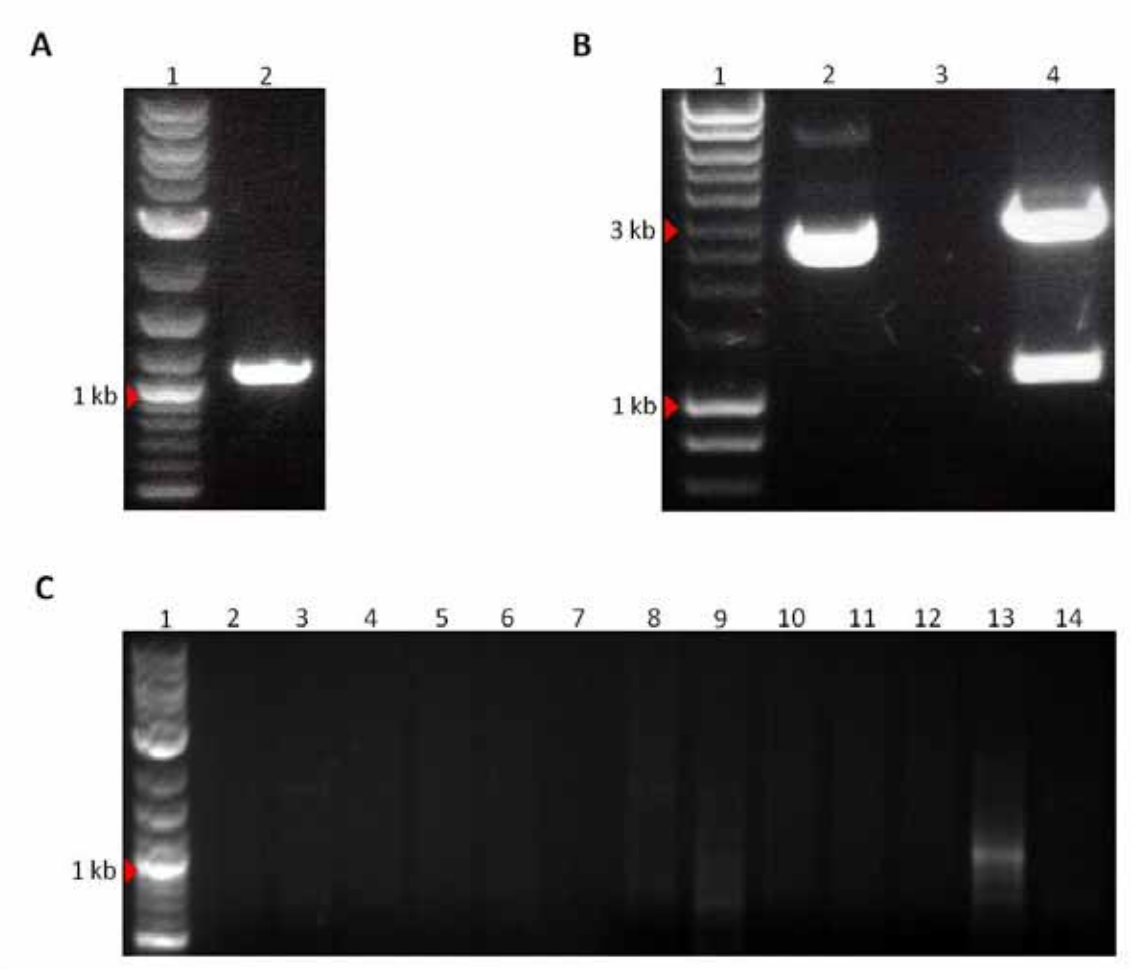


Figure 4.1 – Construction of the pGADT7-*TaPHS1* activation domain prey plasmid. A) The full-length *TaPHS1* ORF was successfully amplified and isolated. Lane 1, molecular weight marker; lane 2, *TaPHS1* amplified with 5´ *Eco*R1 cut-site adaptor primer and 3´ *Cla*1 cut-site adaptor primer. B) Double-digest of the pGEM®T-Easy-*TaPHS1* plasmid using *Eco*RI and *Cla*I liberated the 1085 bp *TaPHS1* insert for directional cloning into the pGADT7 plasmid that was also double-digested with the same enzymes. Lane 1, molecular weight marker; lane 2, undigested pGEM®T-Easy-*TaPHS1* plasmid control; lane 3, blank; lane 4, *Eco*R1 and *Cla*1 double-digested pGEM®T-Easy-*TaPHS1*. C) Colony PCR screening of bacterial colonies transformed with the pGADT7-*TaPHS1* plasmid identified a single positive colony (lane 13) that was then sequenced with Big Dye chemistry for confirmation.

4.3.2 – Isolation and bioinformatic analyses of *TaRAD50*

Upon successful amplification of a PCR product of the expected length, the product was cloned and sequenced. To ensure that the wheat *RAD50* homologue was correctly isolated, the wheat sequence was used in BLASTn searches of the NCBI and TIGR databases that resulted in positive hits for RAD50 homologues in

other organisms. The full-length coding sequence of *TaRAD50* was successfully isolated with a 3591 bp ORF (Figure 4.2 A) (refer to Appendix C2, Figure 2 for the full-length *TaRAD50* ORF sequence). Primers designed to add restriction enzyme cut-sites 5´ and 3´ of the *TaRAD50* coding sequence were then used to re-amplify the *TaRAD50* product (Figure 4.2 B).

TaRAD50 encodes a protein product with a predicted molecular weight of 152.227 kD and a pI of 7.76. The level of conservation between the predicted protein products of *TaRAD50* and its homologues are detailed in Table 4.2. cDART analysis of the predicted *Ta*RAD50 protein detected the presence of the RAD50 conserved domain within the sequence (Figure 4.3 A). As expected, *Ta*RAD50 shares much higher levels of similarity with its homologues in the plant kingdom compared to human and yeast equivalents. Phylogenetic analyses of the predicted protein sequences further confirmed this with clear branching of the three kingdoms seen within the neighbour-joining tree (Figure 4.3 B).

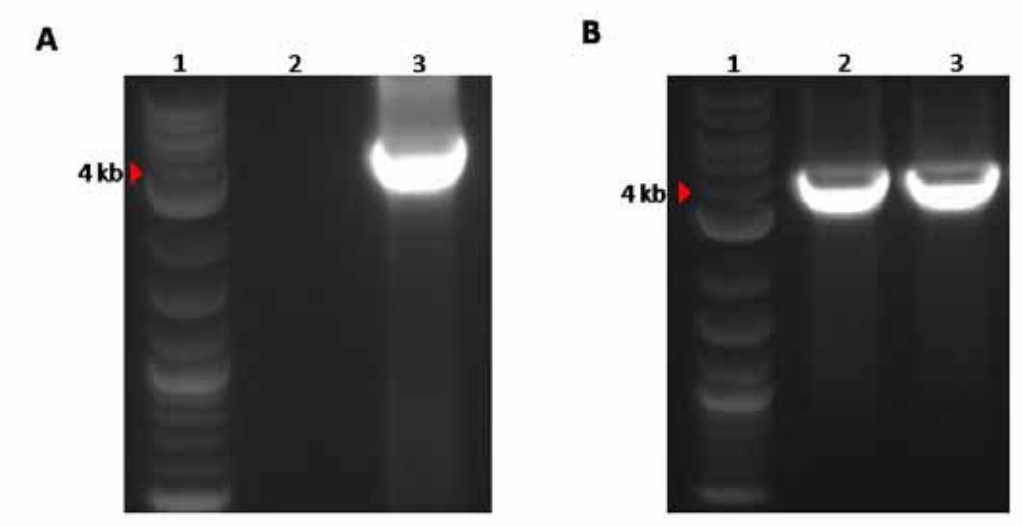


Figure 4.2 – Isolation and re-amplification of *TaRAD50*. A) The full-length *TaRAD50* ORF was successfully isolated from meiotic spike Chinese Spring cDNA. Lane 1, molecular weight marker; lane 2, negative control; lane 3, *TaRAD50*. B) Re-amplification of *TaRAD50* using primers designed to add restriction enzyme cut-sites to either end of the *TaRAD50* coding sequence was successful. Lane 1, molecular weight marker; lane 2, *TaRAD50* amplified with 5´ *Eco*R1 cut-site adaptor primer and 3´ *Pst*1 cut-site adaptor primer; lane 3, *TaRAD50* amplified with 5´ *Eco*R1 cut-site adaptor primer and 3´ *Nde*1 cut-site adaptor primer.

Table 4.2 – Summary of RAD50 sequence conservation. This table highlights the varying level of sequence conservation between five RAD50 homologues from three different kingdoms. Levels of conservation also varied between the two classes within the plant kingdom (dicotyledons and monocotyledons).

***RAD50* homologue**	**Accession number**	**Number of identical predicted amino acid residues**	**Pairwise % identity**
Oryza sativa RAD50	TA52048_4530 (TIGR Plant transcript assembly)	1075 (80.5%)	80.3%
Arabidopsis thaliana RAD50	TA_45294_3702 (TIGR Plant transcript assembly)	861 (64.5%)	64%
Homo sapiens RAD50	NM_005732 (GenBank)	397 (14.8%)	40.1%
Saccharomyces cerevisiae RAD50	NM_001183088 (GenBank)	394 (14.7%)	40.1%

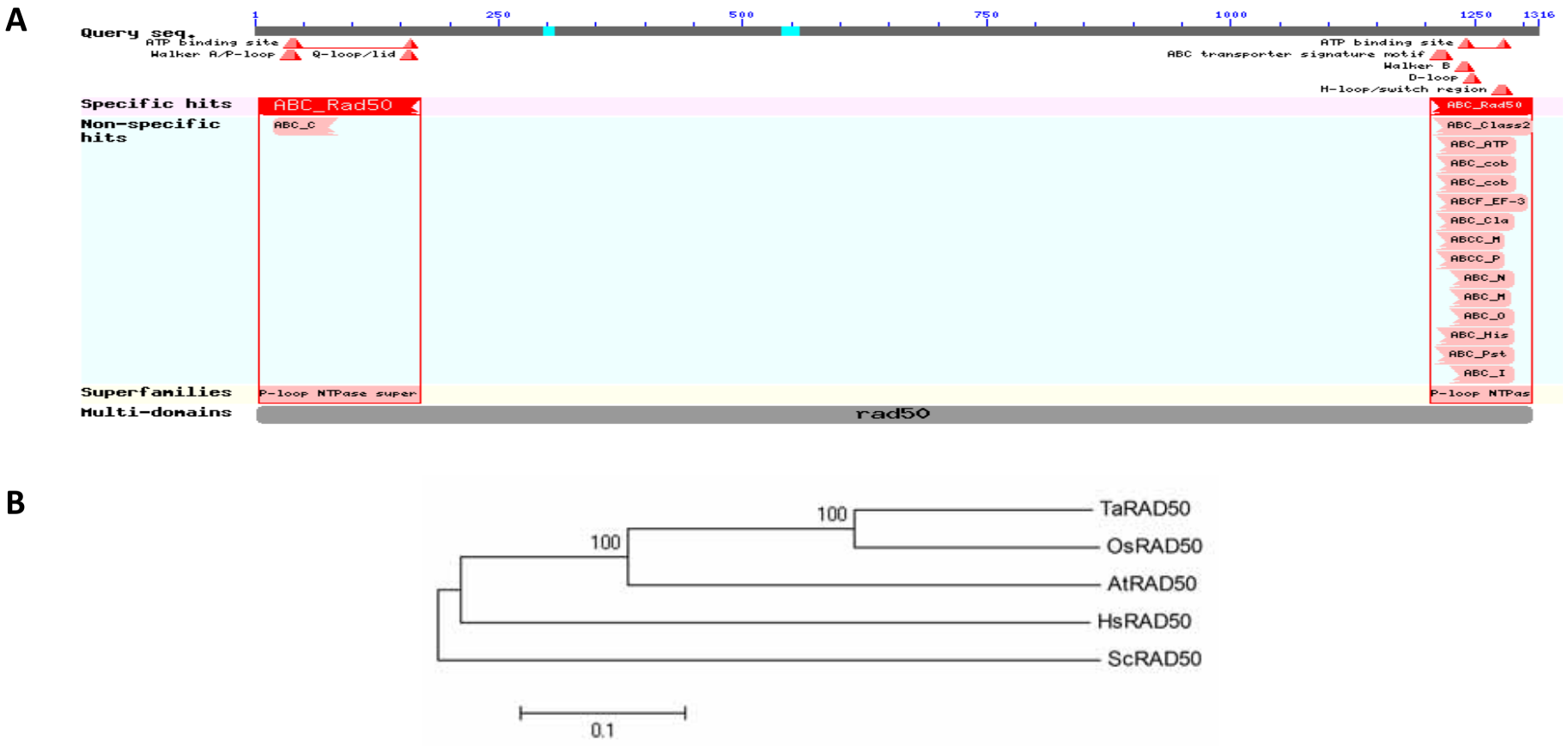


Figure 4.3 – Conserved domain and phylogenetic analyses of *Ta*RAD50. A) Conserved domain analysis revealed that *Ta*RAD50 possesses the ATP Binding Cassette (ABC)_Rad50 domain that is common in other RAD50 proteins previously characterised. B) The phylogenetic tree of the RAD50 protein from various organisms was constructed with the evolutionary history inferred using the Neighbor-Joining method. The bootstrap consensus tree inferred from 10000 replicates is taken to represent the evolutionary history of the five taxa analysed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (10000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 1106 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4.

4.3.3 – Construction of the pGBKT7-*TaRAD50* bait plasmid

The construction of the pGBKT7-*TaRAD50* plasmid was unsuccessful although multiple attempts were made. The two *TaRAD50* inserts (one with a 5′ *Eco*R1 cut-site overhang and a 3′ *Pst*1 cut-site overhang, and the other with a 5′ *Nde*1 cut-site overhang and a 3′ *Eco*R1 cut-site overhang) were liberated from their respective pGEM®T-Easy plasmids successfully using double-digests of their respective restriction enzymes (Figure 4.4). However, attempts to ligate either of the *TaRAD50* inserts into the pGBKT7 vector (digested with the same restriction enzymes as those used for liberation of the respective *TaRAD50* inserts) were unsuccessful (results not shown). As highlighted in the materials and methods, various methods of optimisation were attempted to facilitate ligation of the large *TaRAD50* insert into the pGBKT7 plasmid (see section 4.2.5.2).

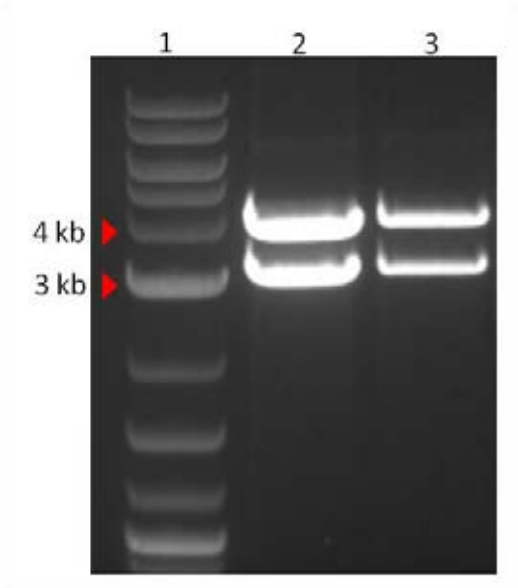


Figure 4.4 – Excision of both *TaRAD50* inserts from their respective pGEM®T-Easy clones. Two *TaRAD50* inserts with different flanking restriction enzyme cut-sites were prepared for directional cloning into appropriately prepared pGADT7. The linearised pGEM®T-Easy vector is 3015 bp in length with the *TaRAD50* insert excised. Lane 1, molecular weight marker; lane 2, *Eco*R1 and *Pst*1 double-digested pGEM®T-Easy-*TaRAD50*; lane 3, *Nde*1 and *EcoR*1 double-digested pGEM®T-Easy-*TaRAD50*.

4.4 – Discussion

Sequence analysis of the *TaRAD50* sequence as well as its protein product revealed that it shared high levels of similarity with rice compared to Arabidopsis. This is not unexpected as rice, like wheat, is a monocotyledonous plant whereas Arabidopsis is a dicotyledonous plant. Indeed, this high level of sequence similarity especially at both the 5′ and 3′ ends of the coding sequence enabled the isolation of the *TaRAD50* using primers based on *OsRAD50*.

Unlike the construction of the pGADT7-*TaPHS1* prey plasmid that was relatively straightforward, construction of the pGBKT7-*TaRAD50* plasmid proved much harder. Although significant amounts of time and effort were invested in optimising the ligation reaction, construction of the pGBKT-*TaRAD50* was unsuccessful. In total, three full-length *TaRAD50* inserts (with different flanking restriction enzyme overhangs) were generated for ligation reactions to construct the bait vector. One possible reason for the ligation failure is the large size of the *TaRAD50* inserts (all approximately 3610-3620 bp in length). Optimisation steps that included increased insert:vector ratios and higher concentrations of T4 DNA ligase failed to overcome this problem. Due to time constraints, no further approaches were investigated.

The logic behind persisting with the full-length *TaRAD50* coding sequence was to ensure that all the conserved domains of the RAD50 protein would be present in the resulting protein product used in the Y2H assay. In addition, using the entire amino acid sequence should allow the structural elements of RAD50 to hold its domains in its native 3-dimensional conformation. While this approach has been unsuccessful, other approaches such as co-immunoprecipitation of PHS1 protein from plant meiocyte cell lysates using the anti-*Ta*PHS1 antibody reported

in Khoo *et al.* (2011) could still answer whether interactions occur between RAD50 and PHS1.

While the work presented in this addendum has not succeeded in answering whether or not RAD50 and PHS1 directly interact with one another, the foundation for future Y2H experiments has been laid with the successful construction of the pGADT7-*TaPHS1* prey plasmid. This plasmid could be used in Y2H experiments in conjunction with bait vectors constructed with portions of the RAD50 sequence. Whether or not this approach would yield conclusive results remains to be seen as the results may be confounded by the use of partial RAD50 peptides. The RAD50 protein may possibly require more than just a single domain to interact with PHS1. Another approach would be to utilise the pGADT7-*Ta*PHS1 prey vector in a Y2H experiment to screen a bait vector library constructed from wheat meiotic cDNA in an effort to uncover other potential partners of *Ta*PHS1.

Chapter 5 – The isolation and characterisation of bread wheat *Molecular ZIPper 1* (*TaZYP1*)

5.1 – Introduction

While the synaptonemal complex has been studied extensively since the 1950s, it has only been in the past decade that the first plant synaptonemal complex (SC) genes were identified and characterised in Arabidopsis (*AtZYP1a* and *AtZYP1b*). These were named after the *Saccharomyces cereviseae* homologue, *Molecular ZIPper 1* (*ScZIP1*). *AtZYP1a* and *AtZYP1b* arose from a gene duplication event and encode proteins that share structural and functional similarities to *Sc*ZIP1 (Dong and Roeder 2000, Higgins *et al.* 2005). Both *At*ZYP1a and *At*ZYP1b form the transverse filaments of the proteinaceous tripartite SC structure that holds together sister chromosomes during the early stages of meiosis and are functionally redundant (Higgins *et al.* 2005). More recently, two other plant homologues of ZYP1 have also been studied in *Secale cereale* (*Sc*ZYP1) and *Oryza sativa* (*Os*ZEP1) (Mikhailova *et al.* 2006, Wang *et al.* 2010). While plant SC proteins share very little conservation at the primary sequence level with their equivalents in other organisms such as the mouse SYCP1 and *C. elegans* SYP2, they retain a high degree of tertiary structural similarity in addition to sharing a common function.

The immuno-localisation profiles of the ZYP1 homologues differ slightly from species to species, with ZYP1 signal first appearing as foci in leptotene stage meiocytes of Arabidopsis and rice but as short linear tracts in those of rye (Higgins *et al.* 2005, Mikhailova *et al.* 2006, Wang *et al.* 2010). These ZYP1 foci are only observed after ASY1 has been loaded onto the axial elements of the

chromatin (Higgins *et al.* 2005, Mikhailova *et al.* 2006). Using the Arabidopsis *spo11* and *dmc1* mutants, Higgins and colleagues (2005) deduced that ZYP1 is recruited to DSB sites during the early stages of single-stranded DNA invasion and that the loading of ZYP1 is independent of recombination initiation. However, lengthening of the ZYP1 signal along the chromatin is dependent on successful recombination. In order for recombination to occur, at least in Arabidopsis, ASY1 mediates the loading of DMC1 recombinase onto the chromatin (Sánchez-Morán *et al.* 2007). In rice and Arabidopsis, the ZYP1 signal lengthens from zygotene to pachytene and localises to the central region of the SC where it is sandwiched between ASY1 signals that are associated with the axial elements on both sides (Higgins *et al.* 2005, Wang *et al.* 2010).

In bread wheat, ASY1 (*TaASY1*) has previously been characterised as an axial element-associated protein that is required for synapsis and promoting homologous chromosome interactions (Boden *et al.* 2009, Boden *et al.* 2007). While the main objective of the work presented in this chapter was to isolate and characterise the bread wheat *ZYP1* (*TaZYP1*) homologue; the signal of *TaASY1* was also monitored. Having isolated the *TaZYP1* coding sequence, analysis of its predicted amino acid sequence and the stage-specific *TaZYP1* gene expression profile in both wild-type and the *ph1b* mutant plants was conducted. *In vitro* competitive DNA-binding assays to characterise the function of *Ta*ZYP1 were also performed while the temporal and spatial localisation profile of *Ta*ZYP1 within the meiocytes of two wild-type cultivars and the *ph1b* mutant were investigated using 3-dimensional dual immuno-fluorescence localisation techniques. Due to the importance of ASY1 in synapsis, maintenance of homologous chromosome interactions, and its reported role in mediating DMC1-

dependent homologous recombination, *Ta*ZYP1 was also investigated across five *Taasy1* mutants.

5.2 – Materials & Methods

5.2.1 – Primer design

5.2.1.1 – Primer design

A rice (*Oryza sativa*) EST (TIGR Rice Annotation number LOC_Os04g37960.1|12004.m08785) showing sequence similarities to both *AtZYP1a* and *AtZYP1b* (GenBank accession numbers NM_102076 and NM_102078 respectively) was identified through a TIGR Rice Genome Annotation Web BLAST search. Multiple forward and reverse primers were designed based on this rice EST to amplify *TaZYP1* (Table 5.1). Upon successfully isolating approximately 2kb of *TaZYP1* sequence, multiple nested primer sets were designed to confirm the sequence within the 2kb product. A 3′RACE primer was then designed to isolate the remainder of the *TaZYP1* sequence.

Table 5.1 – List of primers used to isolate and characterise *TaZYP1*. Primers required for standard PCR, 3′ RACE PCR, nested primers for determining nucleic acid sequences in the large *TaZYP1* ORF product, ORF-isolation primers, Q-PCR primers, and primers for amplifying the *TaZYP1* probe used in Southern blot experiments are all shown.

Primer name	**Primer sequence (5´ → 3´)**	**T_m (°C)**
***TaZYP1* gene isolation primers**		
*TaZYP1*_F1	ATGCAGAAGCTGGGTTTATCGGG	60
*TaZYP1*_R1	AATGCTCCTTGCTTCTCCTCCTTTGACT	
***TaZYP1* 3′ RACE primers**		
*TaZYP1*_3′R F1	TCAGAAGTCAAAGGAGGAGAAGCAAAGAGCATT	62
GeneRacer™ 3′ Primer	GCTGTCAACGATACGCTACGTAACG	76
GeneRacer™ 3′ Nested Primer	CGCTACGTAACGGCATGACAGTG	72

***TaZYP1* gene expression primers (to amplify ORF for protein production)**		
*TaZYP1*_ORF_F1	ATGCAGAAGCTGGGTTTATCGGG	58
*TaZYP1*_ORF_R1	CTAGGCAAATGCATAAGGGTCATCAGC	
***TaZYP1* nested primers (to obtain full-length *TaZYP1* sequence)**		
*TaZYP1*_cF1	ACAGTTGGAGGGTTCAGTTGAAGA	55
*TaZYP1*_cR1	GCCTCGGTAAGTTGACATTCTG	
*TaZYP1*_cF2	ATCCCGCTTATTGTGTGCTGACT	55
*TaZYP1*_cR2	CCTGAAGCATGAGATCGTACTGTT	
Plasmid vector sequencing primers		
GW1 (pCR®8/GW/TOPO®)	GTTGCAACAAATTGATGAGCAATGC	50
GW2 (pCR®8/GW/TOPO®)	GTTGCAACAAATTGATGAGCAATTA	
T7 (pDEST17)	TAATACGACTCACTATAGGG	
Quantitative real-time PCR (Q-PCR) primers		
*TaZYP1*_QF1	GCTTCAGTTGCCAGGTCCAG	57
*TaZYP1*_QR1	CAATGACTTCTGAGTATTCGGTTCC	
Southern blot probe primers		
*TaZYP1*_SF1	ACAAAAGTTACAGATCCAAGCATCA	58
*TaZYP1*_SR1	AATGCTCTTTGCTTCTCCTCCTTTGACT	

5.2.2 – Preparation of meiotic cDNA and 3′RACE libraries

5.2.2.1 – Synthesis of libraries

Staging and collection of Chinese Spring, Bob White MPB26, *ph1b*, *Taasy1-1.9*, *Taasy1-1.9.2*, *Taasy1-2.2*, *Taasy1-2.2.2*, and *Taasy1-2.2.3* meiotic spike tissue from plants grown in a controlled-environment room programmed with a 16/8 hr photoperiod at approximately 23°C was performed as per section 2.2.1. RNA extraction, quantitation and cDNA synthesis were performed as outlined in sections 3.2.3.3 and 3.2.3.4. Chinese Spring and *ph1b* stage-specific cDNA libraries were also made using RNA isolated from staged anther tissue. A 3′ RACE library was synthesised from 2.5 μg of Chinese Spring total RNA per reaction using a GeneRacer Kit (Invitrogen) according to the manufacturer's instructions.

5.2.3 – Preparation of *ZYP1* clones

5.2.3.1 – High fidelity amplification and isolation of *ZYP1* from meiotic cDNA

Initial PCR reactions contained 0.2 μL of template meiotic cDNA, 3.2 μL of 5 mM dNTPs, 1 μL of 10 μM forward and reverse primers (refer to Table 3.1), 2.5 μL of 10× high fidelity PCR buffer +$MgCl_2$ (Roche), 5 μL of GC-rich solution (Roche), 0.2 μL of Fast Start Taq polymerase (Roche), and 11.9 μL of autoclaved nanopure water. PCR cycle conditions were as follows; initial denaturation step at 95°C for 5 min, then 35 cycles of 96°C for 30 sec, T_m°C for 30 sec, 72°C for 2 min 30 sec, followed by a final extension step at 72°C for 10 min (refer to Table 5.1 for T_m of primers). 3′RACE PCR was performed as per section 3.2.3.5. Agarose gel electrophoresis and gel purification of the separated PCR products was conducted as per section 3.2.3.6.

5.2.3.2 – Production and sequencing of pCR®8/GW/TOPO®-*ZYP1* clones

Production and sequencing of pCR®8/GW/TOPO®-ZYP1 clones was performed as outlined in section 3.2.3.7 to 3.2.3.11. Sequence information obtained from the *ZYP1* clones derived from Chinese Spring were used to form a consensus sequence for *TaZYP1*. The *TaZYP1* sequence was then compared to the *OsZEP1* and *ZmZYP1* sequence to ensure that the correct coding sequence was isolated. One positive pCR®8/GW/TOPO®-*TaZYP1* ORF clone was selected for the construction of the *Ta*ZYP1 heterologous protein expression plasmid (refer to section 5.2.7).

5.2.4 – Chromosomal location of *TaZYP1*

5.2.4.1 – Southern blot analysis using nullisomic-tetrasomic membranes

The procedure was conducted as per section 3.2.6.1 using a 561 bp fragment (nucleotide positions 1554-2114) of the *TaZYP1* sequence as the probe.

5.2.5 – Expression analysis of *TaZYP1*

5.2.5.1 – Quantitative real-time PCR (Q-PCR)

Q-PCR was conducted in triplicate according to Crismani *et al.* (2006). cDNA sets screened were Chinese Spring and Bob White MPB26 wild-type, *ph1b*, *Taasy1-1.9*, *Taasy1-1.9.2*, *Taasy1-2.2*, *Taasy1-2.2.2* and *Taasy1-2.2.3*. For stage-specific Q-PCR analysis of *TaZYP1* in the Chinese Spring wild-type and *ph1b* backgrounds, cDNA from anther tissue of four pooled stages of meiosis were used: PM – pre-meiotic interphase; L-P – leptotene to pachytene; D-A – diplotene to anaphase I; and IP – immature pollen. For Q-PCR analysis of *TaZYP1* in the Bob White MPB26 wild-type and *Taasy1* mutants, cDNA from whole meiotic spike tissue was used. Product amplification was completed using gene specific Q-PCR primers (see Table 5.1). The optimal acquisition temperature for *TaZYP1* was 79°C. Q-PCR of *TaASY1* was conducted as per Boden *et al.* (2009). Correlation between *TaASY1-TaZYP1* gene expression was calculated using normalised transcript values of three technical replicates from Bob White MPB26 and all five *Taasy1* mutants.

5.2.6 – Analyses of *TaZYP1* and its protein product

5.2.6.1 – Comparative amino acid and conserved domain analyses of *Ta*ZYP1

The *TaZYP1* ORF sequence was translated into its corresponding amino acid sequence and the predicted physical properties of *Ta*ZYP1 were obtained using VectorNTI's protein analysis function. The *Ta*ZYP1 sequence was used in comparative amino acid and conserved domain analyses against other sequences. Amino acid alignments and comparisons of full-length ZYP1 sequences (obtained from various BLAST searches using the NCBI, TIGR, and PredictProtein [http://www.predictprotein.org/; (Rost *et al.* 2004)] databases), and subsequent construction of the phylogenetic tree (neighbour-joining method) (Saitou and Nei 1987) was completed using Molecular Evolutionary Genetics Analysis (MEGA) software (version 4.0) (Tamura *et al.* 2007). Default parameters were used except for the following: the pairwise deletion option was used, the internal branch test bootstrap value was set at 10,000 resamplings, and the model setting was amino acid: Poisson correction with predicted gamma parameters set at 2.0. Accession numbers of the sequences used were: *Ta* – *Ta*ZYP1; *Bd* – Bradi5g12010 (Phytozome); *Os* – LOC_Os04g0452500 (MSU Rice Genome Annotation); *Zm* – HQ116413 (GenBank); *Rc* – XP_002513917.1 (GenBank); *Vv* – CBI19158.1 (UniProt/TrEMBL); *At-a* – NP_173645.3 (GenBank); *At-b* – NP_564164.1 (GenBank); *Bo* – ABO69625.1 (GenBank); *Mm* – NP_035646.2 (GenBank); and *Sc* – NP_013498 (GenBank).

5.2.7 – *Ta*ZYP1 protein production, extraction and purification

5.2.7.1 – Construction of pDEST17-*TaZYP1* ORF

Construction of the pDEST17-*TaZYP1* ORF clone, bacterial transformation with the pDEST17-*TaZYP1* plasmid, colony PCR screening, directional sequencing, contig construction, and clone selection were performed as outlined in section 3.2.7.

5.2.7.2 – Protein production of *Ta*ZYP1

A pDEST17-*TaZYP1* ORF clone BL21-A1 *E. coli* (Invitrogen) cell line was established, which was followed by culturing and protein production using the protocols in section 3.2.8.

5.2.7.3 – Protein extraction, purification and SDS-PAGE analyses

Protein extractions and purifications were performed as outlined in section 3.2.9 and 3.2.10 while SDS-PAGE was performed as per 3.2.9.3. Visualisation of proteins was conducted by staining with only coomassie blue as performed in section 2.2.4.

5.2.8 – Mass-peptide identification of *Ta*ZYP1

5.2.8.1 – Sample preparation, MS, and peptide identification

Upon confirmation of positive SDS-PAGE results, the *Ta*ZYP1 band was excised from the gels with a scalpel. Sample preparation and mass spectrometry were performed as per sections 3.2.1 and 3.2.2. The MS spectra were searched with

Bioworks 3.3 (Thermo Electron Corp, San Jose, CA) using the Sequest algorithm against the Rice Genome Annotation (Build 6.1).

5.2.9 – Functional analysis of *Ta*ZYP1

5.2.9.1 – Competitive DNA-binding assay

Recombinant *Ta*ZYP1 extracted under native conditions were quantified using the Bradford assay (1976). The competitive DNA binding assay was conducted as described by Pezza *et al.*(2006) with modifications as per Khoo *et al.*(2008). The DNA binding ability of *Ta*ZYP1was tested with ΦX174 circular single-stranded DNA (ssDNA) (virion) (30 μM per nucleotide) (NEB) and ΦX174 linear double-stranded DNA (dsDNA) (RFI form *PstI*-digested) (15 μM per base pair) (NEB).

5.2.10 – Production of the anti-*Ta*ZYP1 polyclonal antibody

5.2.10.1 – Polyclonal antibody production against whole *Ta*ZYP1 protein

5.2.10.1.1 – Protein sample clean-up

The elution samples obtained in section 5.2.7.3 were put through a 4 mL column of Ni-NTA agarose beads (QIAGEN) equilibrated with 3 mL of Lysis buffer B. The elution-bead slurry was left to incubate on a daisy-wheel for 1 hr at 4°C. The beads were then spun down at 300 rpm and the supernatant was collected as Wash1 (W1). For the washing procedure, 15 mL of Wash buffer C2 was added to the bead slurry, which was then vortexed briefly at low speed, and left to incubate on ice for 3 min. The beads were subsequently spun down at 500 rcf for 2 min and the supernatant aspirated and kept as Wash 2 (W2). The wash procedure was repeated using 15 mL of Wash buffer D, with the supernatant aspirated and kept

as Wash 3 (W3). Elution of the purified 6×His-tagged proteins was then conducted six times, each time using 1 mL of Elution buffer E using the same procedure as the washing steps with each elution sample kept separate (E1 – E6). All solutions used in this protocol were identical to those in section 3.2.9.2.

5.2.10.1.2 – Determining protein sample quality and quantity

The purity and quality of the purified *Ta*ZYP1 sample was determined by SDS-PAGE and visualised by coomassie staining as per section 2.2.4. Quantification of the purified *Ta*PHS1 protein content in each sample was determined using a Bradford assay as per section 2.2.3. The values of the dilution series were used to generate a BSA standard curve, with the absorbance readings of the eluted samples then used to estimate protein concentration.

5.2.10.1.3 – Separation of purified protein samples and polyclonal antibody production

The purified *Ta*ZYP1 sample was separated using 11cm IPG+1 well SDS-PAGE gels (BIO-RAD). The *Ta*ZYP1 band was excised from the gel and washed with autoclaved nanopure water. It was then cut into equal length pieces, packaged and sent to IMVS for immunisation of the animals. Five mice (*Mus musculus*) were immunised with the *Ta*ZYP1 antigen. Immunisation and housing of the animals was conducted by IMVS, Gilles Plains Veterinary Division, Adelaide, South Australia.

5.2.10.1.4 – Clean-up of polyclonal antibody raised against full-length *Ta*ZYP1 antigen

A 250 mL culture of *E. coli* BL21 cells was grown overnight in at 37°C shaking incubator and the cells harvested by centrifugation at 7500 rpm. The cells were then washed with PBS solution twice and resuspended in saline (0.9% NaCl) at a concentration of 1 g ml^{-1}. 4 mL of acetone was added for every 1 mL of cell suspension and the mixture was shaken vigorously then incubated on ice for 30 min. The precipitate was collected by centrifugation at 8940 rpm for 10 min. The pellet was resuspended in fresh acetone, mixed vigorously then incubated on ice for 10 min. The precipitate was then pelleted by centrifugation at 8940 rpm for 10 min. The pellet was transferred to a piece of filter paper and air-dried before being crushed to a fine powder. Bacterial extract powder was then added to a final concentration of 1% (w/v) to a PBS solution containing 20% glycerol. The sera obtained from terminal bleeds of the immunised animals were then diluted 1:10 with the bacterial extract-PBS solution, vortexed, incubated on ice for 30 min before being centrifuged at maximum speed for 10 min. The supernatant was collected and used as the cleaned sera.

5.2.10.2 – Polyclonal antibody production against a partial *Ta*ZYP1 peptide

5.2.10.2.1 – Partial peptide synthesis and polyclonal antibody production

The amino acid sequence of *Ta*ZYP1 was assessed for hydrophobic and antigenic regions using Kite-Doolittle and Hopp plots. Peptides deemed suitable were assessed by Mimetopes (Mimetopes, Clayton, Victoria, Australia). The peptide sequence chosen for peptide synthesis by Mimetopes was: CLRAYHKEELQRIRS. The partial peptide was conjugated to keyhole limpet

using a hemocyanin maleimidocaproyl-N-hydroxysuccinimide (MCS) linker. Synthesis and purification of the peptide was performed by Mimotopes. The samples were delivered to IMVS for immunisation of two rats (*Rattus rattus*).

5.2.11 – Immuno-localisation of *Ta*ZYP1 by fluorescence microscopy

5.2.11.1 – Fluorescence immuno-localisation of *Ta*ZYP1 and *Ta*ASY1

Fluorescence immunolocalisation of *Ta*ZYP1 with *Ta*ASY1 was performed as per Franklin *et al.* (1999) and Boden *et al.* (2009) with the following modifications: anthers were fixed with 2% paraformaldehyde and cells permeabilised for 3 h. Anthers were harvested from Chinese Spring and Bob White MPB26 wild-type, *ph1b* and five *Taasy1* knock-down mutants (*Taasy1-1.9*, *Taasy1-1.9.2*, *Taasy1-2.2*, *Taasy1-2.2.2*, and *Taasy-2.2.3*). For detecting the localisation pattern of *Ta*ZYP1 using the mouse anti-full-length *Ta*ZYP1 antibody (1:100 and 1:10), an AlexaFluor® 488 conjugated donkey anti-mouse antibody (1:50; Molecular Probes, Invitrogen) was used while detection of the rat anti-partial peptide *Ta*ZYP1 antibody (1:100 and 1:10) was performed with an AlexaFluor® 488 conjugated donkey anti-rat antibody (1:50; Molecular Probes, Invitrogen).

Optical sections (90–120 per nucleus) of meiocytes were collected using a Leica TCS SP5 Spectral Scanning Confocal Microscope (Leica Microsystems, http://www.leicamicrosystems.com/) equipped with an oil immersion HCX Plan Apochromat 63 × /1.4 lens, a 405 nm pulsed laser and an Argon laser using an excitation wavelength of 468 nm. All images were processed using Leica Application Suite Advanced Fluorescence software (LAS-AF; version 1.8.2, build 1465, Leica Microsystems) to generate maximum intensity projections of each nucleus.

5.3 – Results

5.3.1 – Isolation and characterisation of the *TaZYP1* coding sequence

The full-length 2592 bp coding sequence of *TaZYP1* was successfully isolated using a combination of standard and 3′ RACE PCR techniques (Figure 5.1). This transcript encodes a protein product 863 aa in length with a predicted MW of 98.541 kD and an overall mean pI of 6.4. Southern blot analysis using membranes prepared with digested DNA from nullisomic-tetrasomic wheat plants showed that *TaZYP1* is located on chromosome group 2 with a copy on the A, B, and D genome respectively (Figure 5.2).

The *Ta*ZYP1 protein shares high levels of conservation with its homologues in close relatives, including the rice transverse element protein (GenBank accession: ADD69817) (Identities = 80%, Positivies = 91%, *E*-value = 0.00) and maize ZYP1 (GenBank accession: HQ116413) (Identities = 75.9%, Positives = 88%, *E*-value = 0.0). When compared with the Arabidopsis homologues, *At*ZYP1b (GenBank accession: NP_564164.1) (Identities = 40%, Positives = 64%, *E*-value = 1e-156) and *At*ZYP1a (GenBank accession: NP_173645.3) (Identities = 39%; Positives = 63%, *E*-value = 1e-152); the level of sequence conservation was reduced. Comparisons with its homologues in non-plant species showed sequence conservation was modest at best, a result reflected in the phylogenetic analysis (Figure 5.3).

Amino acid sequence analysis revealed that both the N- and C-terminal regions of *Ta*ZYP1 (aa positions 1-68 and 723-863 respectively) have high pI values of 10.22 and 10.05. In addition, 18.75% of the C-terminus consisted of

arginine and lysine residues. A putative DNA-binding S/TPXX motif was also found within this region (aa positions 761 to 764). Conserved domain analysis revealed that the *Ta*ZYP1 amino acid sequence retained modest similarities with two known Structural Maintenance of Chromosomes (SMC) conserved domains characterised in archaea, namely SMC_prok_B.and SMC_prok_A (*E*-values = 3.65e-09 and 3.00e-04 respectively) (Figure 5.4). 3-dimensional protein modelling predicted that the central region of *Ta*ZYP1 forms a coiled-coil domain structure (from aa position 69 to 722) (data not shown).

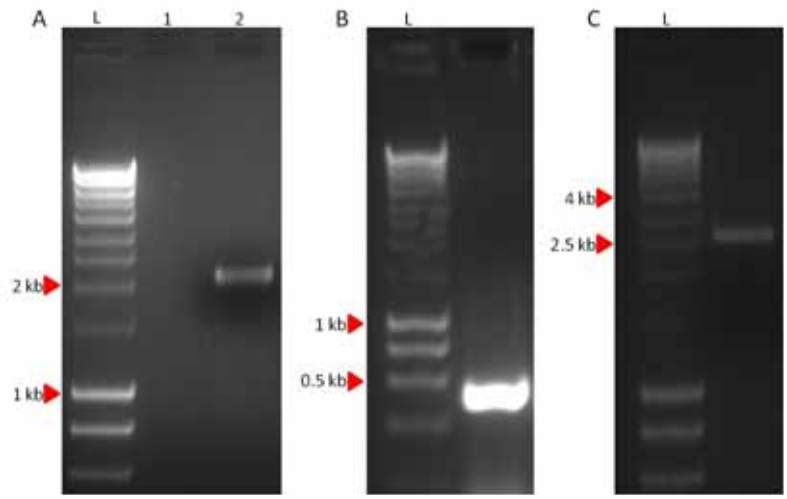


Figure 5.1 – Isolation of the full-length coding sequence of *TaZYP1*. A) Using primers designed based on a rice EST that shared sequence similarity to *AtZYP1a* and *AtZYP1b*, standard PCR was used to amplify a 2 kb fragment of *TaZYP1*. L – Molecular weight marker; Lane 1 – blank; Lane 2 –*TaZYP1* 2 kb fragment. B) 3′RACE PCR was used to amplify the remainder of the *TaPHS1* coding sequence. L – Molecular weight marker. C) Full-length *TaZYP1* ORF transcript. L – Molecular weight marker.

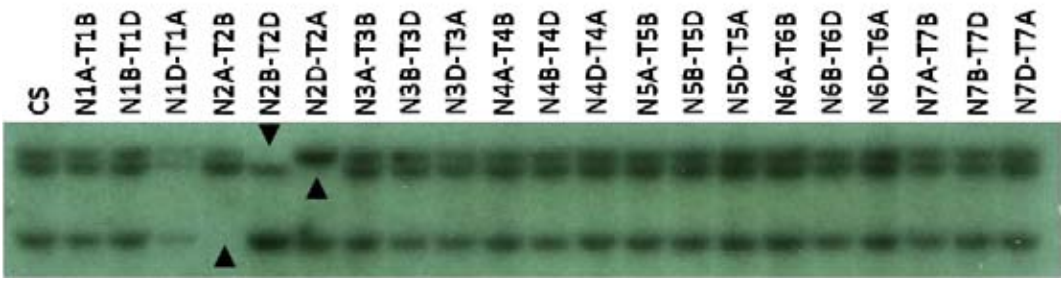


Figure 5.2 – *Ta*ZYP1 resides on chromosome group 2. Southern blot analysis showed the absence of bands in lanes containing DNA from plants nullisomic for chromosome group 2 thus indicating the location of *TaZYP1* within the genome. Arrow heads denote the missing bands. Loading was uneven in lane 4 (N1D-T1A) hence the faint signal.

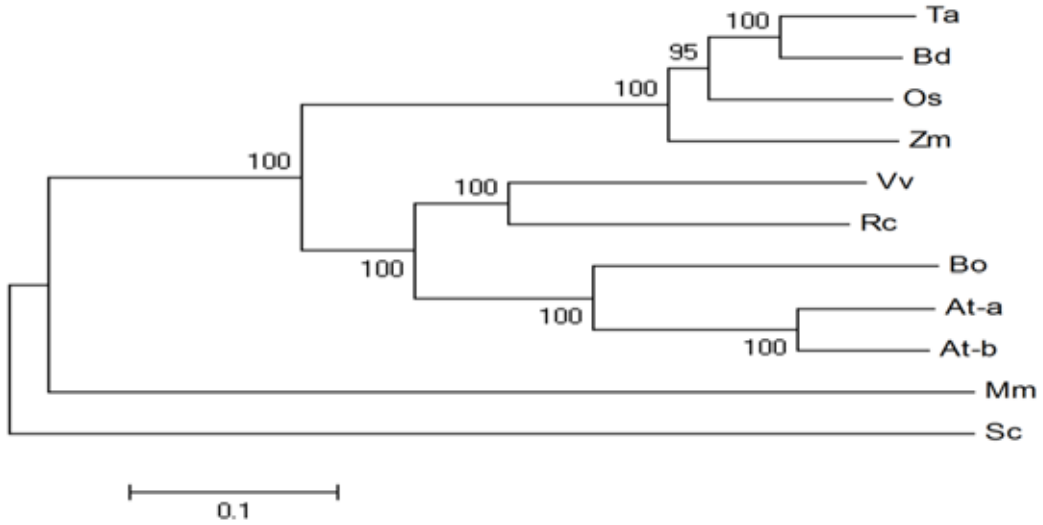


Figure 5.3 – Phylogenetic analysis of *Ta*ZYP1 and its homologues. The evolutionary history was inferred using the Neighbor-Joining method. The bootstrap consensus tree inferred from 10000 replicates is taken to represent the evolutionary history of the 11 taxa analysed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (10000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons. There were a total of 1019 positions in the final dataset. Phylogenetic analysis was conducted using MEGA4. Ta – *Triticum aestivum*; Bd – *Brachypodium distachyon*; Os – *Oryza sativa*; Zm – *Zea mays*; Vv – *Vitus vinifera*; Rc – *Ricinus communis*; Bo – *Brassica oleracea*; At-a – *Arabidopsis thaliana* ZIP1a; At-b – *Arabidopsis thaliana* ZIP1b; Mm – *Mus musculus*; Sc – *Saccharomyces cereviseae*.

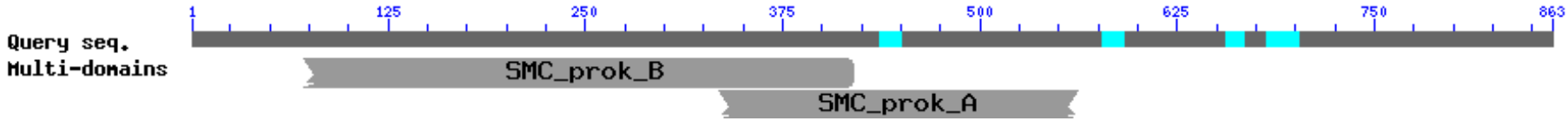


Figure 5.4 – Conserved domain analysis of *Ta*ZYP1. *Ta*ZYP1 shared modest sequence conservation with two known Structural Maintenance of Chromosomes (SMC) domains previously characterised in DNA-binding proteins found in multiple species of Archaea.

5.3.2 – Characterisation of the *TaZYP1* protein product *in vitro*

5.3.2.1 – Heterologous expression of *Ta*ZYP1

The 6×His-tagged *Ta*ZYP1 protein was heterologously expressed in *E. coli* and then purified by nickel affinity chromatography. This was successfully performed under both denaturing and native conditions. The protein product migrated close to its predicted molecular weight within denaturing SDS-PAGE gels (Figure 5.5).

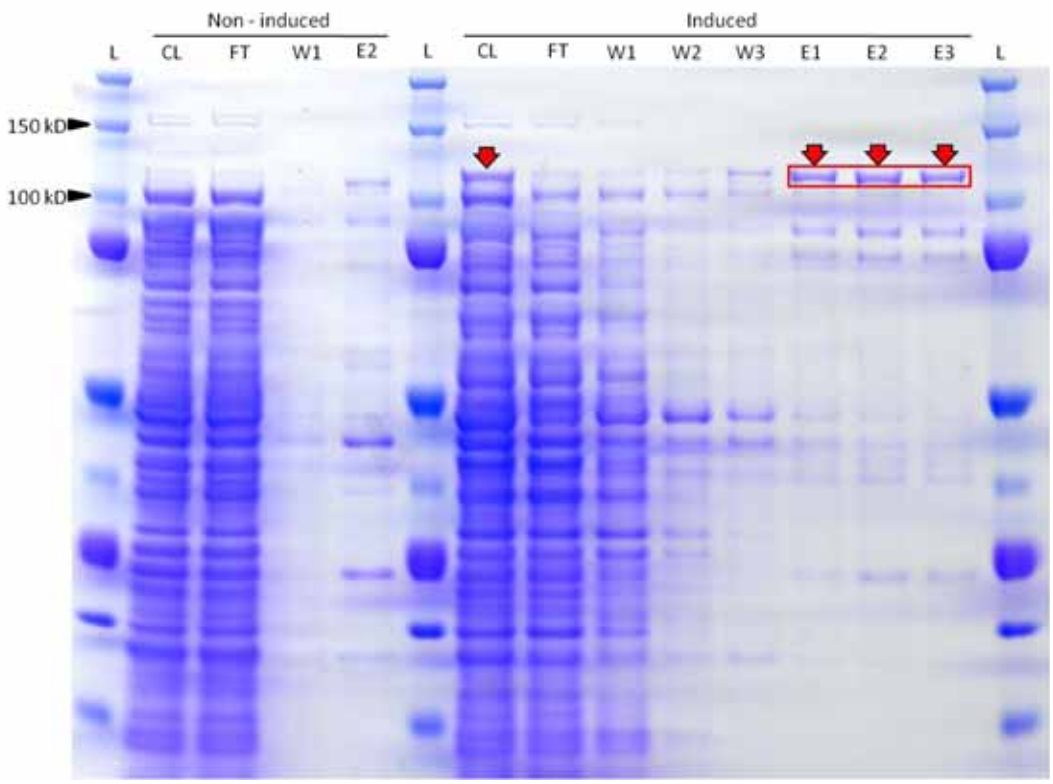


Figure 5.5 – SDS-PAGE analysis of *Ta*ZYP1. Heterologously expressed 6×His-tagged *Ta*ZYP1 was successfully isolated under denaturing conditions. L – BIO-RAD Precision Plus molecular weight marker, Cell lysate (CL), Flow-through (FT), Wash 1 (W1), Wash 2 (W2), Wash 3 (W3), Elution 1 (E1), Elution 2 (E2) and Elution 3 (E3) samples are shown for the induced cell culture while W2, W3, E1 and E3 were omitted from the non-induced cell culture as they were not required for the analysis of the result. The red arrows highlight the 6×His-tagged *Ta*ZYP1 product and the red rectangle highlight the bands identified by MS/MS as *Ta*ZYP1.

Mass peptide identification of peptides obtained from tryptic digestion of *Ta*ZYP1 revealed the peptides were identical to regions within a rice synaptonemal complex protein (TIGR Rice locus identifier: LOC_Os04g37960), with a total amino acid coverage percentage of 6.54% spreading from the N-terminus to the C-terminus (Figure 5.6).

Protein:

MQKLGLSGLR GLEGFRSLAG STSTAAKAPN PKPSSDIGGS TYGSFANLKI
TAEKLVKEQA SVKTDLEMTV YFSAPTIIYG NPTGWVYGNI IPTVVIPHGA
NVLVISWIKE AKLQQAVNEN AKLKVKQTED SKLWQGLDSK VSSTKTLCNQ
LTETLQQLAS QTERAEEDKK PFEEILGKNS KAFEEFNCLL HDSSIKLECA
EQMIISGKQE MLRIKQEKEE MDQSYKEQLY ASDTTIREKN SLIKQLEDSI
EQNKARLLYV DSRLECMEQE LKLKEDVCIC LKENLASTES EKNDLKLRNE
GYTLEVQKLS KDNKELNELL SGFTVKVTEL DKEHTSISSH VTQLISSFER
YDGKVHEEKM LMIKSAKDKF EHLQNQYVNL ISENNALQTE IEELKSRIIE
LQKTQEIVMV QHVEECQVAE DKIRRLESEA EISASNISQL EKVASDLEGR
VQKLLEDSRS AENHKQELLQ KILKLESDNQ KLLGRVQSVL NEKSNDTESL
QQEIAKRDQQ VETLENQVNQ LCSIIDEKEQ LHTCAVEREK NLEEQKLQVQ
ASLAATESQL TEAKKQYDIM LEGKKIELSK HLKELSLKND QAINEIRRKY
ELEKVEIINI EKEKAEKLIK EMENKCNEKI SENRQDSESS FEMVARIQQD
NEHKESTLRA YHKEELQRIQ SQAENELRER LSSLRKDHEI QMKSLTKKHE
ENCQKLQDEL ELQKSKEEKQ RALLQLQWKV MQETQQVDQE VNSKKEYSVS
SIKRRDPYIR KEHQLQLVSP ETKRKDVNLS GIIQSPITNM LRKVEKGTQD
IPKHRKVTHH EYEVETANGR ITKRRKTKST VMFGEPNTQK SLHDTADKDP
TKMKKVVAGS HPHPANIGEL FSEGSLNPYA EDPYAPG

Protein Coverage:

Sequence	MH+	% Mass	AA	% AA
LGLSGLR	715.45	0.70	4 - 10	0.79
GLEGFRSLAGSTSTAAK	1652.86	1.62	11 - 27	1.92
LWQGLDSK	946.50	0.93	133 - 140	0.90
EQLYASDTTIR	1296.64	1.27	227 - 237	1.24
RLESEAEISASNISQLEK	2004.02	1.97	425 - 442	2.03
LESDNQELLGR	1273.64	1.25	475 - 485	1.24
VQSVLNEK	916.51	0.90	486 - 493	0.90
KQYDIMLEGK	1224.63	1.20	565 - 574	1.13
QYDIMLEGK	1096.53	1.08	566 - 574	1.01
AYHKEELQR	1173.60	1.15	660 - 668	1.01
LQDELELQK	1115.59	1.09	706 - 714	1.01
ALLQLQWK	999.60	0.98	722 - 729	0.90
STVMFGEPNTQK	1338.64	1.31	829 - 840	1.35
Totals:	6803.51	6.68	58	6.54

Figure 5.6 – Mass-peptide identification summary of *Ta*ZYP1. The purified *Ta*ZYP1 samples ran on SDS-PAGE gels were washed, digested with trypsin and then identified by ion-tap mass-spectrometry. Peptides identified from the MS/MS data of the digested *Ta*ZYP1 samples that were identical to those within the *Os*ZYP1 sequence (TIGR rice locus identifier: LOC_Os04g37960) are highlighted (yellow) in the cartoon as well as shown in the rice synaptonemal complex protein amino acid sequence (red font). Total sequence coverage was 6.54% of the protein spread over its entire length.

5.3.2.2 – *Ta*ZYP1 interacts with DNA

*Ta*ZYP1 protein extracted under native conditions was used in competitive DNA-binding assays to investigate whether *Ta*ZYP1 interacted with DNA as had previously been reported for its homologues. Competitive DNA-binding assays showed that *Ta*ZYP1 appears to interact with both ssDNA and dsDNA without

any preferences (Figure 5.7). This is evidenced by the equal rates of retardation of both DNA species regardless of the *Ta*ZYP1 concentration used in the assays.

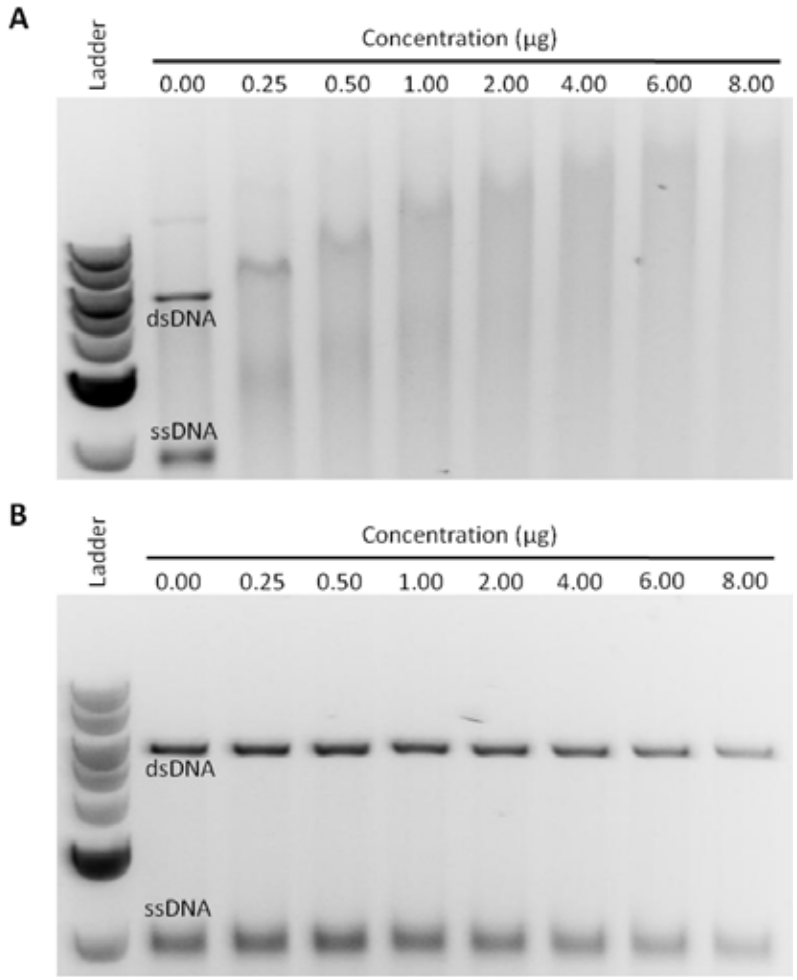


Figure 5.7 – *Ta*ZYP1 interacts with both ss- and dsDNA equally. *Ta*ZYP1interactions with both DNA species appear equal as the severity of retardation of both DNA species appears identical in each lane regardless of the concentration of *Ta*ZYP1 present. (A) Induced native *Ta*ZYP1 protein extract; (B) Non-induced native *Ta*ZYP1 protein extract.

5.3.3 – Transcript expression analysis of *TaZYP1*

5.3.3.1 – Q-PCR analysis of *TaZYP1* in wild-type and the *ph1b* mutant

Q-PCR was conducted to determine the transcript expression levels of *TaZYP1* across various stages of meiosis in both the Chinese Spring wild-type and the *ph1b* mutant (Figure 5.8). The data showed that under wild-type conditions,

TaZYP1 was highly expressed during pre-meiotic interphase. This high level of *TaZYP1* expression persisted as the cells progressed through the early stages of meiosis up to pachytene. Transcript numbers then steadily declined from diplotene to anaphase I. By the stage of immature pollen formation, *TaZYP1* transcript levels were only approximately 11% that of the pre-meiotic interphase stage transcript levels.

In the *ph1b* mutant, *TaZYP1* transcript levels are present at increased levels throughout all stages of meiosis examined but the overall expression pattern of *TaZYP1* remains similar to that of wild-type. The largest differences in *TaZYP1* transcript copy numbers were seen in the L-P and D-A pooled stages with an approximate 1.27-fold and 1.30-fold increase seen in the *ph1b* mutant.

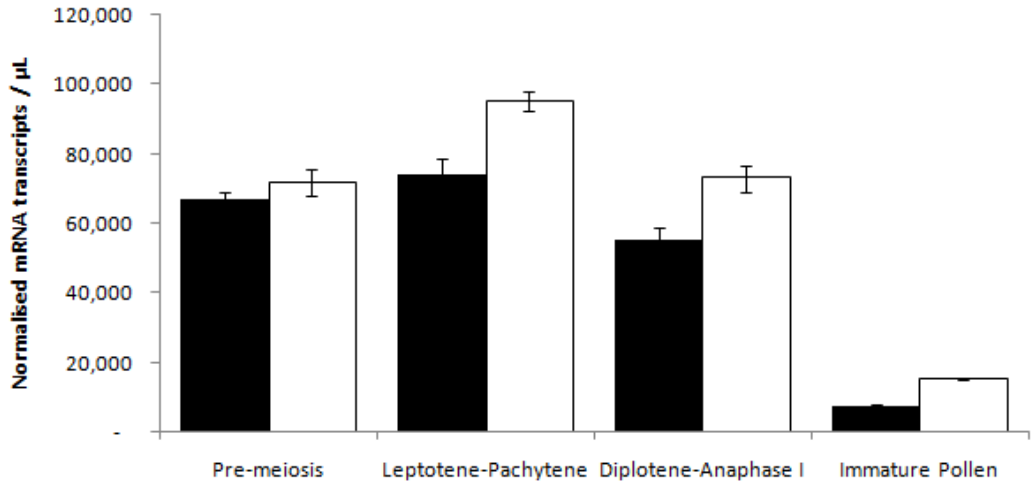


Figure 5.8 – Transcript expression analysis of *TaZYP1*. Q-PCR analysis of *TaZYP1* transcript numbers revealed that *TaZYP1* is highly expressed during the early stages of meiosis but reduces significantly at the later stages examined in Chinese Spring wild-type (black bar). While the general expression profile of *TaZYP1* in the *ph1b* mutant (open bar) is similar to that of the wild-type, *TaZYP1* transcript numbers are elevated with a 1.3-fold increase seen during leptotene to pachytene when compared to the wild-type. Error bars represent the ± standard deviation of three replicate experiments.

5.3.3.2 – Q-PCR analysis of *TaZYP1* in *Taasy1* mutant lines

TaZYP1 gene expression levels were also investigated in five individual RNAi knock-down *Taasy1* mutant lines; two from the T_1 generation and three from the T_2 generation (Figure 5.9 A). Q-PCR analysis revealed that there were no statistically-significant differences between the *TaZYP1* expression levels of the wild-type and *Taasy1* mutants except for *Taasy1*-2.2.3 where transcript levels were 4.17-fold higher than wild-type. In contrast, *TaASY1* levels were significantly lower in all five *Taasy1* mutants when compared to wild-type levels except for *Taasy1-2.2.3* which unexpectedly showed a 3.69-fold increase. Even though no statistically-significant differences were observed in the *TaZYP1* expression levels of four of the *Taasy1* mutants, the overall fluctuations in the *TaZYP1* levels within the wild-type and *Taasy1* mutants mirrored the fluctuations seen in *TaASY1* expression with a high correlation coefficient of r = 0.9604 obtained (Figure 5.9 B).

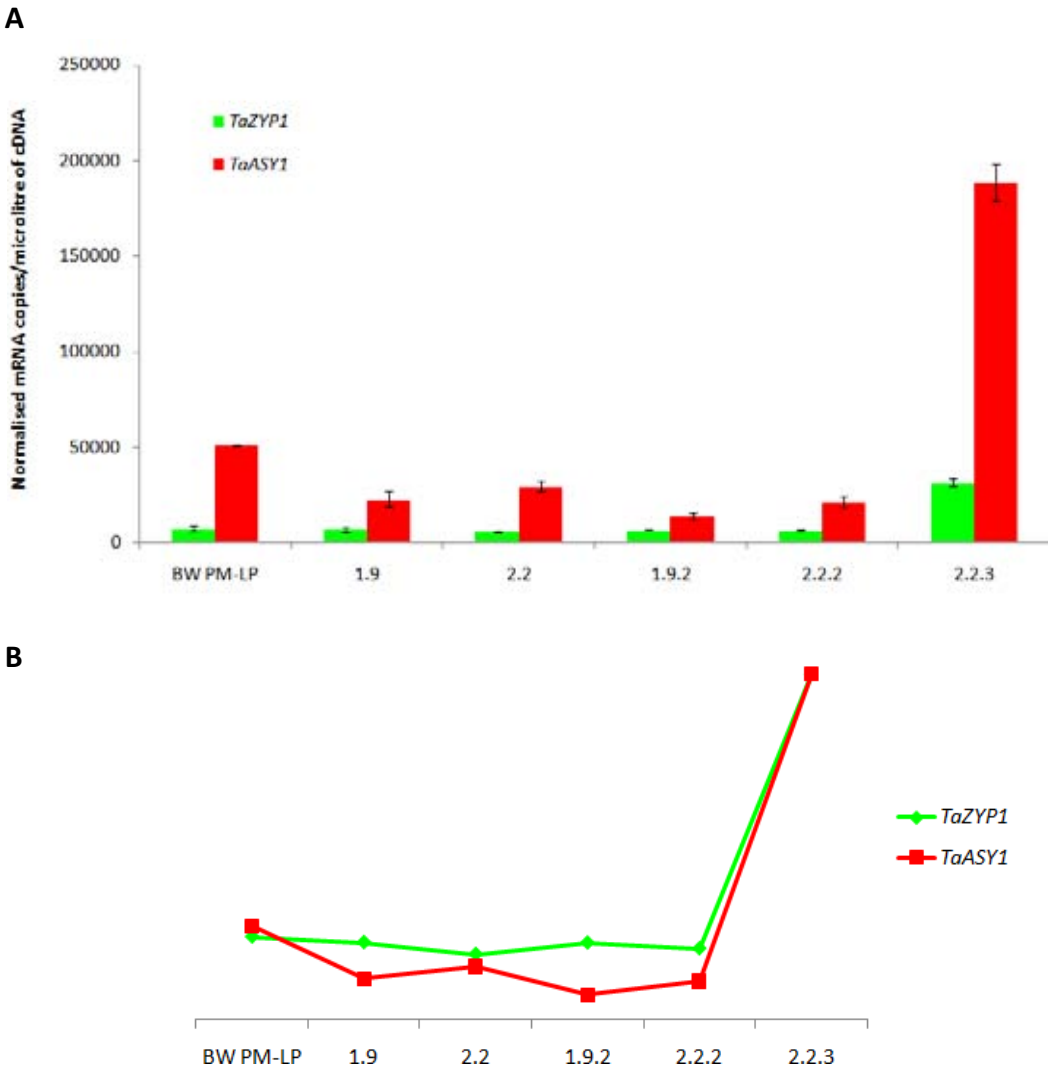


Figure 5.9 – Q-PCR analysis of T_1 and T_2 *Taasy1* mutant lines. A) Expression profiling of *TaZYP1* revealed no significant differences in expression levels between Bob White MPB26 (BW) wild-type and the *Taasy1* mutants apart from *Taasy1-2.2.3* which showed a 4.17-fold increase. *TaASY1* expression levels were significantly reduced in all the mutants except for *Taasy1-2.2.3*, which showed a 3.69-fold increase compared to wild-type. Error bars represent the standard deviation of three replicate experiments. B) A high correlation coefficient (r = 0.9604) was seen between the expression profiles of *TaZYP1* and *TaASY1* in the 6 lines examined.

5.3.4 – Fluorescence immuno- localisation of *Ta*ZYP1

5.3.4.1 – Localisation profile of *Ta*ZYP1 in wild-type bread wheat

3-dimensional dual immuno-fluorescence localisation experiments using staged wheat meiocytes embedded within poly-acrylamide pads were performed to investigate the localisation profile of *Ta*ZYP1 during early meiosis, with *Ta*ASY1 localisation used as a control. Initial experiments using the polyclonal antibody raised against the full-length *Ta*ZYP1 protein yielded poor results due to non-specific binding of the antibody to the cell cytoplasm. Optimisation of this antibody included performing the immunolocalisation experiment with increased antibody concentrations (1:10 vs 1:100) as well as pre-treating the immune-sera with bacterial acetone powder to remove antibodies that did not recognise the *Ta*ZYP1 protein. However, these optimisation experiments did not overcome the non-specific binding of the polyclonal antibody (Figure 5.10).

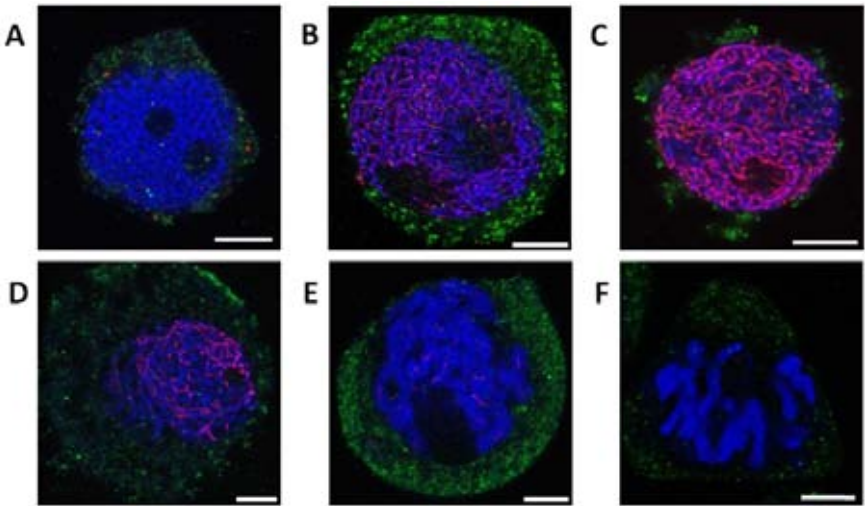


Figure 5.10 – Immuno-localisation results using the polyclonal antibody raised against the full-length *Ta*ZYP1 antigen. (A) pre-meiotic interphase; (B) leptotene – zygotene transition; (C) zytogene – pachytene transition; (D) pachytene; (E) diplotene; (F) diakinesis. The polyclonal antibody raised against the full-length *Ta*ZYP1 antigen displayed non-specific binding even after the immune sera was pre-adsorbed with bacterial acetone powder to remove antibodies that did not bind specifically to *Ta*ZYP1 (green). *Ta*ASY1 was used as a marker for early meiotic events. Chromatin was counter-stained with DAPI (blue). Scale bar, 7.5 µm.

Dual immuno-localisation assays conducted with the polyclonal antibody raised against the *Ta*ZYP1 partial peptide yielded superior results compared to those using the initial polyclonal antibody raised against the full length *Ta*ZYP1 protein. *Ta*ASY1 localisation was used as a control in these experiments as it has previously been shown to be a good marker of events that occur during early meiosis in wheat. In addition, its localisation profile has previously been characterised in both wild-type and the *ph1b* mutant. *Ta*ASY1 localisation in these experiments was as expected with signal first being detected during pre-meiotic interphase (Figure 5.11 row 1, panel 1). The signal lengthened along the chromatin during leptotene and then condensed as synapsis and pairing progressed during zygotene (Figure 5.11 row 1, panels 2 and 3 respectively). The *Ta*ASY1 signal also revealed the presence of synapsis forks that appeared during late-zygotene; while in pachytene the *Ta*ASY1 signal thickened further as synapsis was completed and the chromatin condensed further (Figure 5.11 row 1, panel 4). By diplotene, almost all *Ta*ASY1 protein had been unloaded from the chromatin and only residual amounts of signal was observed on the chromatin (Figure 5.11 row 1, panel 5). *Ta*ASY1 signal was completely absent from cells in the diakinesis stage (Figure 5.11 row 1, panel 6).

Unlike the *Ta*ASY1 signal that first appeared during pre-meiotic interphase, *Ta*ZYP1 was first observed as faint punctuate foci that localised to chromatin during early-leptotene within the wild-type cells examined (Figure 5.11 row 1, panel 2). These foci then formed short linear tracts along the chromatin as the cells progressed into zygotene (Figure 5.11 row 1, panel 3). From mid-zygotene to pachytene, the *Ta*ZYP1 signal continued to lengthen but also

condensed mirroring the condensation of the chromatin. During late pachytene, the *Ta*ZYP1 signal appeared as long tracts that populated regions of chromatin even after *Ta*ASY1 was unloaded (Figure 5.11 row 1, panel 4). The *Ta*ZYP1 signal then quickly dissipated from late-pachytene to diplotene where only low residual levels of *Ta*ZYP1 signal were observed (Figure 5.11 row 1, panel 5). No *Ta*ZYP1 signal was present during diakinesis (Figure 5.11 row 1, panel 6). Interestingly, the *Ta*ZYP1 signal appeared on regions of chromatin only after *Ta*ASY1 first appeared within those regions, indicating that synapsis was completed before *Ta*ZYP1 installation between the homologous chromosomes. In addition, although *Ta*ZYP1 and *Ta*ASY1 signals were sometimes seen localised on regions of chromatin simultaneously, they did not co-localise with one another.

5.3.4.2 – Characterising *Ta*ZYP1 localisation in the *ph1b* bread wheat mutant

Having determined that *TaZYP1* expression levels were slightly elevated in the *ph1b* mutant (Figure 5.8), dual immune-fluorescence localisation assays were performed to determine whether *Ta*ZYP1 loading was affected within the *ph1b* mutant. As previously reported by Boden *et al*. (2009), the *Ta*ASY1 signal was much stronger in the *ph1b* mutant with the nucleoplasm of some observed meiocytes completely saturated even at pre-meiotic interphase (Figure 5.11 row 3, panel 1). Furthermore, the *Ta*ASY1 signal appeared disordered on the condensing chromatin and persisted into diplotene. Large aggregates of *Ta*ASY1 protein, thought to be polycomplexes, were also observed as large foci in nucleolar regions of the *ph1b* meiocytes (Figure 5.11 row 3, panel 5)

Similar to wild-type, *Ta*ZYP1 was first observed as faint punctuate foci during leptotene in the *ph1b* mutant (Figure 5.11 rows 2 and 3, panel 2).

However, unlike wild-type, lengthening of the signal from foci to form short tracts was delayed until pachytene (Figure 5.11 rows 2 and 3, panel 4). In addition, *Ta*ZYP1 signal persisted at higher levels in diplotene compared to wild-type, possibly as a consequence of the delayed loading of *Ta*ZYP1 in the mutant (Figure 5.11 rows 2 and 3, panel 5).

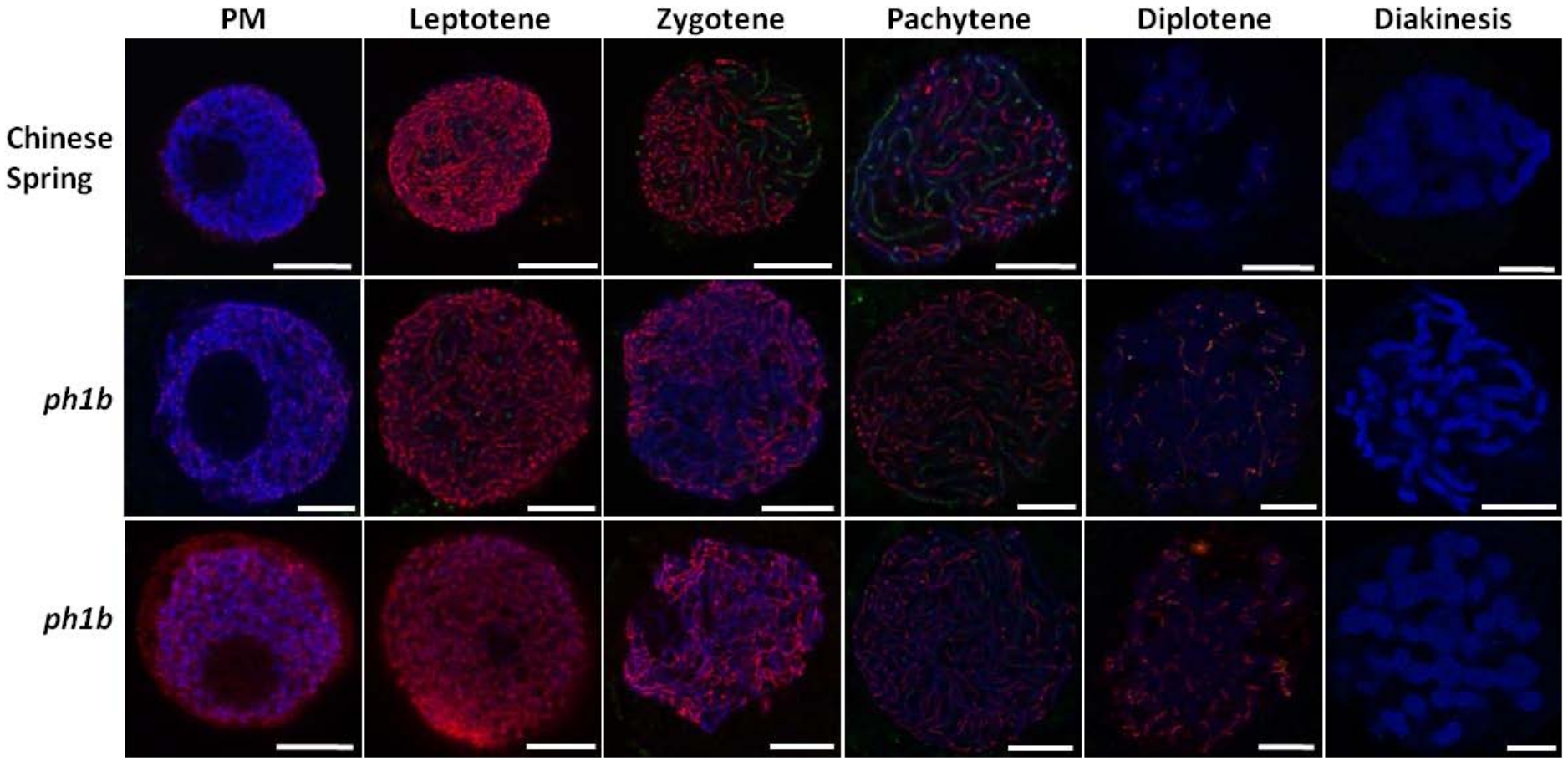


Figure 5.11 –*Ta*ZYP1 localisation during sub-stages of prophase I in both wild-type bread wheat as well as the *ph1b* mutant. In Chinese Spring wild-type meiocytes, *Ta*ZYP1 (green) first appears as faint punctuate foci that localise to chromatin during leptotene. *Ta*ZYP1 signal expands from these foci to form short tracts in zygotene before lengthening and condensing to form long thick tracts along the synapsing homologous chromosomes during pachytene. The *Ta*ZYP1 signal diminishes significantly by diplotene and is absent in diakinesis. Spatial localisation of *Ta*ZYP1 in the *ph1b* mutant was similar to wild-type but differed temporally. The *Ta*ZYP1 signal remained as punctuate foci until the zygotene-pachytene transition before short tracts were observed. The signal of *Ta*ZYP1 persisted into diplotene at higher levels when compared to wild-type. *Ta*ASY1 (red) was used as a marker of early meiotic events. Chromatin (blue) was counter-stained with DAPI. Scale bars, 7.5 µm.

5.3.4.3 – Characterising *Ta*ZYP1 localisation in *Taasy1* bread wheat mutants

Dual immuno-fluorescence localisation assays were also performed on meiocytes of five *Taasy1* mutant lines with a Bob White MPB26 control included as the *Taasy1* mutants were derived from a Bob White MPB26 genotype background (Figure 5.12). In general, *Ta*ASY1 signals in the mutant lines appeared similar to wild-type both spatially and temporally but appeared weaker and more diffuse. In addition, *Ta*ASY1 persisted at much higher levels in diplotene compared to the wild-type (compare Figure 5.12 row 1, panel 5 with *Taasy1* mutant panels down the column).

*Ta*ZYP1 localisation was identical both temporally and spatially in the Bob White MPB26 and Chinese Spring wild-types (compare Figure 5.11 row 1 with Figure 5.12 row 1). Amongst the *Taasy1* mutants, *TaZYP1* signals were generally similar to wild-type, first appearing as faint punctate foci on regions of chromatin during leptotene and lengthening into short tracts during zygotene (Figure 5.12 rows 2 to 6, panels 1 to 3). The *Ta*ZYP1 signal then condensed and lengthened further during pachytene, populating regions of chromatin where the *Ta*ASY1 signal was being unloaded (Figure 5.12 rows 2 to 6, panel 4). *Ta*ZYP1 signal in the *Taasy1* mutants were only present at residual levels during diplotene (as was seen in the wild-type) (Figure 5.12 row 1, panel 5 with panels down that column). Similar to observations of both wild-type meiocytes (Figures 5.11, 5.12), *Ta*ZYP1 and *Ta*ASY1 signals were sometimes seen populating regions of chromatin simultaneously but were never co-localised.

Amongst the *Taasy1* mutants examined, two anomalies were evident (which were consistently observed across at least 80 to 100 meiocytes from two independent immune-fluorescence experiments, with each of the lines

investigated). First, *Ta*ZYP1 loading appeared more reduced in the *Taasy1-1.9.2* mutant compared to the other mutants examined. While the *Ta*ZYP1 signal was at its most abundant during pachytene in both wild-types (Figures 5.11, 5.12), as well as all the other *Taasy1* mutants examined, the pachytene stage *Taasy1-1.9.2* meiocytes displayed a very weak and diffuse *Ta*ZYP1 signal (Figure 5.12 row 3, panel 4). The signal appeared as short diffused and disrupted tracts instead of the long condensed tracts that were typically observed in wild-type meiocytes. In some cases, *Ta*ZYP1 signal was almost undetectable (Appendix D1 Figure 2).

Secondly, contrary to the Q-PCR data which showed that *TaZYP1* and *TaASY1* expression were 4.17-fold and 3.69-fold higher than wild-type, the immuno-fluorescence localisation data for both proteins in *Taasy1-2.2.3* meiocytes did not show increased signal of either protein within the cells (when compared to the level of loading observed across the other mutants and wild-type) (Appendix D1 Figure 3).

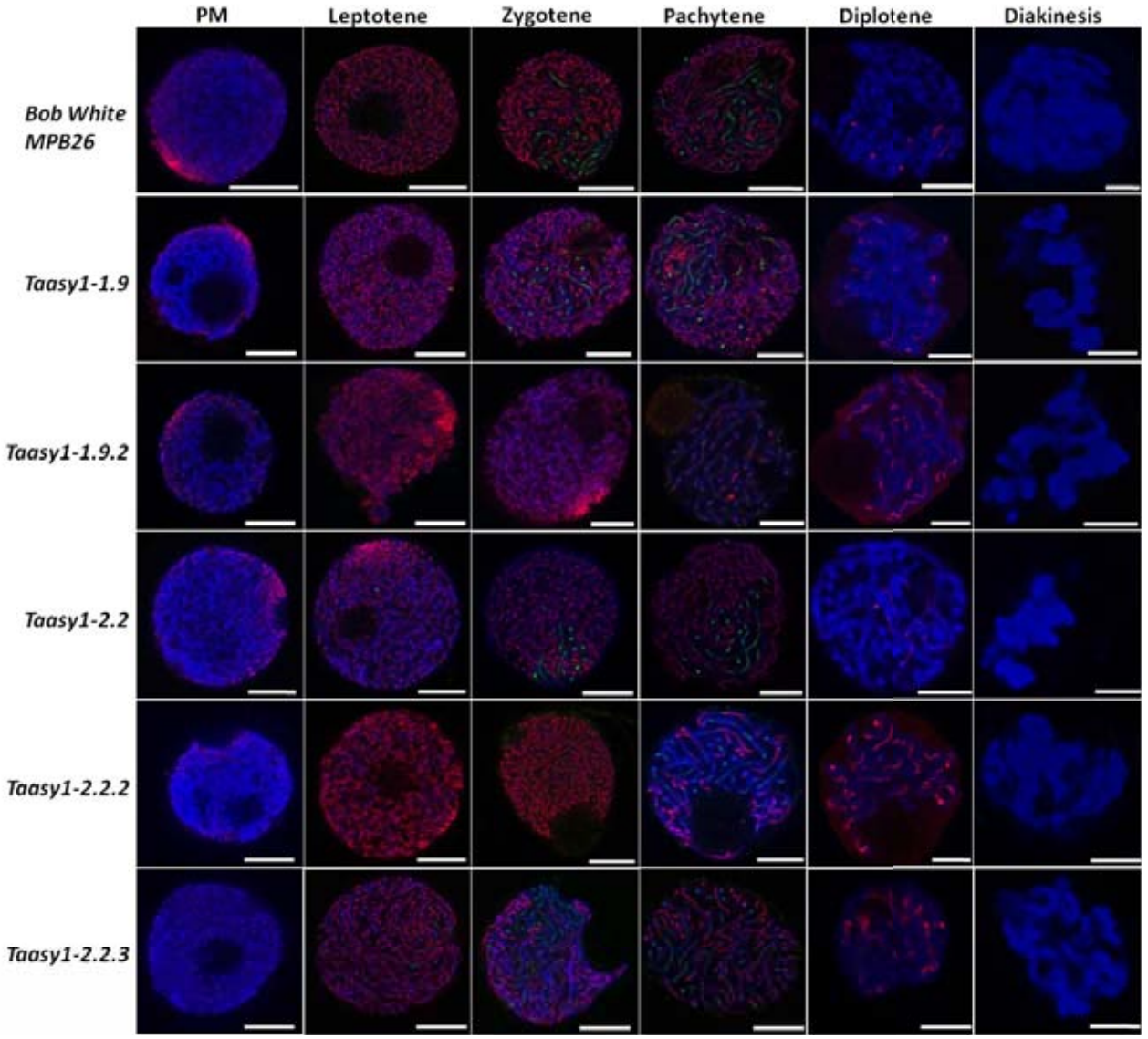


Figure 5.12 - Localisation of *Ta*ZYP1 and *Ta*ASY1 across five *Taasy1* RNAi knock-down mutants. In general, *Ta*ZYP1 (green) localisation within the *Taasy1* mutants was similar to that of wild-type with diffuse disrupted tracts and punctate foci observed during pachytene (except for *Taasy1-1.9.2*, which displayed lower levels of *Ta*ZYP1 in general). Compared to wild-type, *Ta*ASY1 (red) signals were diffuse and slightly weaker in the mutants and persisted at higher levels through to diplotene. Chromatin (blue) was counter-stained with DAPI. Scale bar, 7.5 µm.

5.4 – Discussion

The work presented in this chapter has characterised the ZYP1 bread wheat homologue at both the genetic and protein levels. This was achieved by first isolating the full-length coding sequence of the wheat *ZYP1* homologue using sequence information from a rice EST that showed high levels of similarity to the *AtZYP1a* and *AtZYP1b* genes. Higgins and colleagues (2005) reported that both *AtZYP1a* and *AtZYP1b* encode protein products highly similar to one another and which were functionally redundant. However, unlike Arabidopsis, allohexaploid bread wheat appears to have only one *ZYP1* gene, with copies on chromosome 2A, B and D. This was not unexpected as Higgins *et al.* (2005) previously hypothesised that the two Arabidopsis *ZYP1* genes were caused by a duplication event that occurred after the divergence of the Arabidopsis and brassica genera and that the *AtZYP1* genes were more closely-related to each other than they were to the only *ZYP1* homologue found in *Brassica oleracea*. More recent data strengthens this hypothesis, as both *Oryza sativa* and *Zea mays* have been reported to possess only a single *ZYP1* homologue (Golubovskaya *et al.* 2011, Wang *et al.* 2010).

Phylogenetic analysis revealed that *Ta*ZYP1 shares relatively high levels of sequence conservation with its homologues in other plant species, particularly with its monocotyledonous relatives (Figure 5.3). Levels of sequence conservation between *Ta*ZYP1 and its equivalents in non-plant species were modest at best. This was not unexpected as SC proteins have been previously reported to share conserved function but lack sequence conservation (Higgins *et al.* 2005).

Based on comparisons of similarities between the ZYP1 yeast homologue and other proteins with known tertiary structures, Sym and colleagues

hypothesised that ZYP1 proteins may form rod-like homodimers that are capped at the ends by globular N- and C-termini (Sym *et al.* 1993, Sym and Roeder 1995). This was later confirmed by Dong and Roeder in their observations of the yeast ZYP1 homologue protein by electron microscopy which also showed that the globular N-terminus was required for ZYP1 homodimer formation (Dong and Roeder 2000). Further work has shown that ZYP1 molecules attach to the lateral elements of sister chromosomes via their C-termini and extend outward allowing the N-termini of ZYP1 molecules on sister chromosomes to interact with one another and thus form the central element of the SC (reviewed by Page and Hawley 2004; and references therein). Amino acid sequence analysis of *Ta*ZYP1 revealed that both the N and C-termini of the protein are basic with high pI values, while predictive protein modeling revealed that the central region of *Ta*ZYP1 has a coiled-coil central region. While no dimerisation domains were found within the *Ta*ZYP1 N-terminus, future work involving alanine-scanning mutagenesis or partial deletions of the N-terminus may reveal whether such dimerisation domains exist within the region. Alternatively, yeast-two-hybrid assays conducted with the *Ta*ZYP1 N-terminus could also be performed to determine whether the *Ta*ZYP1 N-terminus self-interacts.

The high pI value of the C-terminus is particularly interesting as it may allow the lysine and arginine residues within it to bind DNA when the physiological pH of the cell is more acidic than the pI of the C-termini (which is 10.03). Positively-charged lysine and arginine residues have previously been shown to bind DNA in a non-specific manner in leucine zipper and helix-loop-helix motifs (Landschulz *et al.* 1988, Voronova and Baltimore 1990). In addition, the presence of a S/TPXX motif, previously characterised by Suzuki (1989),

within the *Ta*ZYP1 C-terminus provides further evidence that the *Ta*ZYP1 should also have a similar localisation pattern and arrangement within the SC as its previously characterised homologues.

cDART database searches with *Ta*ZYP1 returned moderate hits to two partial SMC conserved domains previously characterised in archaea (Figure 5.4). Both these domains have previously been reported in proteins that interact with DNA (Hirano and Hirano 2002, Jensen and Shapiro 2003). Phylogenetic analysis performed by Cobbe and Hecke (2004) have also shown that while these domains have been characterised in various species of archaea, these domains are also present in higher eukaryotic species.

To determine whether *Ta*ZYP1 is able to bind DNA, purified *Ta*ZYP1 protein heterologously-expressed in *E. coli* was used in competitive DNA-binding assays. The equal levels of retardation of both the ss- and dsDNA species indicates that *Ta*ZYP1 is able to bind both species of DNA indiscriminately under *in vitro* conditions (Figure 5.7). The observation that *Ta*ZYP1 is able to interact with ssDNA was not unexpected as studies in both mice and *C. elegans zyp1* homologue deletion mutants have shown that the recruitment of the ZYP1 homologues to DNA recombination sites is essential for the unloading of the recombination machinery during meiosis within both organisms (Colaiacovo *et al.* 2003, de Vries *et al.* 2005). Under wild-type conditions in mice and *C. elegans*, ZYP1 signal precedes the removal of the RAD51 foci signal (thus indicating that ZYP1 recruitment to regions of chromatin with RAD51 recombinase, and hence regions of DNA where single-stranded DNA invasion have occurred, is required before the recombination machinery can be removed). Work conducted by Higgins and colleagues (2005) on the Arabidopsis *spo11* and *dmc1* knock-out

mutants show that the Arabidopsis ZYP1 homologues failed to localise to chromatin in the *spo11* mutant but were localised as foci to sites of unsynapsed chromosome axes in the *dmc1* mutant. Taken together, these data indicate that ZYP1 localisation to chromatin in Arabidopsis is dependent on double-stranded break (DSB) formation but independent of recombination initiation, thus inferring that ZYP1 is most likely recruited to sites of recombination during the very early stages of single-strand DNA invasion (Higgins *et al.* 2005). However, ZYP1 polymerisation along the length of the chromosome is dependent on recombination occurring (Higgins *et al.* 2005).

Q-PCR profiling showed that *TaZYP1* expression fits the profile of a protein required during early meiosis (Figure 5.8). The relatively high expression level during the stages of pre-meiosis through to pachytene correlated well with observations in the immune-fluorescence localisation assays where *Ta*ZYP1 signal was first observed in leptotene and peaked at pachytene (Figure 5.11 row 1). The reduction in expression post-pachytene was also reflected in the immunolocalisation loading with the *Ta*ZYP1 signal quickly diminishing post-pachytene.

Comparisons of the *TaZYP1* expression profiles between the wild-type and *ph1b* mutant showed that while the overall expression profiles were similar, transcript levels were elevated in the *ph1b* mutant at all stages examined (Figure 5.8). Whether this increased *TaZYP1* expression can be linked to the deletion of the *Ph1* locus or whether it is a physiological reaction to the increased *Ta*ASY1 protein levels within the *ph1b* mutant, is yet to be determined. Indeed, the average *TaZYP1* expression levels were slightly lower (though these values were not statistically-significant) in all *Taasy1* mutants which showed reduced *TaASY1*

expression (Figure 5.9). However, *TaZYP1* expression was increased in the *Taasy1-2.2.3* mutant that also showed a large unexpected increase of *TaASY1* expression. With the correlation coefficient of *TaASY1-TaZYP1* gene expression between the *Taasy1* mutants and wild-type being highly significant (r = 0.9604), this may also indicate a regulatory connection between the two gene products.

The large difference in *TaZYP1* expression between the stage-specific wild-type in the Q-PCR data and the *Taasy1* mutant data is most likely an artifact of the experimental design which utilised cDNA from different tissue types (compare *TaZYP1* Figure 5.8 with Figure 5.9). While the stage-specific *TaZYP1* data was generated using cDNA synthesised from RNA extracted from anther tissue to generate the most accurate stage-specific data possible, cDNA synthesised from RNA extracted from whole spike tissue was used in the Q-PCR profiling of the Bob White MPB26 wild-type and the *Taasy1* mutants. The proportion of *TaZYP1* mRNA within the total RNA samples extracted from whole spike tissue would be significantly less than that of the anther tissue RNA samples and may have resulted in less *TaZYP1* mRNA copies being synthesised into cDNA. The unexpected increase in transcript levels of *TaASY1* and *TaZYP1* in the *Taasy1-2.2.3* mutant could also be attributed to the use of meiotic spike tissue. It is possible that the spike used for RNA extraction contained a larger proportion of pre-meiotic to pachytene stage anthers than the other spikes used for the mutants and wild-type. Furthermore, previous Q-PCR profiling of *TaASY1* in this mutant conducted by Dr Scott Boden (2008) showed that *TaASY1* levels were lower when compared to wild-type. The immuno-fluorescence data presented in this study also appears to agree with this explanation as there was no observable increase in

the intensity of either the *Ta*ASY1 or *Ta*ZYP1 signals in meiocytes of the *Taasy1-2.2.3* mutant compared to wild-type (compare Figure 5.12 row 1 with row 6).

Another plausible explanation is variation in gene silencing levels caused by the RNAi silencing pathway in wheat plants. Anand and colleagues (2003) have previously shown that variation in gene-silencing can occur within different tillers of a single RNAi knock-down wheat plant. While this may explain variation in the *TaASY1* expression, this explanation is not favoured as it fails to adequately explain the significantly increased level of *TaZYP1* expression in the *Taasy1-2.2.3* mutant.

Having characterised the DNA-binding ability of *Ta*ZYP1 *in vitro* as well as its expression profile in wild-type, *ph1b* and *Taasy1* mutants, immuno-fluorescence localisation experiments were conducted to investigate the localisation profile of *Ta*ZYP1. The polyclonal antibody raised against the full-length *Ta*ZYP1 protein initially used displayed high levels of non-specific binding. Optimisation steps taken such as purification of the sera by depleting the sera of antibodies that were not specific for *Ta*ZYP1 using bacterial acetone powder; and varying the antibody concentration, all proved unsuccessful (Figure 5.10). In an effort to overcome this, a polyclonal antibody raised against a partial peptide of *Ta*ZYP1 was generated. This new antibody yielded significantly better results with high levels of binding specificity.

The temporal and spatial localisation profile of *Ta*ZYP1 within both the Chinese Spring and Bob White MPB26 wild-types were identical with faint punctate foci associated with chromatin first observed during leptotene (Figure 5.11 row 1 and Figure 5.12 row 1). *Ta*ZYP1 signal then spread along the condensing chromatin and reached maximum intensity as long condensed signals

during pachytene before dissipating in diplotene. This data confirmed the Q-PCR profiling data, while association of the *Ta*ZYP1 signal with chromatin confirmed the results of the DNA-binding assay.

Although *Ta*ZYP1 and *Ta*ASY1 were both present on chromatin, they did not co-localise (Figure 5.11 and Figure 5.12). Even in the few instances where both signals were observed on the same regions of chromatin simultaneously, no co-localisation signals were detected. This observation indicates that *Ta*ZYP1, like its homologues in other species, appears to be a transverse filament protein and would therefore not be expected to co-localise with *Ta*ASY1 (which associates with the axial and lateral elements of chromatin). Furthermore, the observed polarity in the localization of *Ta*ZYP1 to the ends of chromosomes during zygotene could possibly predispose these regions of chromatin to recombination and account for the distal localization of chiasmata seen in bread wheat. Another important observation is that the *Ta*ZYP1 signal appears to be loaded onto regions of chromatin as *Ta*ASY1 is being unloaded. This data fits well with the hypothesis that *Ta*ASY1 may somehow directly/indirectly mediate recruitment of *Ta*ZYP1 onto the chromatin. Interestingly, this aspect of the *Ta*ZYP1 localisation pattern is in contrast to the localisation of its homologues in *Arabidopsis thaliana* and *Secale cereale*, where the ZYP1 homologues load onto the chromatin while ASY1 signal is still present, forming a co-aligned ZYP1 signal that is sandwiched between the ASY1 signal associated to axial elements on both sides (Higgins *et al.* 2005, Mikhailova *et al.* 2006). It is also noteworthy that the rye ZYP1 homologue signal already appears as linear tracts during leptotene (Mikhailova *et al.* 2006), unlike the punctate foci observed during the leptotene stage of both Arabidopsis (Higgins *et al.* 2005), rice (Wang *et al.* 2010)

and bread wheat in this study. This clearly shows that differences are present within the temporal loading patterns of the ZYP1 protein in different species.

Taken together, the DNA-binding data showing that *Ta*ZYP1 interacts with ssDNA, the fact that the Arabidopsis ZYP1 homologues are recruited to sites of recombination during the very early stages of single-stranded DNA invasion, and that the same observation of punctate foci were observed during leptotene in Arabidopsis, rice and bread wheat; suggest that *Ta*ZYP1 may be recruited to chromatin in a temporal and spatial pattern similar to Arabidopsis ZYP1. However, the absence of tracts of tripartite signals consisting of a *Ta*ZYP1 signal sandwiched on both sides by *Ta*ASY1 as well as the observation that *Ta*ZYP1 signal appears to lengthen into regions where *Ta*ASY1 is being unloaded is at odds with this conclusion. This raises the question of when and where exactly *Ta*ZYP1 is recruited onto chromatin in bread wheat.

The presence of relatively high levels of *Ta*ASY1 signal in meiocytes of the *ph1b* and *Taasy1* mutants during diplotene (Figure 5.11 and Figure 5.12) is also of interest. These increased levels of *Ta*ASY1 observed are probably the result of different mechanisms; as *Ta*ASY1 protein levels are significantly higher in the *ph1b* mutant but significantly lower in the *Taasy1* knock-down mutants (Boden *et al.* 2009). Boden and colleagues (2009) previously hypothesised that the *Ph1* locus is responsible for regulation of *TaASY1* and showed that in its absence, *TaASY1* expression is up-regulated approximately 20-fold compared to wild-type, with high levels of *Ta*ASY1 protein seen in the *ph1b* mutant meiocytes. Therefore, the *Ta*ASY1 signal seen in diplotene *ph1b* meiocytes could possibly be a result of too much *Ta*ASY1 protein present within the cells,

requiring more time for the cellular machinery to unload the *Ta*ASY1 molecules which leads to persistent *Ta*ASY1 signal in later stages of meiosis.

In contrast, the persisting *Ta*ASY1 signal in the *Taasy1* mutants could possibly be explained by a cellular compensation mechanism that is delaying the unloading of the *Ta*ASY1 protein. Previous work in *Taasy1* and *asy1* homologue mutants have shown that chromosomes in these mutants lack a true pachytene stage and that levels of chromosome synapsis were reduced (Boden *et al.* 2009, Nonomura *et al.* 2004, Ross *et al.* 1997). Perhaps cell cycle check-point mechanisms are trying to lengthen the duration for which *Ta*ASY1 is loaded onto the axial elements to allow proper synapsis to occur along the lengths of the homologous chromosomes.

In this study, the *Ta*ZYP1 signal was observed to be loaded onto the chromatin as *Ta*ASY1 signal is being unloaded in Chinese Spring and Bob White MPB26 wild-types as well as the *Taasy1* mutants. This then raises the question of whether *Ta*ZYP1 loading is completed within the *Taasy1* mutants due to presence of *Ta*ASY1 signal in the diplotene stage where *Ta*ZYP1 signal was no longer observable. Work conducted in the Arabidopsis *asy1* mutant previously showed that the Arabidopsis ZYP1 homologues are loaded onto the chromatin as foci during leptotene but subsequently fail to lengthen into linear tracts (Sánchez-Morán *et al.* 2007). This again is at odds with the observations in this study which clearly show that the *Ta*ZYP1 signal still forms linear tracts in the *Taasy1* mutants analysed. One possible explanation for this is the variation in severity of the *asy1* knock-down effect in the mutants (Boden *et al.* 2009, Caryl *et al.* 2000). While the *ASY1* gene expression data reported in both species are incomparable due to different gene expression analysis techniques being used, inferences could be

made about the severity of the *ASY1* knock-down effect by analysis of the fertility levels reported. While the Arabidopsis *asy1* mutant has a more severe knock-down effect with only 10% fertility, the average fertility of the *Taasy1* T_2 mutants was 77.74%. This possibly indicates that higher levels of ASY1 protein were present in the *Taasy1* mutants compared to that in the *Atasy1* mutant thus leading to a less severe effect on *Ta*ZYP1 loading in the *Taasy1* mutant meiocytes.

The main objective of the work presented in this chapter was to isolate and characterise the bread wheat ZYP1 homologue to further the current knowledge-base of bread wheat meiosis as well as to determine the role of ZYP1 in a complex hexaploid. This was achieved at the gene level by isolating the *TaZYP1* coding sequence, and then conducting Q-PCR expression profiling across wild-type, *ph1b*, and five *Taasy1* mutants. Characterisation of *Ta*ZYP1 *in vitro* determined its DNA-binding capabilities while dual immuno-fluorescence localisation experiments with a polyclonal anti-*Ta*ZYP1 antibody on both Chinese Spring and Bob White MPB26 wild-type meiocytes detailed its localisation profile during the early stages of meiosis both temporally and spatially. *Ta*ZYP1 localisation was also investigated in the *ph1b* and *Taassy1* mutants in an effort to further our understanding of *Ta*ZYP1 in the early stages of meiosis in bread wheat. While detailed characterisation of *Ta*ZYP1 has been achieved, further work is required to fully unravel the localisation profile of *Ta*ZYP1.

Chapter 6 – General discussion

The primary aim of this project was to identify and characterise proteins that are required during the early stages of meiosis in bread wheat. This was achieved by using a two-pronged approach; firstly utilising a large scale proteome-wide strategy to identify novel meiosis candidates, and secondly via gene-targeting to isolate and characterise the wheat homologues of proteins that have previously been reported to have roles in meiosis in other organisms.

6.1 – A large scale proteomics investigation of meiosis

6.1.1 – Optimisation of 2DGE for studying wheat meiosis

The 2DGE results within this study revealed that the largest proportion of protein species isolated from wheat meiotic tissue using the optimised TCA-acetone extraction technique had isoelectric potentials that were within the pH 5-8 range. Optimisation for the various 2DGE parameters were also performed to ensure that the best possible sample focussing was achieved. While the ultra-sensitive DIGE technology was initially investigated as a technique to visualise protein spots on the 2D gels, this technique yielded limited results. Consequently, this was replaced with a mass-spectrometry compatible EBT-silver stain technique. Comparisons of 2D gels loaded with wild-type whole anther protein extracts from various stages of meiosis revealed highly complex proteome profiles with a large number of protein species detected.

When shifting to meiocyte-enriched protein extracts, coupled with the EBT silver-staining technique, six protein spots that were differentially-expressed

between the different stages of meiosis and/or genotypes analysed were identified. Even so, the data obtained from this study may only represent a subset of the total proteomic profiles present; as different protein extraction techniques can yield different protein subsets (Carpentier *et al.* 2005). Furthermore, many more proteins that were present at different abundances between the stages and genotypes could also have been identified had the DIGE visualisation technique been more successful. Indeed, microarray data generated by Crismani *et al.* (2006) suggests that large numbers of transcript species are differentially-expressed at different stages of meiosis which could possibly translate to differences in the levels of the protein products they encode.

6.1.2 – Future directions for bread wheat meiosis using proteomics

Protein extraction techniques used in proteomics experiments play a large role in determining which protein subsets are isolated and analysed. Within this study, an optimised TCA-acetone extraction and precipitation technique that yielded substantial amounts of intact total cellular protein was used. However, a drawback of this technique was the high level of residual salt left within the samples. Ionic salt contamination is a leading cause of poor isoelectric focusing during the first dimension resolution of 2DGE samples. To overcome this, an additional wash step using 90% acetone (v/v) with sterile nanopure water was added to remove the ionic salt contaminants. In repeating such a study in the future, sample dialysis could be investigated to further reduce the residual levels of salt. While this would lead to purer samples, it may also result in substantial protein loss due to precipitation of protein species out of solution during the dialysis process. Due to the large amount of time required to harvest enough starting material (especially

so for meiocyte-enriched samples) and the limited amount of protein available in this study, this approach was non-viable and not investigated.

While optimisation of both horizontal and vertical resolution were deemed sufficient based on the 2D gel results reported in chapter 2, future research could investigate larger gel formats to achieve greater resolution. Although the use of narrow range pH strips gave adequate resolution for the isolation of six differentially-expressed spots, future work could also include using the micro-range pH IEF strips that allow resolution to a single pH unit. These narrow range strips would assist analysis by increasing the distance between each spot thus eliminating in-gel ambiguities such as that seen between candidate spots KK03 and KK04 (which both migrated to almost identical coordinates within the gels). These strips would also be useful for protein isoform analysis.

A setback of the proteomics research in this study was that the DIGE method could not be optimised (within a reasonable timeframe and within budget). Although optimisation steps were made in consultation with DIGE product specialists (Dr Sherif Tawfilis and Dr Tanya Lewanowitsch) and Dr Timothy Chataway (Head of Flinders Proteomics Facility), the results from DIGE visualisation were limited as the addition of the Cy-dye molecules caused inadequate isoelectric focusing to occur. One possible explanation for this is that the Cy-dye molecules were reacting with an unknown substance in the protein samples. While the unknown substance may have been an artefact from the protein extraction protocol used, this would seem unlikely as successful TCA-acetone protein extraction protocols with leaf and root tissue samples coupled with DIGE visualisation have previously been reported in the literature (Keeler *et al.* 2007). While another fluorescent dye visualisation technique (SYPRO® Ruby)

was also explored (data not shown), the levels of protein spot detection sensitivity were similar when compared to that of the EBT-silver stain.

One possible strategy for overcoming all of the problems posed by 2DGE and the inadequacies of the visualisation techniques is the use of gel-free proteomics technology. Multidimensional protein identification technology (MudPIT) (Lohrig and Wolters 2009, Wolters *et al.* 2001) is an automated shotgun proteomics technology that couples multidimensional liquid chromatography and electrospray ionisation tandem mass-spectrometry to identify peptides within complex samples. This system offers the advantage of reduced manual handling of the samples resulting in less chance of manual error, therefore lending itself to highly reproducible proteomics results. Furthermore, MudPIT does not require separation by 2DGE thus eliminating the issues of spot resolution and staining sensitivity. MudPIT protein samples are digested prior to MS/MS and therefore eliminates problems with protein insolubility that sometimes occur when dealing with 2DGE samples. The use of this technology however, will not overcome all the downstream obstacles of peptide identification such as the shifts in mass peaks due to the hundreds of post-translational modifications and their effects on MS/MS which have not been characterised.

6.1.3 – Proteomics and the identification of novel meiosis candidates

Based on the MS/MS data obtained from the six protein spots identified to be differentially-expressed, four candidate mRNA transcripts were isolated using a combination of 5′ and 3′ RACE PCR with full length coding transcripts obtained for KK01, KK03 and KK06 while a partial coding sequence was obtained for KK04. KK01 was characterised as a MATH-BTB speckle-POZ protein with

DNA-binding ability while the KK03 candidate was characterised as a pollen-specific SF21 protein. While the function of the KK03 SF21 protein is not clear, it may be required for pollen tube elongation based on data from previously characterised SF21 proteins (Allen *et al.* 2010, Lazarescu *et al.* 2006). The partial coding sequence of KK04 encodes a portion of a hexose transporter protein while the full-length KK06 candidate was characterised as a putative HSP70-2 plant homologue with a dnaK domain and DNA-binding ability. HSP70-2 variants in both humans and mice have previously been shown to form part of the SC and are also required for the disassembly of the SC (Allen *et al.* 1996, Dix *et al.* 1996, Dix *et al.* 1997, Zakeri *et al.* 1988).

Q-PCR profiling was conducted to investigate whether the gene expression profiles of these four candidates could explain the presence of their respective spots in the proteomics data. Comparisons of the transcriptomics and proteomics data revealed that the gene expression levels for three of the four candidates correlated with their protein expression profiles. Competitive DNA-binding assays with the KK01 and KK06 full-length protein products revealed that both these proteins interact with DNA, with KK01 preferentially binding to ssDNA and KK06 binding ssDNA and dsDNA equally.

Taking into account the data presented in Chapters 2 and 3, it is possible to conclude that a 2DGE proteomics approach is a viable *albeit* time-consuming strategy for identifying proteins with possible roles during the early stages of bread wheat meiosis. The successful isolation and characterisation of two novel DNA-interacting proteins, KK01 and KK06, highlights the success of this approach while future work characterising these two proteins further will increase our understanding of meiosis in bread wheat.

6.1.4 – Future research directions for the protein candidates identified

An important outcome of the research from this thesis was the characterisation of both the KK01 and KK06 proteins. Based on previously published literature, proteins with the MATH-BTB domain, which the KK01 candidate possesses, have been shown to interact with DNA and cause chromatin remodelling and gene expression regulation (Sunnerhagen *et al.* 1997, Wolffe 1997). Meanwhile, the KK06 candidate shares high levels of inter-species amino acid sequence similarity with HSP70-2 variants previously characterised in both human and mouse (Allen *et al.* 1996, Dix *et al.* 1996, Dix *et al.* 1997, Zakeri *et al.* 1988). KK06 therefore appears to be a good candidate as a HSP70-2 bread wheat homologue, with its gene expression profile and DNA-binding activity indicating it has the hallmarks of an SC protein. Future work should focus on investigating the roles of these proteins *in planta* through the use of antibodies raised against these proteins. Furthermore, 3-dimensional immuno-fluorescence localisation experiments of wheat meiocytes will determine the spatial and temporal localisation profiles during early meiosis while co-immunoprecipitation assays should reveal any protein partners that interact with these candidates within the cell.

The antibodies raised against these two proteins could also be used in translational proteomics assays to determine whether the KK01 and KK06 spots identified in Chapter 2 of this study do correspond to the protein spots of the KK01 and KK06 candidates. Further work encompassing localisation assays using meiocytes from the *ph1b* and *ph2a* mutants may also assist in elucidating the functions of these two proteins and whether either (or both) of these candidates behave differently in the *ph* mutants.

6.2 – A gene-targeted approach to study meiosis: *TaPHS1* and *TaZYP1*

6.2.1 – *TaPHS1*

The isolation and characterisation of the wheat PHS1 homologue, *TaPHS1*, revealed that it shares relatively high levels of sequence similarity with its equivalents in other plant species. Bioinformatic analysis of the *Ta*PHS1 amino acid sequence showed that it neither possesses any known nuclear localisation signal nor known conserved DNA-binding domains. Sequence alignments of PHS1 proteins from multiple species detected four conserved regions (termed Regions 1 to 4) within the PHS1 proteins analysed, with repeats found in both the Arabidopsis and wheat PHS1 sequences. However, the function of these oligopeptide repeats is not yet known.

One significant discovery made during the characterisation of the *Ta*PHS1 protein is its ability to interact with DNA *in vitro*. Further characterisation using partial peptides of the four conserved regions identified that Region 1 was responsible for the DNA-binding ability of *Ta*PHS1. Sequence analysis revealed that two S/TPXX DNA-binding motifs are present in Region 1 indicating that Region 1 may contain an uncharacterised DNA-binding domain. Interestingly, Region 1 coincides with the exact region in which the insertion mutations used to generate the *Atphs1-1* and *Zmphs1-0* knockout mutants are located (Ronceret *et al.* 2009, supplementary data).

3-dimensional immuno-fluorescence localisation experiments conducted with an antibody raised against *Ta*PHS1 localised the *Ta*PHS1 protein to the

nucleus of wheat meiocytes with the signal appearing as diffuse tracts with punctate foci that followed the chromatin. This data further supports the DNA-binding results and suggests that *Ta*PHS1 interacts with chromatin *in planta*. The ability to interact with DNA and its localisation to chromatin were not previously observed in the PHS1 localisation profiles of both *At*PHS1 and *Zm*PHS1. Therefore, this raises the question of whether *Ta*PHS1 functions as a direct transporter of meiotic proteins such as RAD50 into the nucleus as was hypothesised (but disfavoured) by Ronceret and colleagues (Ronceret *et al*. 2009). Other differences between the wheat PHS1 compared to both the Arabidopsis and maize PHS1 proteins was that the *Ta*PHS1 signal was observed on the nucleolar periphery; however its function there remains speculative. Data from dual immuno-fluorescence localisation of *Ta*PHS1 and *Ta*ASY1 revealed that while both signals were localised close to one another, the two proteins did not appear to co-localise.

6.2.2 – A model for homology searching in wheat

Based on observations of disrupted RAD50 localisation in the *phs1* mutants of both Arabidopsis and maize, Ronceret and colleagues (2009) hypothesised that PHS1 regulates RAD50 localisation into the nucleus. Considering the localisation profile of *Ta*PHS1, it is plausible to hypothesise that *Ta*PHS1 shuttles its meiotic protein partner(s) (such as RAD50) into the nucleus and may even be involved in the direct loading of protein(s) onto chromatin (as *Ta*PHS1 itself associates with chromatin). The localisation pattern of *Ta*PHS1 (diffuse tracts with punctate foci along regions of the chromatin) could also indicate multiple roles for *Ta*PHS1 during early meiosis (at least in bread wheat).

Firstly, regions of concentrated *Ta*PHS1 protein which form foci may indicate a role in the direct positioning of elements of the recombination machinery (such as RAD50) by *Ta*PHS1 at recombination sites. This is plausible as many recombination proteins such as RAD51, RAD50 and MLH3 (Franklin *et al.* 1999, Jackson *et al.* 2006, Ronceret *et al.* 2009) appear as foci associated to chromatin during the early stages of meiosis.

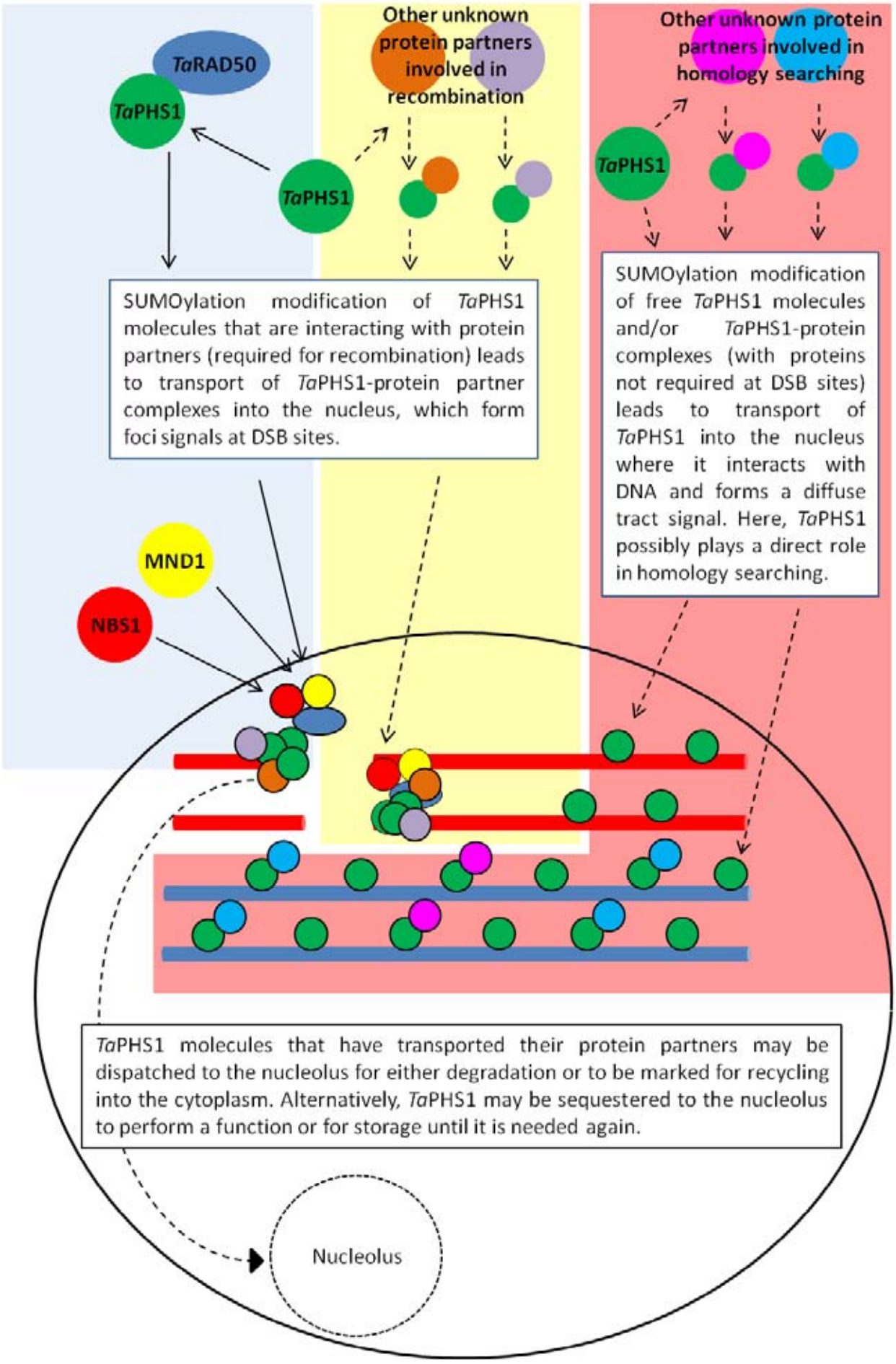
In contrast, the diffuse tracts of *Ta*PHS1 may indicate a direct role in the homology searching mechanism in bread wheat as both homology searching and recombination are tightly-linked processes. It is tempting to suggest that free *Ta*PHS1 molecules marked for nuclear entry by SUMOylation (or another as yet unknown nuclear localisation signal) move into the nucleus and form these diffuse tracts to perform homology searching. *Ta*PHS1 molecules may also interact and transport protein partner(s) which are required for homology searching but are not required at recombination sites. Previously, PHS1 has been indirectly linked with homology searching as the failure of the RAD51 recombinase protein to be recruited to chromatin in the *Zmphs1-0* mutant is caused by disrupted RAD50 localisation into the nucleus. The absence of RAD50 in the nucleus prevents the assembly of the MND1-RAD50-NBS1 complex (collectively known as the MRN complex) thus preventing resection of the DSB and the failed recruitment of the RAD51 recombinase (Pawlowski *et al.* 2004, Ronceret *et al.* 2009). Nuclear magnetic resonance (NMR) imaging of the DNA-RAD51 nucleoprotein filament structure (Nishinaka *et al.* 1998) and *in vitro* DNA pairing assays have shown that RAD51 is capable of homology searching and promotes homologous chromosome pairing over regions of DNA several kilobases in length (Eggler *et al.* 2002). However, it is debatable as to whether the lack of RAD51-based

homology searching is able to account for the 95% of synapsis that occurs between non-homologous chromosomes in the *Zmphs1-0* mutant. Furthermore, with the reach of the RAD51/DMC1 homology searching nucleoprotein filaments limited to only a few kilobases, *Ta*PHS1 molecules that form the diffuse tracts may somehow play a direct role in the homology searching process over longer distances of the chromatin. Additional data substantiating this hypothesis was uncovered by Osman and colleagues (2006) when they reported that both *At*ZYP1 and *Zm*PHS1 possess regions of peptide sequences that resemble bacterial topoisomerase IV domains. The bacterial topoisomerase IV proteins are members of the topoisomerase type IIA family previously hypothesised to have potential roles in inter-homologue chromosome resolution (von Wettstein 1984). This fits well into the model where *Ta*PHS1 may act as a component of the homology searching mechanism in wheat.

An additional homology searching mechanism may be necessary in the case of hexaploid bread wheat due to its large genome size (approximately 16,000 Mb), which contains long regions of repetitive sequence throughout. Based on comparisons of PHS1 function in both maize and Arabidopsis (two species with vastly different genome sizes and complexities), Ronceret and colleagues (2009) hypothesised that the target of PHS1-mediated chromosome pairing control must be in the gene-rich portions of both genomes. However, it is noteworthy that the maize *phs1* mutant showed a more severe phenotype and that maize has a genome approximately 20-fold larger than Arabidopsis. Extrapolation of the *phs1* phenotype based on genome size may prove fatal in hexaploid wheat with a genome approximately 128-fold larger than Arabidopsis; even if only gene-rich regions are considered.

Based on the data presented, could it be that *Ta*PHS1 functions differently from its two other known plant homologues? And is it possible that bread wheat has evolved a different mechanism for homology searching as it has for the control of homologous pairing with the *Ph1* locus?

Figure 6.1 – A model for *Ta*PHS1 homology searching in bread wheat. This model is based on combining data from previous literature as well as the results of this study. Blue region indicates previous function attributed to the maize PHS1 homologue by Ronceret *et al.* (2009) that hypothesised PHS1 regulates nuclear entry of RAD50. While the direct transport of RAD50 by PHS1 was suggested by Ronceret *et al.* (2009), it was disfavoured. However, the immunolocalisation profile of *Ta*PHS1 showing *Ta*PHS1 interacts with chromatin and the high levels of *Ta*PHS1 signal in the nucleolus suggest that *Ta*PHS1 may directly transport RAD50 into the nucleus. Furthermore, *Ta*PHS1 may aid in the loading of RAD50 onto the DSB sites of the chromatin due to its DNA-binding ability thus resulting in the formation of the *Ta*PHS1 foci signal. Within the nucleus, the MRN complex assembles at DSB sites and processes the DSBs to form single-stranded 3′ overhangs. This is followed by the recruitment of the RAD51 and DMC1 homology searching recombinases. The recruitment and direct transport of other unknown protein partners involved in recombination to the DSB sites by *Ta*PHS1 may also contribute to foci signal formation (yellow region). Both transport processes (blue and yellow region) show that *Ta*PHS1 has an indirect role in homology searching. Free *Ta*PHS1 and possible *Ta*PHS1-protein complexes (with protein partners not required at DSB sites) may also be marked for nucleus entry where the *Ta*PHS1 molecules form diffuse tracts along the chromatin to directly perform homology searching over larger distances compared to the RAD51/DMC1 nucleoprotein filament complexes (dull-orange region). *Ta*PHS1 molecules may then be moved to the nucleolus either for degradation, export back into cytoplasm, or sequestration into the nucleolus until they are required again (see following page).


Other unknown protein partners involved in recombination
Other unknown protein partners involved in homology searching
TaRAD50
TaPHS1
TaPHS1
TaPHS1
SUMOylation modification of TaPHS1 molecules that are interacting with protein partners (required for recombination) leads to transport of TaPHS1-protein partner complexes into the nucleus, which form foci signals at DSB sites.
SUMOylation modification of free TaPHS1 molecules and/or TaPHS1-protein complexes (with proteins not required at DSB sites) leads to transport of TaPHS1 into the nucleus where it interacts with DNA and forms a diffuse tract signal. Here, TaPHS1 possibly plays a direct role in homology searching.
MND1
NBS1
TaPHS1 molecules that have transported their protein partners may be dispatched to the nucleolus for either degradation or to be marked for recycling into the cytoplasm. Alternatively, TaPHS1 may be sequestered to the nucleolus to perform a function or for storage until it is needed again.
Nucleolus

6.2.3 – Future directions of *Ta*PHS1 research

While chapter 4 reported on the characterisation of *Ta*PHS1, more research could be conducted to understand the role(s) of this protein within wheat meiocytes. Firstly, temporal localisation of *Ta*PHS1 in relation to other known meiotic proteins (*e.g.* RAD50) should be investigated using dual 3-dimensional immuno-fluorescence localisation. Data generated from such experiments would provide valuable insight into the precise order in which these meiotic proteins are loaded onto the chromatin in bread wheat. Co-immunoprecipitation assays using the anti-*Ta*PHS1 antibody should also be conducted to determine whether *Ta*PHS1 interacts directly with other proteins within the cell *in planta*.

Another possible approach to identify protein partners of *Ta*PHS1 would be to use the pGADT7-*TaPHS1* prey plasmid generated in this study to screen a Y2H bait library constructed with total meiotic cDNA. In tandem, the co-immunoprecipitation assays and Y2H screen could prove complimentary and yield significant results, especially when paired with data from the 3-dimensional immuno-fluorescence experiments proposed. Furthermore, a bait vector library made from portions of the *TaRAD50* coding sequence isolated may determine whether PHS1 directly interacts with RAD50 and causes its transport into the nucleus. Tagging of the *Ta*PHS1 protein with GFP, coupled with live cell imaging, could reveal when *Ta*PHS1 protein molecules are sequestered to the nucleolus and whether these molecules are then degraded or recycled to the cytoplasm for transport of additional proteins into the nucleus.

Other characterisation experiments could revolve around the deletion of the oligopeptide repeat region within the *Ta*PHS1 sequence to determine whether those repeats play a role (be it structural or otherwise) within the *Ta*PHS1 protein.

Last but not least, analysis of *Taphs1* knock-down/knock-out mutants will help further our understanding of the role(s) of *Ta*PHS1 in bread wheat.

6.2.4 – *Ta*ZYP1

The isolation and characterisation of *TaZYP1* revealed that it shared relatively high levels of sequence similarities with its homologues in other plant species while Q-PCR profiling showed that it is highly expressed during early meiosis. There also appeared to be a correlation between the expression profiles of *TaZYP1* and *TaASY1*. Competitive DNA-binding assays showed that *Ta*ZYP1 interacted with both ss- and dsDNA with no preference for either DNA species *in vitro*, while 3-dimensional fluorescence immuno-localisation suggests that *Ta*ZYP1 was first loaded onto chromatin during leptotene as foci that then lengthened into linear tracts during zygotene and pachytene before dissipating by diplotene. Analysis of the *ph1b* and five *taasy1* knock-down mutants showed that *Ta*ZYP1 localisation, in general, was similar to that of the wild-type. However, a delay in unloading of the *Ta*ZYP1 signal was observed as evidenced during diplotene in both the *ph1b* and *taasy1* meiocytes (where higher levels of *Ta*ZYP1 signal were retained when compared to wild-type). In both wild-type and the *Taasy1* mutants, the *Ta*ZYP1 signal appeared to be loading onto regions of chromatin as *Ta*ASY1 was being unloaded. This was in contrast to previous reports in Arabidopsis, rice and rye that showed the ZYP1 signals within these species appearing while ASY1 signal was still present on the chromatin (Higgins *et al.* 2005, Mikhailova *et al.* 2006, Wang *et al.* 2010). While this anomaly can be explained by technical differences (e.g. the degree of anther fixation), and/or personal interpretation of the stage of a given meiocyte; it may also represent a

true result that reflects a different mechanism for SC formation that has evolved in bread wheat. Furthermore, although ZYP1 and its homologues in other organisms are thought to have a conserved function, immuno-fluorescence localisation experiments in Arabidopsis (Higgins *et al.* 2005), rice (Wang *et al.* 2010), and rye (Mikhailova *et al.* 2006) have revealed that the ZYP1 protein is loaded differently onto the chromatin of these species during leptotene.

The differences in the loading pattern of *Ta*ZYP1 compared to its other homologues with regards to ASY1 raises questions as to the precise temporal loading of *Ta*ZYP1 in wheat. In addition, the presence of *Ta*ZYP1 foci during leptotene may suggest that *Ta*ZYP1 is recruited to recombination sites. In light of the different systems regulating SC formation in various species previously studied, these data highlight the need for further work to be conducted to determine whether SC formation in wheat is regulated by recombination initiation.

6.2.5 – Future directions of *Ta*ZYP1 research

Future research involving *Ta*ZYP1 should involve dual-immunofluorescence localisation of *Ta*ZYP1 in conjunction with an anti-RAD51 or anti-DMC1 antibody to determine if and when the *Ta*ZYP1 foci seen during early leptotene localise to recombination sites. Data obtained from such experiments may determine the order in which these proteins are loaded onto chromatin. However, antibodies would first need to be raised against the wheat RAD51 and/or DMC1 proteins. This is because the widely used anti-*Hs*RAD51 antibody previously reported to bind the RAD51 homologues of many plant species (Franklin *et al.* 1999, Li *et al.* 2007, Terasawa *et al.* 1995) does not work in 3D-

immunofluorescence localisation assays using wheat meiocytes (Khoo, *unpublished data*) although the antibody was able to detect the wheat RAD51 protein in western blots (Khoo *et al.* 2008). Furthermore, the widely-used anti-*At*DMC1 antibody does not bind to the wheat DMC1 homologue either. This is more than likely a consequence of the antibody having been raised against the first 18 amino acids of *At*DMC1, of which only seven residues are identical in the equivalent *Ta*DMC1 sequence (Jolly 2010). Ideally, to answer the question of whether SC formation and *Ta*ZYP1 loading onto chromatin is mediated by recombination initiation, the anti-*Ta*ZYP1 antibody should also be used in fluorescence immuno-localisation experiments using meiocytes from wheat *spo11* as well as *dmc1* knock-out mutants.

Other future experiments using the anti-*Ta*ZYP1 antibody developed could involve co-immunoprecipitation assays to identify protein partners of *Ta*ZYP1. This would lead not only to a better understanding of *Ta*ZYP1 but may also allow the identification of novel uncharacterised proteins that have an important role in the formation of the SC in bread wheat.

6.3 – The bigger picture

While the isolation and characterisation of genes involved in bread wheat meiosis may appear to be purely fundamental, such studies contribute to the ever-growing body of knowledge which increases our understanding of this highly complex cell division process. For bread wheat researchers, a desire to understand meiosis is further driven by the promise that tools (i.e. bridging lines, new mutants in addition to *ph1b*) will one day lead to increasing the genetic variation within current breeding lines much more easily than what is currently possible. In order

to do so, inter-specific chromosomes must be able to interact with one another thereby allowing genetic recombination to occur thus leading to the acquisition of new genes (or the retrieval of genes previously lost through breeding practices) that may prove beneficial for current and future wheat breeding programs. The ability to control such introgression ‘events’ into current breeding lines would eventually enable superior wheat varieties to be generated so that increasing food demands can be met in an ever-changing agricultural environment caused by climate change and an increasing human population.

References

Able, J.A., Crismani, W. and Boden, S.A. (2009) Understanding meiosis and the implications for crop improvement. *Functional Plant Biology*, **36**, 575-588.

Able, J.A. and Langridge, P. (2006) Wild sex in the grasses. *Trends in Plant Science*, **11**, 261-263.

Able, J.A., Langridge, P. and Milligan, A.S. (2007) Capturing diversity in the cereals: many options but little promiscuity. *Trends in Plant Science*, **12**, 71-79.

Abranches, R., Beven, A.F., Aragon-Alcaide, L. and Shaw, P.J. (1998) Transcription sites are not correlated with chromosome territories in wheat nuclei. *Journal of Cell Biology*, **143**, 5-12.

Agarwal, S. and Roeder, G.S. (2000) Zip3 provides a link between recombination enzymes and synaptonemal complex proteins. *Cell*, **102**, 245-255.

Akhunov, E.D., Goodyear, A.W., Geng, S., Qi, L.L., Echalier, B., Gill, B.S., Miftahudin, Gustafson, J.P., Lazo, G., Chao, S.M., Anderson, O.D., Linkiewicz, A.M., Dubcovsky, J., La Rota, M., Sorrells, M.E., Zhang, D.S., Nguyen, H.T., Kalavacharla, V., Hossain, K., Kianian, S.F., Peng, J.H., Lapitan, N.L.V., Gonzalez-Hernandeiz, J.L., Anderson, J.A., Choi, D.W., Close, T.J., Dilbirligi, M., Gill, K.S., Walker-Simmons, M.K., Steber, C., McGuire, P.E., Qualset, C.O. and Dvorak, J. (2003) The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. *Genome Research*, **13**, 753-763.

Al-Kaff, N., Knight, E., Bertin, I., Foote, T., Hart, N., Griffiths, S. and Moore, G. (2008) Detailed dissection of the chromosomal region containing the *Ph1* locus in wheat *Triticum aestivum*: With deletion mutants and expression profiling. *Annals of Botany*, **101**, 863-872.

Allen, A.M., Lexer, C. and Hiscock, S.J. (2010) Characterisation of *sunflower-21* (*SF21*) genes expressed in pollen and pistil of *Senecio squalidus* (Asteraceae) and their relationship with other members of the *SF21* gene family. *Sexual Plant Reproduction*, **23**, 173-186.

Allen, J.W., Dix, D.J., Collins, B.W., Merrick, B.A., He, C., Selkirk, J.K., PoormanAllen, P., Dresser, M.E. and Eddy, E.M. (1996) HSP70-2 is part of the synaptonemal complex in mouse and hamster spermatocytes. *Chromosoma*, **104**, 414-421.

Anand, A., Trick, H.N., Gill, B.S. and Muthukrishnan, S. (2003) Stable transgene expression and random gene silencing in wheat. *Plant Biotechnology Journal*, **1**, 241-251.

Anfinsen, C. and Haber, E. (1961) Studies on reduction and re-formation of protein disulfide bonds. *Journal of Biological Chemistry*, **236**, 1361-1363.

Aragón-Alcaide, L., Reader, S., Miller, T. and Moore, G. (1997) Centromeric behaviour in wheat with high and low homoeologous chromosomal pairing. *Chromosoma*, **106**, 327-333.

Armstrong, S.J., Caryl, A.P., Jones, G.H. and Franklin, F.C.H. (2002) Asy1, a protein required for melotic chromosome synapsis, localizes to axis-

associated chromatin in *Arabidopsis* and *Brassica*. *Journal of Cell Science*, **115**, 3645-3655.

Armstrong, S.J., Franklin, F.C.H. and Jones, G.H. (2001) Nucleolus-associated telomere clustering and pairing precede meiotic chromosome synapsis in *Arabidopsis thaliana*. *Journal of Cell Science*, **114**, 4207-4217.

Armstrong, S.J., Franklin, F.C.H. and Jones, G.H. (2003) A meiotic time-course for *Arabidopsis thaliana*. *Sexual Plant Reproduction*, **16**, 141-149.

Armstrong, S.J. and Jones, G.H. (2003) Meiotic cytology and chromosome behaviour in wild-type *Arabidopsis thaliana*. *Journal of Experimental Botany*, **54**, 1-10.

Arumuganathan, K. and Earle, D.E. (1991) Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter*, **9**, 208-218.

Bass, H.W., Marshall, W.F., Sedat, J.W., Agard, D.A. and Cande, W.Z. (1997) Telomeres cluster de novo before the initiation of synapsis: A three-dimensional spatial analysis of telomere positions before and during meiotic prophase. *Journal of Cell Biology*, **137**, 5-18.

Bishop, D.K. and Zickler, D. (2004) Early decision: Meiotic crossover interference prior to stable strand exchange and synapsis. *Cell*, **117**, 9-15.

Bitton, D.A., Okoniewski, M.J., Connolly, Y. and Miller, C.J. (2008) Exon level integration of proteomics and microarray data. *BMC Bioinformatics*, **9**.

Bleuyard, J.Y., Gallego, M.E., Savigny, F. and White, C.I. (2005) Differing requirements for the Arabidopsis Rad51 paralogs in meiosis and DNA repair. *Plant Journal*, **41**, 533-545.

Bleuyard, J.Y., Gallego, M.E. and White, C.I. (2004) Meiotic defects in the *Arabidopsis rad50* mutant point to conservation of the MRX complex function in early stages of meiotic recombination. *Chromosoma*, **113**, 197-203.

Boden, S.A. (2008) Investigating chromosome pairing in bread wheat using ASYNAPSIS 1. In *School of Agriculture, Food & Wine*. Adelaide: University of Adelaide.

Boden, S.A., Langridge, P., Spangenberg, G. and Able, J.A. (2009) *Ta*ASY1 promotes homologous chromosome interactions and is affected by deletion of *Ph1*. *Plant Journal*, **57**, 487-497.

Boden, S.A., Shadiac, N., Tucker, E.J., Langridge, P. and Able, J.A. (2007) Expression and functional analysis of *Ta*ASY1 during meiosis of bread wheat (*Triticum aestivum*). *BMC Molecular Biology*, **8**, Article No.: 65.

Bolot, S., Abrouk, M., Masood-Quraishi, U., Stein, N., Messing, J., Feuillet, C. and Salse, J. (2009) The 'inner circle' of the cereal genomes. *Current Opinion in Plant Biology*, **12**, 119-125.

Borner, G.V., Kleckner, N. and Hunter, N. (2004) Crossover/noncrossover differentiation, synaptonemal complex formation, and regulatory surveillance at the leptotene/zygotene transition of meiosis. *Cell*, **117**, 29-45.

Bovill, W.D., Priyanka, D., Sanjay, K. and Able, J.A. (2009) Whole genome approaches to identify early meiotic gene candidates in cereals. *Functional & Integrative Genomics*, **9**, 219-229.

Bowers, J.E., Chapman, B.A., Rong, J.K. and Paterson, A.H. (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*, **422**, 433-438.

Bradford, M.M. (1976) Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding. *Analytical Biochemistry*, **72**, 248-254.

Carlton, P.M. and Cande, W.Z. (2002) Telomeres act autonomously in maize to organize the meiotic bouquet from a semipolarized chromosome orientation. *Journal of Cell Biology*, **157**, 231-242.

Carmo-Fonseca, M., Mendes-Soares, L. and Campos, I. (2000) To be or not to be in the nucleolus. *Nature Cell Biology*, **2**, E107-E112.

Carpentier, S.C., Witters, E., Laukens, K., Deckers, P., Swennen, R. and Panis, B. (2005) Preparation of protein extracts from recalcitrant plant tissues: An evaluation of different methods for two-dimensional gel electrophoresis analysis. *Proteomics*, **5**, 2497-2507.

Caryl, A.P., Armstrong, S.J., Jones, G.H. and Franklin, F.C.H. (2000) A homologue of the yeast *HOP1* gene is inactivated in the Arabidopsis meiotic mutant *asy1*. *Chromosoma*, **109**, 62-71.

Casasoli, M., Spadoni, S., Lilley, K.S., Cervone, F., De Lorenzo, G. and Mattei, B. (2008) Identification by 2-D DIGE of apoplastic proteins regulated by oligogalacturonides in *Arabidopsis thaliana*. *Proteomics*, **8**, 1042-1054.

Cavell, A.C., Lydiate, D.J., Parkin, I.A.P., Dean, C. and Trick, M. (1998) Collinearity between a 30-centimorgan segment of *Arabidopsis thaliana* chromosome 4 and duplicated regions within the *Brassica napus* genome. *Genome*, **41**, 62-69.

Chen, G.A., Gharib, T.G., Huang, C.C., Taylor, J.M.G., Misek, D.E., Kardia, S.L.R., Giordano, T.J., Iannettoni, M.D., Orringer, M.B., Hanash, S.M. and Beer, D.G. (2002) Discordant protein and mRNA expression in lung adenocarcinomas. *Molecular & Cellular Proteomics*, **1**, 304-313.

Chen, Y.K., Leng, C.H., Olivares, H., Lee, M.H., Chang, Y.C., Kung, W.M., Ti, S.C., Lo, Y.H., Wang, A.H.J., Chang, C.S., Bishop, D.K., Hsueh, Y.P. and Wang, T.F. (2004) Heterodimeric complexes of Hop2 and Mnd1 function with Dmc1 to promote meiotic homolog juxtaposition and strand assimilation. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 10572-10577.

Chikashige, Y., Tsutsumi, C., Yamane, M., Kamasa, K., Haraguchi, T. and Hiraoka, Y. (2006) Meiotic proteins Bqt1 and Bqt2 tether telomeres to form the bouquet arrangement of chromosomes. *Cell*, **125**, 59-69.

Clark, B.N. and Gutstein, H.B. (2008) The myth of automated, high-throughput two-dimensional gel analysis. *Proteomics*, **8**, 1197-1203.

Cobbe, N. and Heck, M.M.S. (2004) The evolution of SMC proteins: Phylogenetic analysis and structural implications. *Molecular Biology & Evolution*, **21**, 332-347.

Colaiacovo, M.P., MacQueen, A.J., Martinez-Perez, E., McDonald, K., Adamo, A., La Volpe, A. and Villeneuve, A.M. (2003) Synaptonemal complex assembly in *C. elegans* is dispensable for loading strand-exchange proteins but critical for proper completion of recombination. *Developmental Cell*, **5**, 463-474.

Colas, I., Shaw, P., Prieto, P., Wanous, M., Spielmeyer, W., Mago, R. and Moore, G. (2008) Effective chromosome pairing requires chromatin remodeling at the onset of meiosis. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 6075-6080.

Cowan, C.R. and Cande, W.Z. (2002) Meiotic telomere clustering is inhibited by colchicine but does not require cytoplasmic microtubules. *Journal of Cell Science*, **115**, 3747-3756.

Crismani, W., Baumann, U., Sutton, T., Shirley, N., Webster, T., Spangenberg, G., Langridge, P. and Able, J.A. (2006) Microarray expression analysis of meiosis and microsporogenesis in hexaploid bread wheat. *BMC Genomics*, **7**, 267.

Daoudal-Cotterell, S., Gallego, M.E. and White, C.I. (2002) The plant Rad50-Mre11 protein complex. *Febs Letters*, **516**, 164-166.

de Carvalho, C.E. and Colaiacovo, M.P. (2006) SUMO-mediated regulation of synaptonemal complex formation during meiosis. *Genes & Development*, **20**, 1986-1992.

de Vries, F.A.T., de Boer, E., van den Bosch, M., Baarends, W.M., Ooms, M., Yuan, L., Liu, J.G., van Zeeland, A.A., Heyting, C. and Pastink, A. (2005) Mouse Sycp1 functions in synaptonemal complex assembly, meiotic recombination., and XY body formation. *Genes & Development*, **19**, 1376-1389.

Dernburg, A.F., McDonald, K., Moulder, G., Barstead, R., Dresser, M. and Villeneuve, A.M. (1998) Meiotic recombination in *C. elegans* initiates by a conserved mechanism and is dispensable for homologous chromosome synapsis. *Cell*, **94**, 387-398.

Ding, D.Q., Yamamoto, A., Haraguchi, T. and Hiraoka, Y. (2004) Dynamics of homologous chromosome pairing during meiotic prophase in fission yeast. *Developmental Cell*, **6**, 329-341.

Dix, D.J., Allen, J.W., Collins, B.W., Mori, C., Nakamura, N., PoormanAllen, P., Goulding, E.H. and Eddy, E.M. (1996) Targeted gene disruption of *Hsp70-2* results in failed meiosis, germ cell apoptosis, and male infertility. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 3264-3268.

Dix, D.J., Allen, J.W., Collins, B.W., Poorman-Allen, P., Mori, C., Blizard, D.R., Brown, P.R., Goulding, E.H., Strong, B.D. and Eddy, E.M. (1997) HSP70-2 is required for desynapsis of synaptonemal complexes during meiotic prophase in juvenile and adult mouse spermatocytes. *Development*, **124**, 4595-4603.

Dong, F.G. and Jiang, J.M. (1998) Non-Rabl patterns of centromere and telomere distribution in the interphase nuclei of plant cells. *Chromosome Research*, **6**, 551-558.

Dong, H.J. and Roeder, G.S. (2000) Organization of the yeast Zip1 protein within the central region of the synaptonemal complex. *Journal of Cell Biology*, **148**, 417-426.

Eggler, A.L., Inman, R.B. and Cox, M.M. (2002) The Rad51-dependent pairing of long DNA substrates is stabilized by replication protein A. *Journal of Biological Chemistry*, **277**, 39280-39288.

Felsenstein, J. (1985) Confidence-limits on phylogenies - An approach using the bootstrap. *Evolution*, **39**, 783-791.

Fey, S.J. and Larsen, P.M. (2001) 2D or not 2D. *Current Opinion in Chemical Biology*, **5**, 26-33.

Foote, T., Roberts, M., Kurata, N., Sasaki, T. and Moore, G. (1997) Detailed comparative mapping of cereal chromosome regions corresponding to the *Ph1* locus in wheat. *Genetics*, **147**, 801-807.

Franklin-Tong, V.E., Drobak, B.K., Allan, A.C., Watkins, P.A.C. and Trewavas, A.J. (1996) Growth of pollen tubes of *Papaver rhoeas* is regulated by a slow-moving calcium wave propagated by inositol 1,4,5-trisphosphate. *Plant Cell*, **8**, 1305-1321.

Franklin, A.E., McElver, J., Sunjevaric, I., Rothstein, R., Bowen, B. and Cande, W.Z. (1999) Three-dimensional microscopy of the Rad51 recombination protein during meiotic prophase. *Plant Cell*, **11**, 809-824.

Fung, J.C., Marshall, W.F., Dernburg, A., Agard, D.A. and Sedat, J.W. (1998) Homologous chromosome pairing in Drosophila melanogaster proceeds through multiple independent initiations. *Journal of Cell Biology*, **141**, 5-20.

Futcher, B., Latter, G.I., Monardo, P., McLaughlin, C.S. and Garrels, J.I. (1999) A sampling of the yeast proteome. *Molecular and Cellular Biology*, **19**, 7357-7368.

Ghaemmaghami, S., Huh, W., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K. and Weissman, J.S. (2003) Global analysis of protein expression in yeast. *Nature*, **425**, 737-741.

Gill, K.S., Gill, B.S., Endo, T.R. and Mukai, Y. (1993) Fine physical mapping of *Ph1*, a chromosome-pairing regulator gene in polyploid wheat. *Genetics*, **134**, 1231-1236.

Golubovskaya, I.N., Harper, L.C., Pawlowski, W.P., Schichnes, D. and Cande, W.Z. (2002) The *pam1* gene is required for meiotic bouquet formation and efficient homologous synapsis in maize (*Zea mays* L.). *Genetics*, **162**, 1979-1993.

Golubovskaya, I.N., Wang, C.J.R., Timofejeva, L. and Cande, W.Z. (2011) Maize meiotic mutants with improper or non-homologous synapsis due to problems in pairing or synaptonemal complex formation. *Journal of Experimental Botany*, **62**, 1533-1544.

Görg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R. and Weiss, W. (2000) The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis*, **21**, 1037-1053.

Greenbaum, D., Jansen, R. and Gerstein, M. (2002) Analysis of mRNA expression and protein abundance data: an approach for the comparison of the enrichment of features in the cellular population of proteins and transcripts. *Bioinformatics*, **18**, 585-596.

Grelon, M., Vezon, D., Gendrot, G. and Pelletier, G. (2001) *AtSPO11-1* is necessary for efficient meiotic recombination in plants. *EMBO Journal*, **20**, 589-600.

Griffiths, S., Sharp, R., Foote, T.N., Bertin, I., Wanous, M., Reader, S., Colas, I. and Moore, G. (2006) Molecular characterization of *Ph1* as a major chromosome pairing locus in polyploid wheat. *Nature*, **439**, 749-752.

Gygi, S.P., Corthals, G.L., Zhang, Y., Rochon, Y. and Aebersold, R. (2000) Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 9390-9395.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H. and Aebersold, R. (1999a) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology*, **17**, 994-999.

Gygi, S.P., Rochon, Y., Franza, B.R. and Aebersold, R. (1999b) Correlation between protein and mRNA abundance in yeast. *Molecular and Cellular Biology*, **19**, 1720-1730.

Haber, J.E., Ira, G., Malkova, A. and Sugawara, N. (2004) Repairing a double-strand chromosome break by homologous recombination: revisiting Robin Holliday's model. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **359**, 79-86.

Haering, C.H., Farcas, A.M., Arumugam, P., Metson, J. and Nasmyth, K. (2008) The cohesin ring concatenates sister DNA molecules. *Nature*, **454**, 297-302.

Hamant, O., Ma, H. and Cande, W.Z. (2006) Genetics of meiotic prophase I in plants. *Annual Review of Plant Biology*, **57**, 267-302.

Harper, L., Golubovskaya, I. and Cande, W.Z. (2004) A bouquet of chromosomes. *Journal of Cell Science*, **117**, 4025-4032.

Harry, J.L., Wilkins, M.R., Herbert, B.R., Packer, N.H., Gooley, A.A. and Williams, K.L. (2000) Proteomics: Capacity versus utility. *Electrophoresis*, **21**, 1071-1081.

Hartung, F. and Puchta, H. (2000) Molecular characterisation of two paralogous *SPO11* homologues in *Arabidopsis thaliana*. *Nucleic Acids Research*, **28**, 1548-1554.

Hartung, F. and Puchta, H. (2001) Molecular characterization of homologues of both subunits A (SPO11) and B of the archaebacterial topoisomerase 6 in plants. *Gene*, **271**, 81-86.

Heyting, C. (1996) Synaptonemal complexes: Structure and function. *Current Opinion in Cell Biology*, **8**, 389-396.

Heyting, C. (2005) Meiotic transverse filament proteins: essential for crossing over. *Transgenic Research*, **14**, 547-550.

Higgins, J.D., Sanchez-Moran, E., Armstrong, S.J., Jones, G.H. and Franklin, F.C.H. (2005) The Arabidopsis synaptonemal complex protein ZYP1 is required for chromosome synapsis and normal fidelity of crossing over. *Genes & Development*, **19**, 2488-2500.

Hirano, M. and Hirano, T. (2002) Hinge-mediated dimerization of SMC protein is essential for its dynamic interaction with DNA. *EMBO Journal*, **21**, 5733-5744.

Holm, P.B. (1988) Chromosome-pairing and synaptonemal complex-formation in hexaploid wheat, nullisomic for chromosome-5B. *Carlsberg Research Communications*, **53**, 91-110.

Holm, P.B. and Wang, X.Z. (1988) The effect of chromosome-5B on synapsis and chiasma formation in wheat, *Triticum aestivum* cv Chinese Spring. *Carlsberg Research Communications*, **53**, 191-208.

Hotelier, T., Renault, L., Cousin, X., Negre, V., Marchot, P. and Chatonnet, A. (2004) ESTHER, the database of the alpha/beta-hydrolase fold superfamily of proteins. *Nucleic Acids Research*, **32**, 145-147.

Imin, N., Kerim, T., Weinman, J.J. and Rolfe, B.G. (2001) Characterisation of rice anther proteins expressed at the young microspore stage. *Proteomics*, **1**, 1149-1161.

Isaacson, T., Damasceno, C.M.B., Saravanan, R.S., He, Y., Catala, C., Saladie, M. and Rose, J.K.C. (2006) Sample extraction techniques for enhanced proteomic analysis of plant tissues. *Nature Protocols*, **1**, 769-774.

Islam, N., Tsujimoto, H. and Hirano, H. (2003) Proteome analysis of diploid tetraploid and hexaploid wheat: Towards understanding genome interaction in protein expression. *Proteomics*, **3**, 549-557.

Jackson, N., Sanchez-Moran, E., Buckling, E., Armstrong, S.J., Jones, G.H. and Franklin, F.C.H. (2006) Reduced meiotic crossovers and delayed prophase I progression in *At*MLH3-deficient Arabidopsis. *EMBO Journal*, **25**, 1315-1323.

Jensen, R.B. and Shapiro, L. (2003) Cell-cycle-regulated expression and subcellular localization of the *Caulobacter crescentus* SMC chromosome structural protein. *Journal of Bacteriology*, **185**, 3068-3075.

Jin, Q.W., Trelles-Sticken, E., Scherthan, H. and Loidl, J. (1998) Yeast nuclei display prominent centromere clustering that is reduced in nondividing cells and in meiotic prophase. *Journal of Cell Biology*, **141**, 21-29.

Jolly, H.R. (2010) Isolation and characterisation of wheat genes with early meiotic expression. In *School of Agriculture, Food & Wine*. Adelaide: University of Adelaide.

Katti, M.V., Sami-Subbu, R., Ranjekar, P.K. and Gupta, V.S. (2000) Amino acid repeat patterns in protein sequences: Their diversity and structural-functional implications. *Protein Science*, **9**, 1203-1209.

Keeler, M., Letarte, J., Hattrup, E., Hickman, F. and Haynes, P.A. (2007) Two-dimensional differential in-gel electrophoresis (DIGE) of leaf and roots of *Lycopersicon esculentum*. *Methods in Molecular Biology*, 157-174.

Keeney, S. (2001) Mechanism and control of meiotic recombination initiation. *Current Topics in Developmental Biology*, **52**, 1-53.

Kerzendorfer, C., Vignard, J., Pedrosa-Harand, A., Siwiec, T., Akimcheva, S., Jolivet, S., Sablowski, R., Armstrong, S., Schweizer, D., Mercier, R. and Schlogelhofer, P. (2006) The *Arabidopsis thaliana* MND1 homologue plays a key role in meiotic homologous pairing, synapsis and recombination. *Journal of Cell Science*, **119**, 2486-2496.

Khoo, K.H.P., Jolly, H.R. and Able, J.A. (2008) The *RAD51* gene family in bread wheat is highly conserved across eukaryotes, with *RAD51A* upregulated during early meiosis. *Functional Plant Biology*, **35**, 1267-1277.

Klapholz, S., Waddell, C.S. and Esposito, R.E. (1985) The role of the *Spo11* gene in meiotic recombination in yeast. *Genetics*, **110**, 187-216.

Klimyuk, V.I. and Jones, J.D.G. (1997) *AtDMC1*, the Arabidopsis homologue of the yeast *DMC1* gene: Characterization, transposon-induced allelic variation and meiosis-associated expression. *Plant Journal*, **11**, 1-14.

Klose, J. (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues - novel approach to testing for induced point mutations in mammals. *Humangenetik*, **26**, 231-243.

Klose, J. and Kobalz, U. (1995) 2-dimensional electrophoresis of proteins - An updated protocol and implications for a functional-analysis of the genome. *Electrophoresis*, **16**, 1034-1059.

Krauter-Canham, R., Bronner, R., Evrard, J.L., Hahne, G., Friedt, W. and Steinmetz, A. (1997) A transmitting tissue- and pollen-expressed protein from sunflower with sequence similarity to the human RTP protein. *Plant Science*, **129**, 191-202.

Krauter-Canham, R., Bronner, R. and Steinmetz, A. (2001) SF21 is a protein which exhibits a dual nuclear and cytoplasmic localization in developing pistils of sunflower and tobacco. *Annals of Botany*, **87**, 241-249.

Kreisberg, J.F., Betts, S.D., Haase-Pettingell, C. and King, J. (2002) The interdigitated beta-helix domain of the P22 tailspike protein acts as a molecular clamp in trimer stabilization. *Protein Science*, **11**, 820-830.

Kubista, H., Edelbauer, H. and Boehm, S. (2004) Evidence for structural and functional diversity among SDS-resistant SNARE complexes in neuroendocrine cells. *Journal of Cell Science*, **117**, 955-966.

Landschulz, W.H., Johnson, P.F. and McKnight, S.L. (1988) The leucine zipper - A hypothetical structure common to a new class of DNA-binding proteins. *Science*, **240**, 1759-1764.

Lazarescu, E., Friedt, W., Horn, R. and Steinmetz, A. (2006) Expression analysis of the sunflower *SF21* gene family reveals multiple alternative and organ-specific splicing of transcripts. *Gene*, **374**, 77-86.

Li, J., Harper, L.C., Golubovskaya, I., Wang, C.R., Weber, D., Meeley, R.B., McElver, J., Bowen, B., Cande, W.Z. and Schnable, P.S. (2007) Functional analysis of maize RAD51 in meiosis and double-strand break repair. *Genetics*, **176**, 1469-1482.

Lloyd, A.H., Milligan, A.S., Langridge, P. and Able, J.A. (2007) *TaMSH7*: a cereal mismatch repair gene that affects fertility in transgenic barley (*Hordeum vulgare L.*). *BMC Plant Biology*, **7**, 67.

Lohrig, K. and Wolters, D. (2009) Multidimensional protein identification technology. *Methods in Molecular Biology*, **564**, 143-153.

Mann, M., Hendrickson, R.C. and Pandey, A. (2001) Analysis of proteins and proteomes by mass spectrometry. *Annual Review of Biochemistry*, **70**, 437-473.

Manning, M. and Colon, W. (2004) Structural basis of protein kinetic stability: Resistance to sodium dodecyl sulfate suggests a central role for rigidity and a bias toward beta-sheet structure. *Biochemistry*, **43**, 11248-11254.

Mao, Q.L. and Pawliszyn, J. (1999) Effect of salt concentration on separation patterns in static capillary isoelectric focusing with imaging detection. *Journal of Chromatography B*, **729**, 355-359.

March, T.J., Able, J.A., Schultz, C.J. and Able, A.J. (2007) A novel late embryogenesis abundant protein and peroxidase associated with black point in barley grains. *Proteomics*, **7**, 3800-3808.

Martinez-Perez, E. and Moore, G. (2008) To check or not to check? The application of meiotic studies to plant breeding. *Current Opinion in Plant Biology*, **11**, 222-227.

Martínez-Pérez, E., Shaw, P., Aragon-Alcaide, L. and Moore, G. (2003) Chromosomes form into seven groups in hexaploid and tetraploid wheat as a prelude to meiosis. *Plant Journal*, **36**, 21-29.

Martinez-Pérez, E., Shaw, P. and Moore, G. (2001) The *Ph1* locus is needed to ensure specific somatic and meiotic centromere association. *Nature*, **411**, 204-207.

Martínez-Pérez, E., Shaw, P., Reader, S., Aragón-Alcaide, L., Miller, T. and Moore, G. (1999) Homologous chromosome pairing in wheat. *Journal of Cell Science*, **112**, 1761-1769.

Martínez-Pérez, E., Shaw, P.J. and Moore, G. (2000) Polyploidy induces centromere association. *Journal of Cell Biology*, **148**, 233-238.

Martinez, M., Cunado, N., Carcelen, N. and Romero, C. (2001) The *Ph1* and *Ph2* loci play different roles in the synaptic behaviour of hexaploid wheat *Triticum aestivum*. *Theoretical & Applied Genetics*, **103**, 398-405.

Masterson, J. (1994) Stomatal size in fossil plants - Evidence for polyploidy in majority of angiosperms. *Science*, **264**, 421-424.

Mello-Sampayo, T. (1968) Somatic association between telocentrics for both arms of the same chromosome in *Triticum aestivum*. *Proceedings of the XII International Congress of Genetics, Tokyo, Japan, August 19-28, 1968. Vol. 1. Abstracts of contributed papers.*, 163 p.

Mello-Sampayo, T. (1971) Promotion of homoelogous pairing in hybrids of *Triticum aestivum* x *Aegilops longissima*. *Gentica Iberica*, **23**, 1.

Mello-Sampayo, T. and Canas, A.P. (1973) Suppressors of meiotic chromosome pairing in common wheat. *Proceedings of the Fourth International Wheat Genetics Symposium. Cytogenetics.*, 709-713.

Mikhailova, E.I., Phillips, D., Sosnikhina, S.P., Lovtsyus, A.V., Jones, R.N. and Jenkins, G. (2006) Molecular assembly of meiotic proteins Asy1 and Zyp1 and pairing promiscuity in rye (*Secale cereale* L.) and its synaptic mutant sy10. *Genetics*, **174**, 1247-1258.

Mikhailova, E.I., Sosnikhina, S.P., Kirillova, G.A., Tikholiz, O.A., Smirnov, V.G., Jones, R.N. and Jenkins, G. (2001) Nuclear dispositions of subtelomeric and pericentromeric chromosomal domains during meiosis in asynaptic mutants of rye (*Secale cereale* L.). *Journal of Cell Science*, **114**, 1875-1882.

Moore, G. (2002) Meiosis in allopolyploids - the importance of 'Teflon' chromosomes. *Trends in Genetics*, **18**, 456-463.

Moore, G. and Shaw, P. (2009) Improving the chances of finding the right partner. *Current Opinion in Genetics & Development*, **19**, 99-104.

Moses, M.J. (1969) Structure and function of the synaptonemal complex. *Genetics*, **61**, Suppl:41-51.

Nabeshima, K., Kakihara, Y., Hiraoka, Y. and Nojima, H. (2001) A novel meiosis-specific protein of fission yeast, Meu13p, promotes homologous pairing independently of homologous recombination. *EMBO Journal*, **20**, 3871-3881.

Nag, D.K., Pata, J.D., Sironi, M., Flood, D.R. and Hart, A.M. (2006) Both conserved and non-conserved regions of Spo11 are essential for meiotic recombination initiation in yeast. *Molecular Genetics and Genomics*, **276**, 313-321.

Nas, T.M.S., Sanchez, D.L., Diaz, M.G.Q., Mendioro, M.S. and Virmani, S.S. (2005) Pyramiding of *thermosensitive genetic male sterility* (*TGMS*) genes and identification of a candidate *tms5* gene in rice. *Euphytica*, **145**, 67-75.

Nishinaka, T., Shinohara, A., Ito, Y., Yokoyama, S. and Shibata, T. (1998) Base pair switching by interconversion of sugar puckers in DNA extended by proteins of RecA-family: A model for homology search in homologous genetic recombination. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 11071-11076.

Nonomura, K.I., Nakano, M., Murata, K., Miyoshi, K., Eiguchi, M., Miyao, A., Hirochika, H. and Kurata, N. (2004) An insertional mutation in the rice *PAIR2* gene, the ortholog of Arabidopsis *ASY1*, results in a defect in homologous chromosome pairing during meiosis. *Molecular Genetics and Genomics*, **271**, 121-129.

O'Farrell, P.H. (1975) High-resolution 2-dimensional electrophoresis of proteins. *Journal of Biological Chemistry*, **250**, 4007-4021.

Okuda, T. and Kondoh, H. (1999) Identification of new genes *Ndr2* and *Ndr3* which are related to *Ndr1*/*RTP*/*Drg1* but show distinct tissue specificity and response to N-myc. *Biochemical and Biophysical Research Communications*, **266**, 208-215.

Olson, M.O.J., Dundr, M. and Szebeni, A. (2000) The nucleolus: an old factory with unexpected capabilities. *Trends in Cell Biology*, **10**, 189-196.

Ortega, S., Prieto, I., Odajima, J., Martin, A., Dubus, P., Sotillo, R., Barbero, J.L., Malumbres, M. and Barbacid, M. (2003) Cyclin-dependent kinase 2 is essential for meiosis but not for mitotic cell division in mice. *Nature Genetics*, **35**, 25-31.

Osman, K., Sanchez-Moran, E., Higgins, J.D., Jones, G.H. and Franklin, F.C.H. (2006) Chromosome synapsis in Arabidopsis: analysis of the transverse filament protein ZYP1 reveals novel functions for the synaptonemal complex. *Chromosoma*, **115**, 212-219.

Page, S.L. and Hawley, R.S. (2004) The genetics and molecular biology of the synaptonemal complex. *Annual Review of Cell and Developmental Biology*, **20**, 525-558.

Pawlowski, W.P. and Cande, W.Z. (2005) Coordinating the events of the meiotic prophase. *Trends in Cell Biology*, **15**, 674-681.

Pawlowski, W.P., Golubovskaya, I.N., Timofejeva, L., Meeley, R.B., Sheridan, W.F. and Cande, W.Z. (2004) Coordination of meiotic recombination, pairing, and synapsis by PHS1. *Science*, **303**, 89-92.

Peck, S.C. (2005) Update on proteomics in Arabidopsis. Where do we go from here? *Plant Physiology*, **138**, 591-599.

Peddinti, D., Nanduri, B., Kaya, A., Feugang, J.M., Burgess, S.C. and Memili, E. (2008) Comprehensive proteomic analysis of bovine spermatozoa of varying fertility rates and identification of biomarkers associated with fertility. *BMC Systems Biology*, **2**, Article No.: 19.

Peoples, T.L., Dean, E., Gonzalez, O., Lambourne, L. and Burgess, S.M. (2002) Close, stable homolog juxtaposition during meiosis in budding yeast is dependent on meiotic recombination, occurs independently of synapsis, and is distinct from DSB-independent pairing contacts. *Genes & Development*, **16**, 1682-1695.

Petersen, G., Seberg, O., Yde, M. and Berthelsen, K. (2006) Phylogenetic relationships of Triticum and Aegilops and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). *Molecular Phylogenetics and Evolution*, **39**, 70-82.

Petronczki, M., Siomos, M.F. and Nasmyth, K. (2003) Un menage a quatre: The molecular biology of chromosome segregation in meiosis. *Cell*, **112**, 423-440.

Pezza, R.J., Petukhova, G.V., Ghirlando, R. and Camerini-Otero, R.D. (2006) Molecular activities of meiosis-specific proteins Hop2, Mnd1, and the Hop2-Mnd1 complex. *Journal of Biological Chemistry*, **281**, 18426-18434.

Phillips, D., Mikhailova, E.I., Timofejeva, L., Mitchell, J.L., Osina, O., Sosnikhina, S.P., Jones, R.N. and Jenkins, G. (2008) Dissecting meiosis of rye using translational proteomics. *Annals of Botany*, **101**, 873-880.

Prieto, P., Shaw, P. and Moore, G. (2004) Homologue recognition during meiosis is associated with a change in chromatin conformation. *Nature Cell Biology*, **6**, 906-908.

Riley, R. (1975) Origins of wheat. *Bread. Social, nutritional and agricultural aspects of wheaten bread.*, 27-45.

Riley, R. and Chapman, V. (1958) Genetic control of the cytologically diploid behaviour of hexaploid wheat. *Nature*, **182**, 713-715.

Rolland, A.D., Evrard, B., Guitton, N., Lavigne, R., Calvel, P., Couvet, M., Jegou, B. and Pineau, C. (2007) Two-dimensional fluorescence difference gel electrophoresis analysis of spermatogenesis in the rat. *Journal of Proteome Research*, **6**, 683-697.

Ronceret, A., Doutriaux, M.P., Golubovskaya, I.N. and Pawlowski, W.P. (2009) PHS1 regulates meiotic recombination and homologous chromosome pairing by controlling the transport of RAD50 to the nucleus. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 20121-20126.

Ross, K.J., Fransz, P., Armstrong, S.J., Vizir, I., Mulligan, B., Franklin, F.C.H. and Jones, G.H. (1997) Cytological characterization of four meiotic mutants of Arabidopsis isolated from T-DNA-transformed lines. *Chromosome Research*, **5**, 551-559.

Rost, B., Yachdav, G. and Liu, J.F. (2004) The PredictProtein server. *Nucleic Acids Research*, **32**, W321-W326.

Saitou, N. and Nei, M. (1987) The neighbor-joining method - a new method for reconstructing phylogenetic trees. *Molecular Biology & Evolution*, **4**, 406-425.

San-Segundo, P.A. and Roeder, G.S. (1999) Pch2 links chromatin silencing to meiotic checkpoint control. *Cell*, **97**, 313-324.

Sánchez-Morán , E., Mercier, R., Higgins, J.D., Armstrong, S.J., Jones, G.H. and Franklin, F.C.H. (2005) A strategy to investigate the plant meiotic proteome. *Cytogenetic and Genome Research*, **109**, 181-189.

Sánchez-Morán, E., Santos, J.L., Jones, G.H. and Franklin, F.C.H. (2007) ASY1 mediates *At*DMC1-dependent interhomolog recombination during meiosis in Arabidopsis. *Genes & Development*, **21**, 2220-2233.

Scherthan, H., Eils, R., Trelles-Sticken, E., Dietzel, S., Cremer, T., Walt, H. and Jauch, A. (1998) Aspects of three-dimensional chromosome reorganization during the onset of human male meiotic prophase. *Journal of Cell Science*, **111**, 2337-2351.

Scherthan, H., Weich, S., Schwegler, H., Heyting, C., Harle, M. and Cremer, T. (1996) Centromere and telomere movements during early meiotic prophase of mouse and man are associated with the onset of chromosome pairing. *Journal of Cell Biology*, **134**, 1109-1125.

Schwacha, A. and Kleckner, N. (1997) Interhomolog bias during meiotic recombination: Meiotic functions promote a highly differentiated interhomolog-only pathway. *Cell*, **90**, 1123-1135.

Schwarzacher, T. (1997) Three stages of meiotic homologous chromosome pairing in wheat: cognition, alignment and synapsis. *Sexual Plant Reproduction*, **10**, 324-331.

Sears, E.R. (1952) Homoeologous chromosomes in *Triticum aestivum*. *Genetics*, **37**, 624.

Sears, E.R. (1976) Genetic-control of chromosome-pairing in wheat. *Annual Review of Genetics*, **10**, 31-51.

Sears, E.R. (1977) Induced mutant with homoeologous pairing in common wheat. *Canadian Journal of Genetics and Cytology*, **19**, 585-593.

Sears, E.R. and Okamoto, M. (1958) Intergenomic chromosome relationships in hexaploid wheat. *Proceedings of the X International Congress of Genetics, McGill University, Montreal, Canada*, **2**, 258-259.

Sharma, S., Chataway, T., Burdon, K.P., Jonavicius, L., Klebe, S., Hewitt, A.W., Mills, R.A. and Craig, J.E. (2009) Identification of LOXL1 protein and Apolipoprotein E as components of surgically isolated pseudoexfoliation material by direct mass spectrometry. *Experimental Eye Research*, **89**, 479-485.

Shinohara, A. and Shinohara, M. (2004) Roles of RecA homologues Rad51 and Dmc1 during meiotic recombination. *Cytogenetic and Genome Research*, **107**, 201-207.

Sorrells, M.E., La Rota, M., Bermudez-Kandianis, C.E., Greene, R.A., Kantety, R., Munkvold, J.D., Miftahudin, Mahmoud, A., Ma, X.F., Gustafson, P.J., Qi, L.L.L., Echalier, B., Gill, B.S., Matthews, D.E., Lazo, G.R., Chao, S.M., Anderson, O.D., Edwards, H., Linkiewicz, A.M., Dubcovsky, J., Akhunov, E.D., Dvorak, J., Zhang, D.S., Nguyen, H.T., Peng, J.H., Lapitan, N.L.V., Gonzalez-Hernandez, J.L., Anderson, J.A., Hossain, K., Kalavacharla, V., Kianian, S.F., Choi, D.W., Close, T.J., Dilbirligi, M., Gill, K.S., Steber, C., Walker-Simmons, M.K., McGuire, P.E. and Qualset, C.O. (2003) Comparative DNA sequence analysis of wheat and rice genomes. *Genome Research*, **13**, 1818-1827.

Stogios, P.J., Downs, G.S., Jauhal, J.J.S., Nandra, S.K. and Prive, G.G. (2005) Sequence and structural analysis of BTB domain proteins. *Genome Biology*, **6**.

Sunnerhagen, M., Nilges, M., Otting, G. and Carey, J. (1997) Solution structure of the DNA-binding domain and model for the complex of multifunctional hexameric arginine repressor with DNA. *Nature Structural Biology*, **4**, 819-826.

Sutton, T., Whitford, R., Baumann, U., Dong, C.M., Able, J.A. and Langridge, P. (2003) The *Ph2* pairing homoeologous locus of wheat (*Triticum aestivum*): identification of candidate meiotic genes using a comparative genetics approach. *Plant Journal*, **36**, 443-456.

Suzuki, M. (1989) SPXX, a frequent sequence motif in gene regulatory proteins. *Journal of Molecular Biology*, **207**, 61-84.

Sym, M., Engebrecht, J. and Roeder, G.S. (1993) ZIP1 is a synaptonemal complex protein required for meiotic chromosome synapsis. *Cell*, **72**, 365-378.

Sym, M. and Roeder, G.S. (1995) ZIP1-induced changes in synaptonemal complex structure and polycomplex assembly. *Journal of Cell Biology*, **128**, 455-466.

Szostak, J.W., Orrweaver, T.L., Rothstein, R.J. and Stahl, F.W. (1983) The double-strand-break repair model for recombination. *Cell*, **33**, 25-35.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology & Evolution*, **24**, 1596-1599.

Terasawa, M., Shinohara, A., Hotta, Y., Ogawa, H. and Ogawa, T. (1995) Localization of RecA-like recombination proteins on chromosomes of the lily at various meiotic stages. *Genes & Development*, **9**, 925-934.

Tian, Q., Stepaniants, S.B., Mao, M., Weng, L., Feetham, M.C., Doyle, M.J., Yi, E.C., Dai, H.Y., Thorsson, V., Eng, J., Goodlett, D., Berger, J.P., Gunter, B., Linseley, P.S., Stoughton, R.B., Aebersold, R., Collins, S.J., Hanlon, W.A. and Hood, L.E. (2004) Integrated genomic and proteomic analyses of gene expression in mammalian cells. *Molecular & Cellular Proteomics*, **3**, 960-969.

Unwin, R.D., Smith, D.L., Blinco, D., Wilson, C.L., Miller, C.J., Evans, C.A., Jaworska, E., Baldwin, S.A., Barnes, K., Pierce, A., Spooncer, E. and Whetton, A.D. (2006) Quantitative proteomics reveals posttranslational control as a regulatory factor in primary hematopoietic stem cells. *Blood*, **107**, 4687-4694.

Varshavsky, A. (1996) The N-end rule: Functions, mysteries, uses. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 12142-12149.

Viera, A., Rufas, J.S., Martinez, I., Barbero, J.L., Ortega, S. and Suja, J.A. (2009) CDK2 is required for proper homologous pairing, recombination and sex-body formation during male mouse meiosis. *Journal of Cell Science*, **122**, 2149-2159.

von Wettstein, D. (1984) The synaptonemal complex and genetic segregation. *Symposia of the Society for Experimental Biology*, 195-231.

Voronova, A. and Baltimore, D. (1990) Mutations that disrupt DNA-binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains. *Proceedings of the National Academy of Sciences of the United States of America*, **87**, 4722-4726.

Wall, A.M., Riley, R. and Chapman, V. (1971) Wheat mutants permitting homoeologous meiotic chromosome pairing. *Genetical Research*, **18**, 311-&.

Wang, M., Wang, K.J., Tang, D., Wei, C.X., Li, M., Shen, Y., Chi, Z.C., Gu, M.H. and Cheng, Z.K. (2010) The Central Element Protein ZEP1 of the Synaptonemal Complex Regulates the Number of Crossovers during Meiosis in Rice. *Plant Cell*, **22**, 417-430.

Wang, X.C., Li, X.F. and Li, Y.X. (2007) A modified Coomassie Brilliant Blue staining method at nanogram sensitivity compatible with proteomic analysis. *Biotechnology Letters*, **29**, 1599-1603.

Waters, K.M., Pounds, J.G. and Thrall, B.D. (2006) Data merging for integrated microarray and proteomic analysis. *Briefings in Functional Genomics and Proteomics*, **5**, 261-272.

Waterworth, W.M., Altun, C., Armstrong, S.J., Roberts, N., Dean, P.J., Young, K., Weil, C.F., Bray, C.M. and West, C.E. (2007) NBS1 is involved in DNA repair and plays a synergistic role with ATM in mediating meiotic homologous recombination in plants. *Plant Journal*, **52**, 41-52.

Weiner, B.M. and Kleckner, N. (1994) Chromosome pairing via multiple interstitial interactions before and during meiosis in yeast. *Cell*, **77**, 977-991.

Weschke, W., Panitz, R., Gubatz, S., Wang, Q., Radchuk, R., Weber, H. and Wobus, U. (2003) The role of invertases and hexose transporters in

controlling sugar ratios in maternal and filial tissues of barley caryopses during early development. *Plant Journal*, **33**, 395-411.

Wilson, P.J., Riggs, C.D. and Hasenkampf, C.A. (2005) Plant chromosome homology: hypotheses relating rendezvous, recognition and reciprocal exchange. *Cytogenetic and Genome Research*, **109**, 190-197.

Wold, F. (1981) *In vivo* chemical modifications of proteins (post-translational modification). *Annual Review of Biochemistry*, **50**, 783-814.

Wolffe, A.P. (1997) Transcriptional control - Sinful repression. *Nature*, **387**, 16-17.

Wolters, D.A., Washburn, M.P. and Yates, J.R. (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Analytical Chemistry*, **73**, 5683-5690.

Wu, F.S. and Wang, M.Y. (1984) Extraction of proteins for sodium dodecyl-sulfate polyacrylamide-gel electrophoresis from protease-rich plant tissues. *Analytical Biochemistry*, **139**, 100-103.

Yoder, M.D., Lietzke, S.E. and Jurnak, F. (1993) Unusual structural features in the parallel beta-helix in pectate lyases. *Structure*, **1**, 241-251.

Ytterberg, A.J. and Jensen, O.N. (2010) Modification-specific proteomics in plant biology. *Journal of Proteomics*, **73**, 2249-2266.

Zakeri, Z.F., Wolgemuth, D.J. and Hunt, C.R. (1988) Identification and sequence-analysis of a new member of the mouse *HSP70* gene family and characterisation of its unique cellular and developmental pattern of expression in the male germ line. *Molecular and Cellular Biology*, **8**, 2925-2932.

Zickler, D. and Kleckner, N. (1998) The leptotene-zygotene transition of meiosis. *Annual Review of Genetics*, **32**, 619-697.

Zickler, D. and Kleckner, N. (1999) Meiotic chromosomes: Integrating structure and function. *Annual Review of Genetics*, **33**, 603-754.

Zuckerkandl, E. and Pauling, L. (1965) Evolutionary divergence and convergence in proteins. In *Evolving genes and proteins* (Bryson, V. and Vogel, H.J. eds). New York: Academic Press, pp. 97-166.

Appendix A

Synthesis of 2% aceto-carmine

A solution of 55% acetic acid made using autoclaved nanopure water and glacial acetic acid (Chem-Supply, Gillman, South Australia, Australia) was heated using a Bunsen burner until boiling. Carmine natural red 4 powder (Sigma-Aldrich, St Louis, Montana, USA) (2% w/v) was added with the solution boiled for a few moments longer and then cooled. The stain solution was then filtered using moist Grade 2 Whatman Qualitative circle paper (Whatman International Ltd., Maidstone, England).

Appendix B1

Output files for KK01 protein spot sequencing

KK09B01.RAW Page 1

```
SRF File: E:\data\KELVIN\09-07-21\KK09B01.srf
Database... indexed - rice_all.pep.fasta.hdr (7/23/2009)
Filter(s)... xc (± 1,2,3)=1.50,2.00,2.50 ; peptide probability<=1e-003
Mods: (M* +15.99492) (STY# +79.96633)

Reference                                                          P      Score   Coverage   Accession
  Scan(s)       Sequence                                MH+     z    P       XC  DeltaCn   Sp   RSp  Ions

protein retrotransposon protein, putative, Ty3-gypsy subclass      4e-005  18.1     0.0       0
  8430        K.YKLKVVAADDVGSS#SSQDGK.G                2134.01   3  4e-005  2.747  0.014   210.5  2   20/114
  8430        K.YKLKVVAADDVGS#SSSQDGK.G                2134.01   3  4e-005  2.709  0.103   190.1  10  19/114

  2 of 3 peptide matches reported, 1 removed due to filtering


protein NFD4, putative, expressed                                  0.0005  10.1     0.0       0
  9539        R.T#DPAFLAM*FAITLATGACAVVGSIGSK.S        2708.31   3  0.0005  2.802  0.008   192.4  24  19/156

  1 of 2 peptide matches reported, 1 removed due to filtering


protein transposon protein, putative, unclassified                 0.0003  10.1     0.0       0
  8894        R.TLQIDLLLK.L                            1056.67   2  0.0003  2.040  0.019   442.4  1   10/16

  1 of 5 peptide matches reported, 4 removed due to filtering


protein retrotransposon, putative, centromere-specific             0.0001  10.1     0.0       0
  7536        K.ESAENLDIS#T#R.G                        1394.52   2  0.0001  2.033  0.134   67.4   64  8/40

  1 of 1 peptide matches reported, 0 removed due to filtering


protein MAR-binding filament-like protein 1, putative, express     0.0005  8.1      0.0       0
  9133        K.LDVTSNQLVS#TM*EAR.E                    1759.79   3  0.0005  2.565  0.014   209.7  35  16/84

  1 of 3 peptide matches reported, 2 removed due to filtering


protein histone acetyltransferase HAC5, putative, expressed        6e-005  8.1      0.0       0
  10619       K.SDKLRS#WY#QNLVK.K                      1796.81   2  6e-005  2.227  0.027   179.4  15  12/48

  1 of 1 peptide matches reported, 0 removed due to filtering


protein expressed protein                                          7e-005  8.1      0.0       0
  10843       K.MKNLDMYWY#SGTNQINK.K                   2185.94   2  7e-005  2.023  0.036   146.9  1   12/48

  1 of 3 peptide matches reported, 2 removed due to filtering


protein transposon protein, putative, unclassified                 0.0004  2.1      0.0       0
  11079       R.EPPT#IPTIPEITDISER.L                   1987.96   2  0.0004  2.083  0.102   193.0  9   12/48

  1 of 2 peptide matches reported, 1 removed due to filtering
```

SRF File: E:\data\KELVIN\09-07-21\KK09B01.srf
Database... indexed - rice_all.pep.fasta.hdr (7/23/2009)
Filter(s)... xc (± 1,2,3)=1.50,2.00,2.50 ; peptide probability<=1e-003 ; # different peptides>=2
Mods: (M* +15.99492) (STY# +79.96633)

Reference Scan(s)	Sequence	MH+	z	P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein retrotransposon protein, putative, Ty3-gypsy subclass				4e-005	18.1	0.0	0			
8430	K.YKLKVVAADDVGSS#SSQDGK.G	2134.01	3	4e-005	2.747	0.014	210.5	2	20/114	
8430	K.YKLKVVAADDVGS#SSSQDGK.G	2134.01	3	4e-005	2.709	0.103	190.1	10	19/114	

Reference: LOC_Os10g13790.1|13110.m01214|protein retrotransposon protein, pu
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 288 Monoisotopic MW: 31764.0 pI: 5.52

Protein:

MSSVKDKFMN SFKETHQGIV VQK**YKLKVVA ADDVGSSSSQ DGK**GAVDASG
DKGDNLQDGV IEDLGEDGAE VQEEPPRRMN FQDQVDYTVQ HALINQSGVL
VNTLTNMIKS VIDDTIDEHQ TTGPVYLPGV VFPNYKSLVT SNQQNTSNAP
QVQPTAAAST LAPVAPSTAQ KQTINPRLLS REQPQHAVQQ TPPREQAVQP
IQQQTVQPIQ QQTVQQMPTG LRTPNNRQTH QNVTQVIPEH MVHTVQPDQS
VAAQIIPEHL VRNIQPNFQN YQQGNLNYQY QLLTLKIL

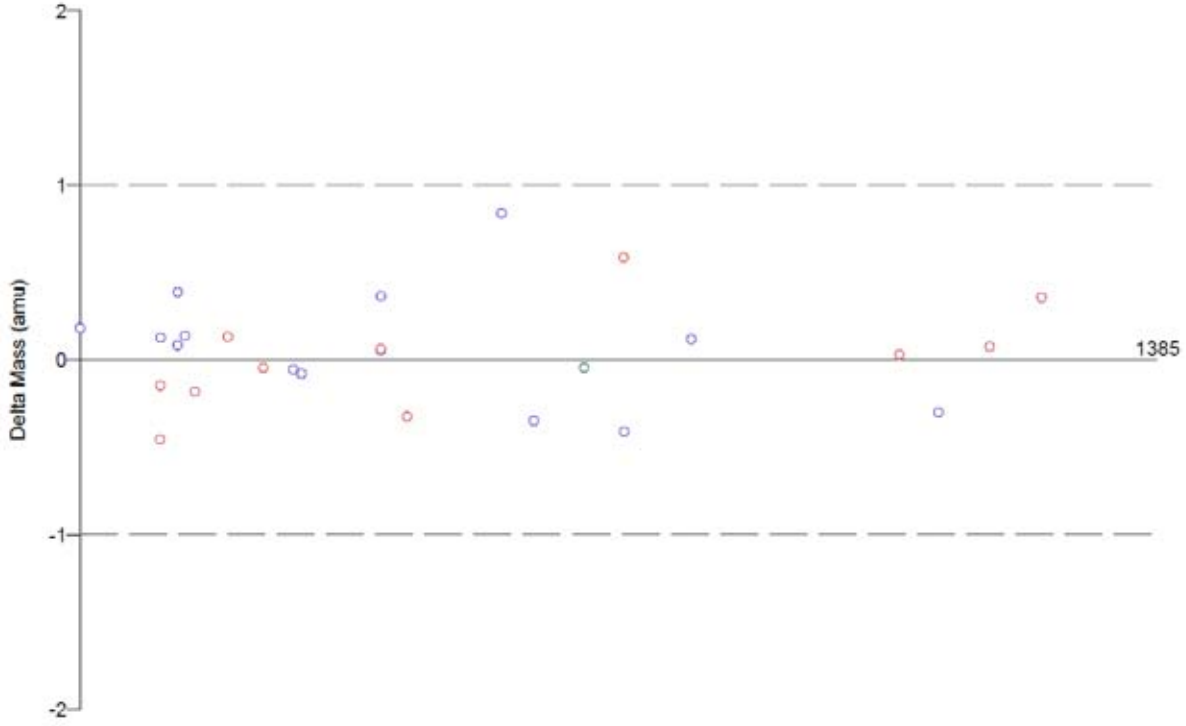
Protein Coverage:

Sequence		MH+	% Mass	AA	% AA
YKLKVVAADDVGSSSSQDGK		2054.04	6.47	24 - 43	6.94
	Totals:	2054.04	6.47	20	6.94

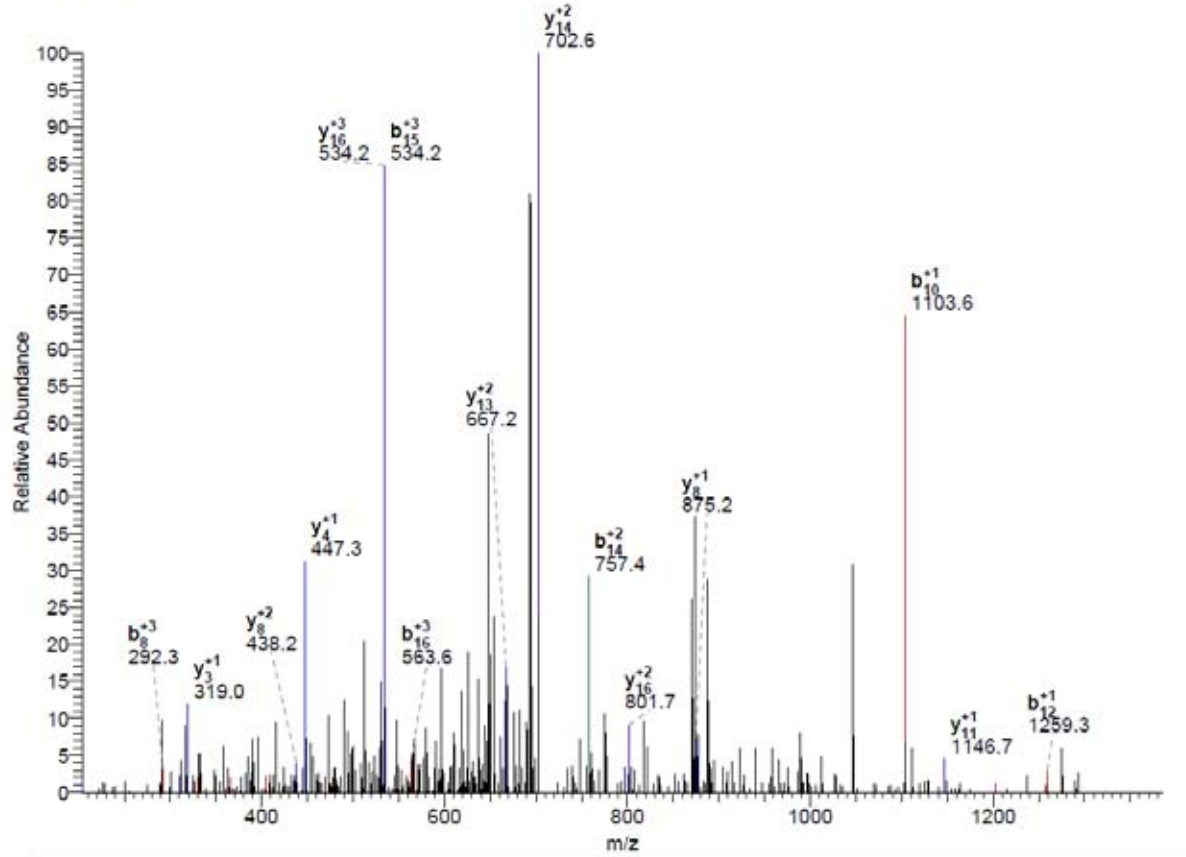
DTA for scan: 8430
Precursor ion: 712.11
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
Y		164.07							
K		**292.17**				1970.94			
L		**405.25**				1842.85			
K		533.34				1729.76			
V		632.41				1601.67			
V		731.48				1502.60			
A		802.52				1403.53			
A		873.56				1332.50			
D		988.58				1261.46			
D		**1103.61**				**1146.43**			
V		**1202.68**				1031.40			
G		**1259.70**				932.34			
S		1346.73				**875.31**			
S#		1513.73				788.28			
S		1600.76				621.28			
S		1687.79				**534.25**			
Q		1815.85				**447.22**			
D		1930.88				**319.16**			
G		1987.90				**204.13**			
K						147.11			



#8430-8430 RT:47.69-47.69 NL: 1.58E2

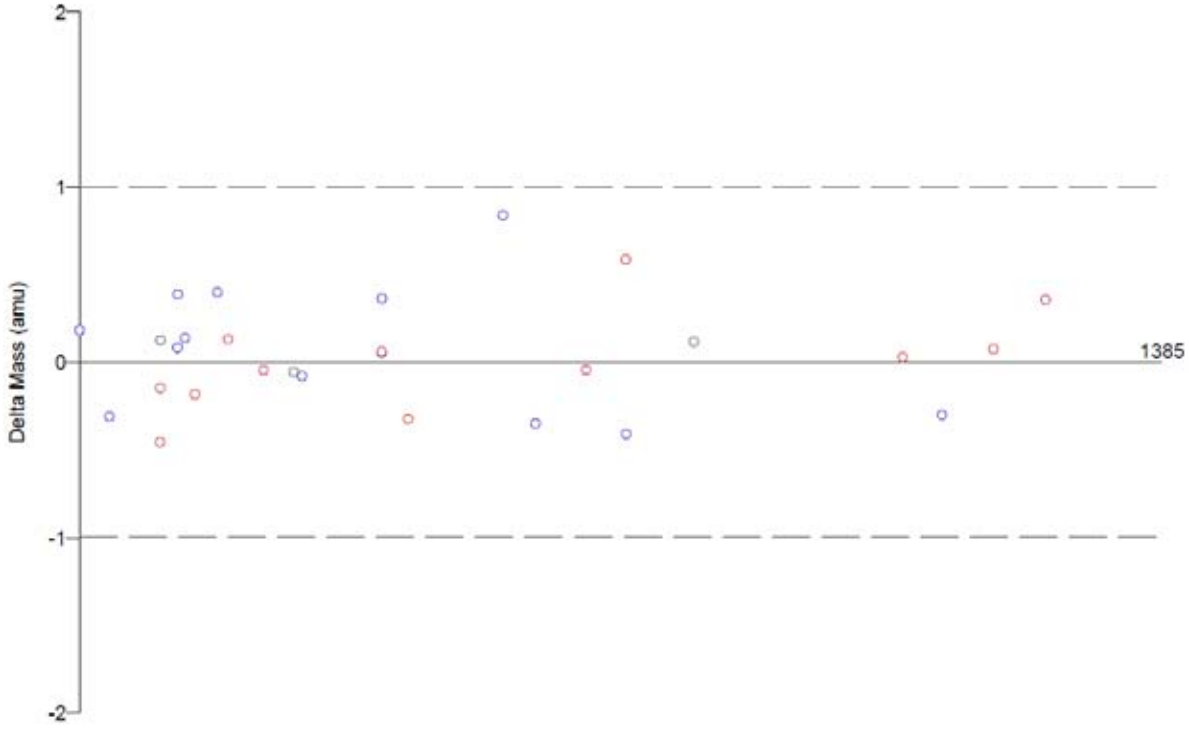


BioworksBrowser rev. 3.3 25/08/2009 13:00:49

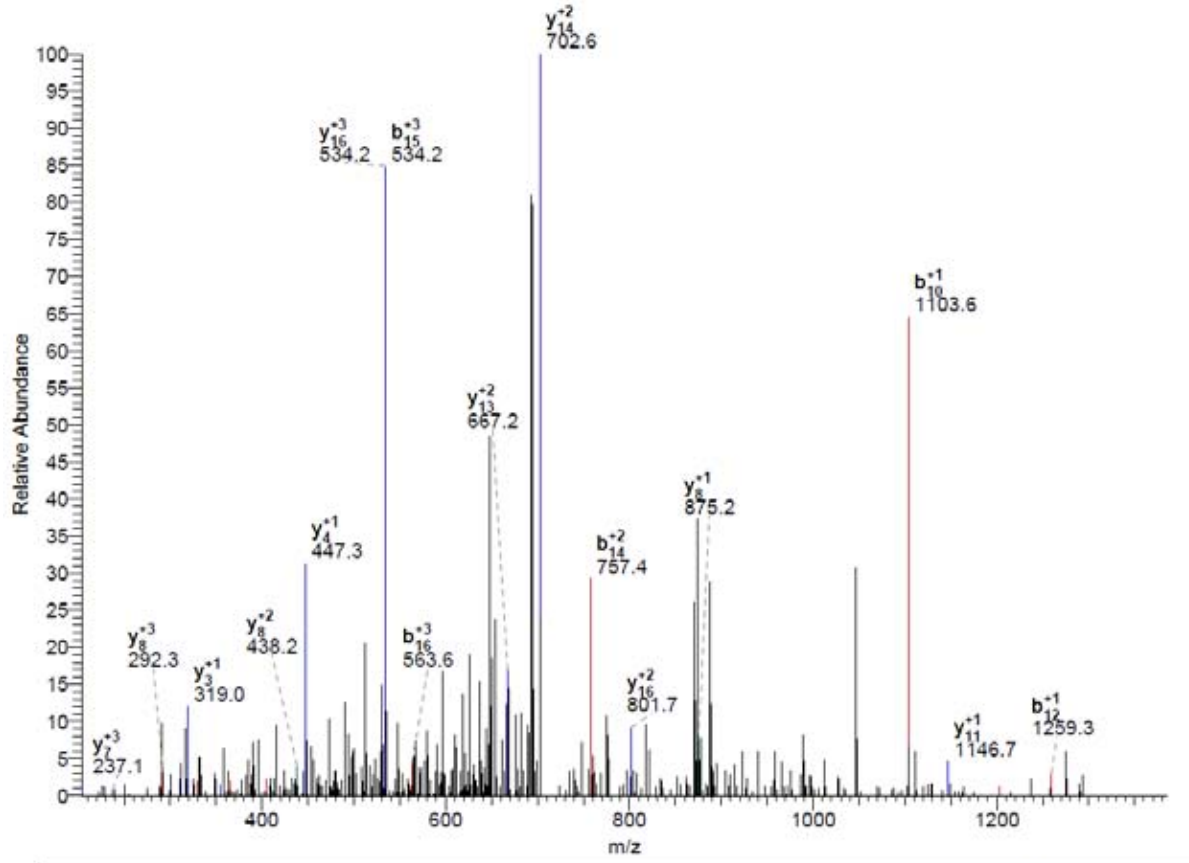
DTA for scan: 8430
Precursor ion: 712.11
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

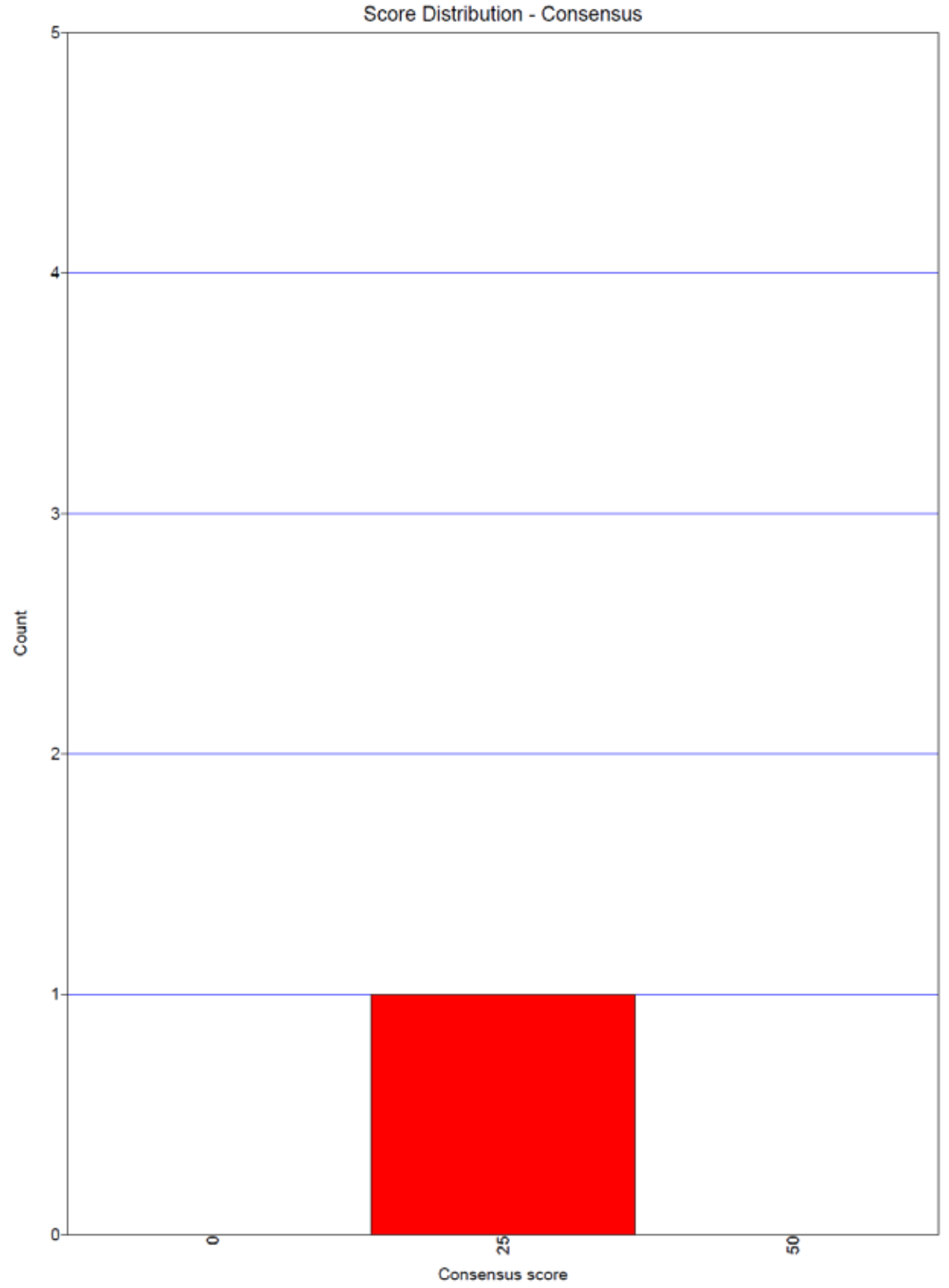
AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
Y		164.07							
K		292.17				1970.94			
L		405.25				1842.85			
K		533.34				1729.76			
V		632.41				1601.67			
V		731.48				1502.60			
A		802.52				1403.53			
A		873.56				1332.50			
D		988.58				1261.46			
D		1103.61				1146.43			
V		1202.68				1031.40			
G		1259.70				932.34			
S#		1426.70				875.31			
S		1513.73				708.32			
S		1600.76				621.28			
S		1687.79				534.25			
Q		1815.85				447.22			
D		1930.88				319.16			
G		1987.90				204.13			
K						147.11			



#8430-8430 RT:47.69-47.69 NL: 1.58E2



BioworksBrowser rev. 3.3 25/08/2009 13:00:49


Score Distribution - Consensus
5
4
3
2
1
0
Count
0
25
50
Consensus score

```
[SEQUEST]
first_database_name = C:\Xcalibur\database\rice_all.pep.fasta.hdr
second_database_name =
peptide_mass_tolerance = 1.0000
peptide_mass_units = 0              ; 0=amu, 1=mmu, 2=ppm
ion_series = 0 1 1 0.0 1.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0
fragment_ion_tolerance = 0.5000          ; for trap data leave at 1.0, for accurate mass data use values < 1.0
fragment_ion_units = 0              ; 0=amu, 1=mmu
num_output_lines = 10                ; # peptide results to show
num_results = 250                    ; # results to store
num_description_lines = 5           ; # full protein descriptions to show for top N peptides
show_fragment_ions = 0              ; 0=no, 1=yes
print_duplicate_references = 0      ; number of duplicate references reported
enzyme_info = Trypsin(KR) 1 1 KR -
max_num_differential_per_peptide = 3          ; max # of diff. mod in a peptide
diff_search_options = 15.994915 M 79.966331 STY 0.000000 M 0.000000 X 0.000000 T 0.000000 Y
term_diff_search_options = 0.000000 0.000000
nucleotide_reading_frame = 0        ; 0=protein db, 1-6, 7 = forward three, 8-reverse three, 9=all six
mass_type_parent = 1                ; 0=average masses, 1=monoisotopic masses
mass_type_fragment = 1              ; 0=average masses, 1=monoisotopic masses
normalize_xcorr = 0                 ; use normalized xcorr values in the out file
remove_precursor_peak = 0           ; 0=no, 1=yes
ion_cutoff_percentage = 0.0000          ; prelim. score cutoff % as a decimal number i.e. 0.30 for 30%
max_num_internal_cleavage_sites = 2         ; maximum value is 12
protein_mass_filter = 0 0           ; enter protein mass min & max value ( 0 for both = unused)
match_peak_count = 0                ; number of auto-detected peaks to try matching (max 5)
match_peak_allowed_error = 1        ;  number of allowed errors in matching auto-detected peaks
match_peak_tolerance = 1.0000           ; mass tolerance for matching auto-detected peaks
partial_sequence =
sequence_header_filter =
digest_mass_range = 600.0000 3500.0000

add_Cterm_peptide = 0.0000              ; added to each peptide C-terminus
add_Cterm_protein = 0.0000              ; added to each protein C-terminus
add_Nterm_peptide = 0.0000              ; added to each peptide N-terminus
add_Nterm_protein = 0.0000              ; added to each protein N-terminus
add_G_Glycine = 0.0000                  ; added to G
add_A_Alanine = 0.0000                  ; added to A
add_S_Serine = 0.0000                   ; added to S
add_P_Proline = 0.0000                  ; added to P
add_V_Valine = 0.0000                   ; added to V
add_T_Threonine = 0.0000                ; added to T
add_C_Cysteine = 0.0000                 ; added to C
add_L_Leucine = 0.0000                  ; added to L
add_I_Isoleucine = 0.0000               ; added to I
add_X_LorI = 0.0000                     ; added to X
add_N_Asparagine = 0.0000               ; added to N
add_O_Ornithine = 0.0000                ; added to O
add_B_avg_NandD = 0.0000                ; added to B
add_D_Aspartic_Acid = 0.0000            ; added to D
add_Q_Glutamine = 0.0000                ; added to Q
add_K_Lysine = 0.0000                   ; added to K
add_Z_avg_QandE = 0.0000                ; added to Z
add_E_Glutamic_Acid = 0.0000            ; added to E
add_M_Methionine = 0.0000               ; added to M
add_H_Histidine = 0.0000                ; added to H
add_F_Phenylalanine = 0.0000            ; added to F
add_R_Arginine = 0.0000                 ; added to R
add_Y_Tyrosine = 0.0000                 ; added to Y
add_W_Tryptophan = 0.0000               ; added to W
add_J_user_amino_acid = 0.0000          ; added to J
add_U_user_amino_acid = 0.0000          ; added to U
```

Output files for KK02 protein spot sequencing

KK09B02.RAW Page 1

SRF File: E:\data\KELVIN\09-07-21\KK09B02.srf
Database... indexed - rice_all.pep.fasta.hdr (7/23/2009)
Filter(s)... xc (± 1,2,3)=1.50,2.00,2.50 ; peptide probability<=1e-003
Mods: (M* +15.99492) (STY# +79.96633)

Reference Scan(s)	Sequence	MH+	z	P P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions
protein pectate lyase precursor, putative, expressed				0.0003	10.1	0.0	0		
9845	R.LT#QELLVASDKTIDGR.G	1838.93	2	0.0003	2.542	0.029	230.7	5	11/45
1 of 3 peptide matches reported, 2 removed due to filtering									
protein retrotransposon protein, putative, unclassified				2e-006	8.1	0.0	0		
8236	K.M*NEIVAGNEMQRLGVEKK.M	2062.04	2	2e-006	2.187	0.053	270.0	2	15/34
1 of 1 peptide matches reported, 0 removed due to filtering									

SRF File: E:\data\KELVIN\09-07-21\KK09B02.srf
Database... indexed - rice_all.pep.fasta.hdr (7/23/2009)
Filter(s)... xc (± 1,2,3)=1.50,2.00,2.50 ; peptide probability<=1e-003
Mods: (M* +15.99492) (STY# +79.96633)

Reference Scan(s)	Sequence	MH+	z	P P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein pectate lyase precursor, putative, expressed				0.0003	10.1	0.0	0			
9845	R.LT#QELLVASDKTIDGR.G	1838.93	2	0.0003	2.542	0.029	230.7	5	11/45	

1 of 3 peptide matches reported, 2 removed due to filtering

Reference: LOC_Os02g12300.1|13102.m01331|protein pectate lyase precursor, pu
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 446 Monoisotopic MW: 49196.6 pI: 9.68

Protein:

MDDQRLQWRK PGSFLLVAGV FLAAAAAVSN AGIGEFDEHW EKRRAAAEAA
AEEVYKPDPF NVTNEFNHAV IRSTERGVLR RELSGKNSKY KGPCLATNPI
DRCWRCRKDW ATDRKRLARC AMGFGRGATG GVRGKIYVVT DPGDGDAANP
RYGTLRWGAM QAAPLWITFA KSMVIRLTQR LLVASDKTID GRGAQVHIAR
GGAGITVQFA RNVIITSLHV HDVKHSDGGA VRDSPTHIGP RTRADGDGIS
LFAATDVWVD HVSMSMCEDG LIDVVQGSTG VTISNSHFTN HNDVMLFGAS
DSYPQDKVMQ ITVAFNHFGR GLVQRMPRCR WGFFHVVNND YTHWLMYAIG
GGMSPTILSQ GNRYIAPPNI AAKLITRHYA PEWEWKNWAW RSDGDLFMNG
AYFQASNGAI NRKVKGSDMV KPKPGSYVRR LTRFAGALSC RPGEPC

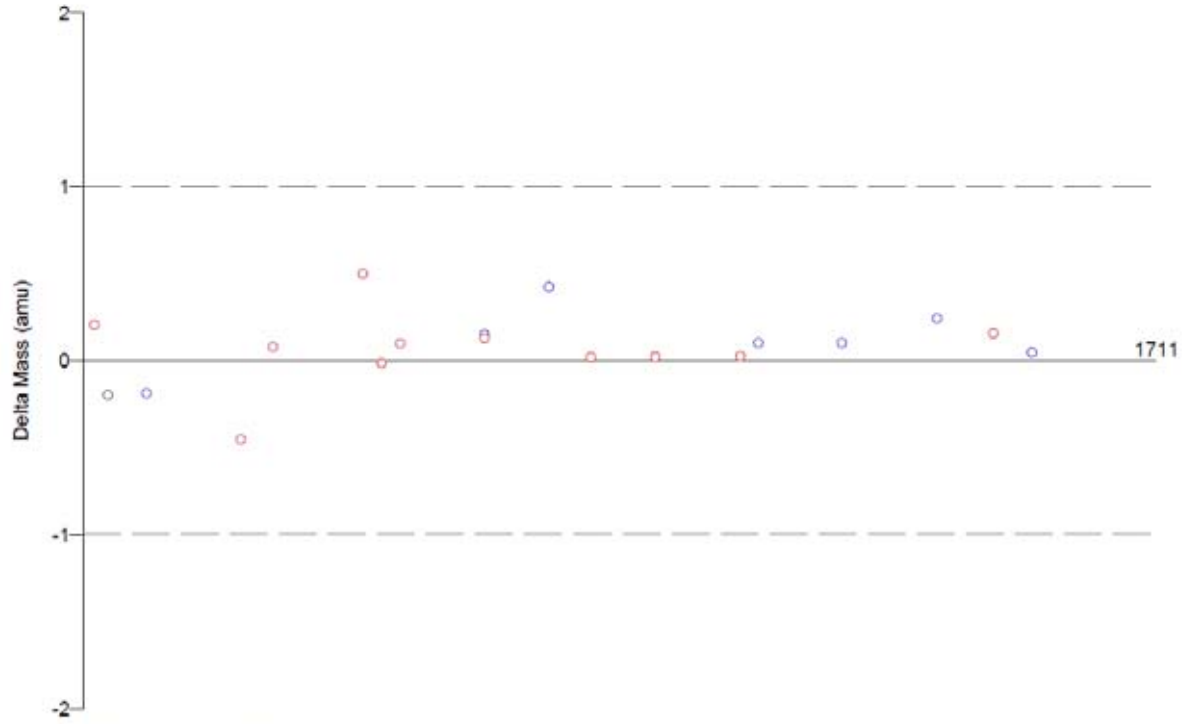
Protein Coverage:

Sequence		MH+	% Mass	AA	% AA
LTQELLVASDKTIDGR		1758.96	3.58	177 - 192	3.59
VMQITVAFNHFGR		1519.78	3.09	308 - 320	2.91
	Totals:	1758.96	3.58	16	3.59

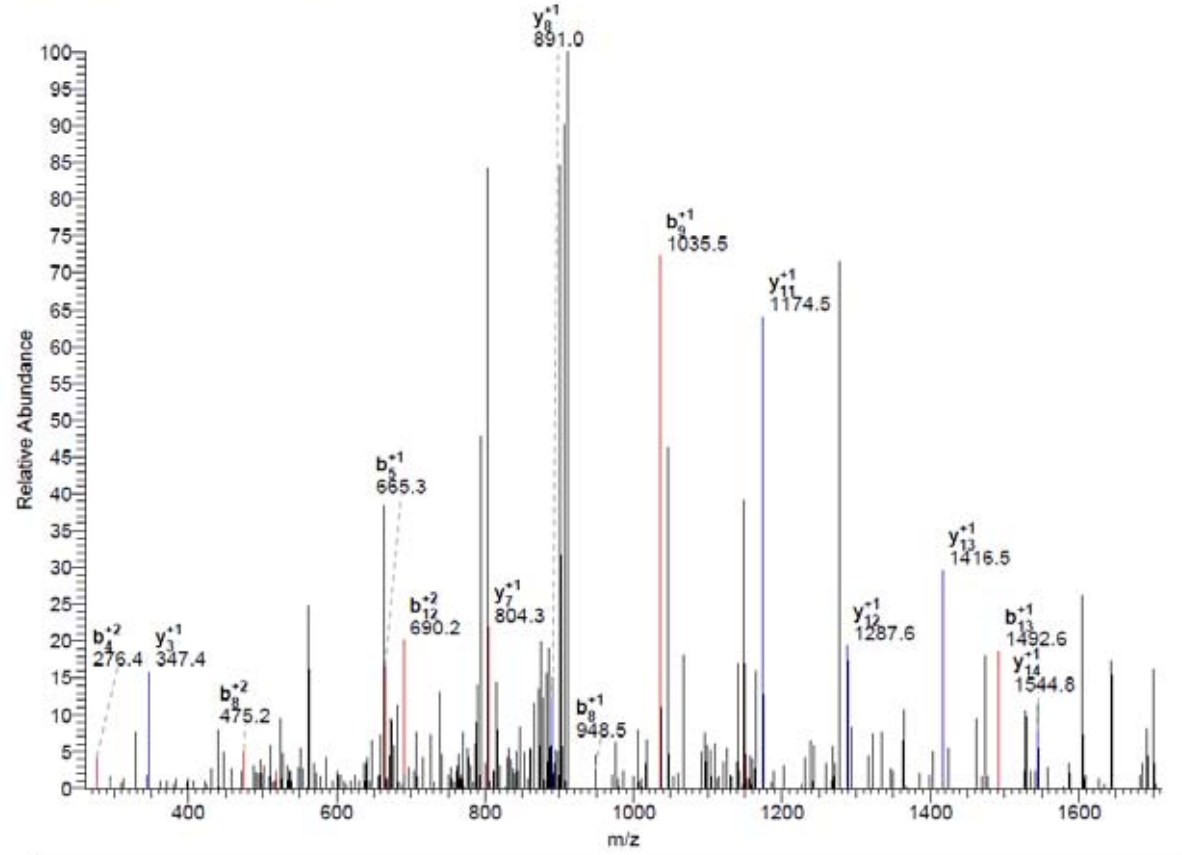
DTA for scan: 9845
Precursor ion: 919.81
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
L		114.09							
T#		**295.11**				1725.84			
Q		423.16				**1544.83**			
E		552.21				**1416.77**			
L		**665.29**				**1287.73**			
L		778.37				**1174.64**			
V		877.44				1061.56			
A		**948.48**				962.49			
S		**1035.51**				**891.45**			
D		**1150.54**				**804.42**			
K		1278.63				689.39			
T		1379.68				561.30			
I		**1492.77**				460.25			
D		1607.79				**347.17**			
G		1664.81				232.14			
R						175.12			



#9845-9845 RT:55.33-55.33 NL: 8.61E1



Reference Scan(s)	Sequence	MH+	z	P P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein retrotransposon protein, putative, unclassified				2e-006	8.1	0.0	0			
8236	K.M*NEIVAGNEMQRLGVEKK.M	2062.04	2	2e-006	2.187	0.053	270.0	2	15/34	

Reference: LOC_Os01g19300.1|13101.m02099|protein retrotransposon protein, pu
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 81 Monoisotopic MW: 9367.6 pI: 4.78

Protein:

MMTECRRLFE NTSREVGAQM DRLTARAISS DGRVAQLEME LEVARDDLQK
MNEIVAGNEM QRLGVEKKMN DLQDHIPSIR D

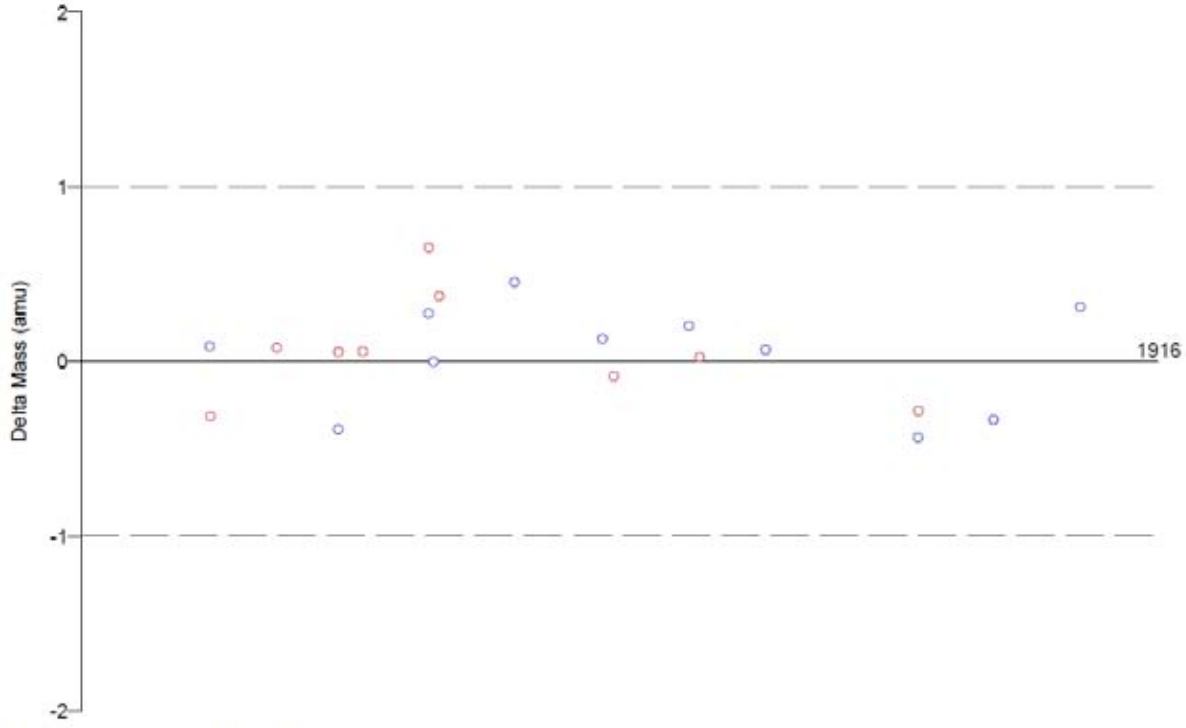
Protein Coverage:

Sequence		MH+	% Mass	AA	% AA
MNEIVAGNEMQRLGVEKK		2046.05	21.84	51 - 68	22.22
	Totals:	2046.05	21.84	18	22.22

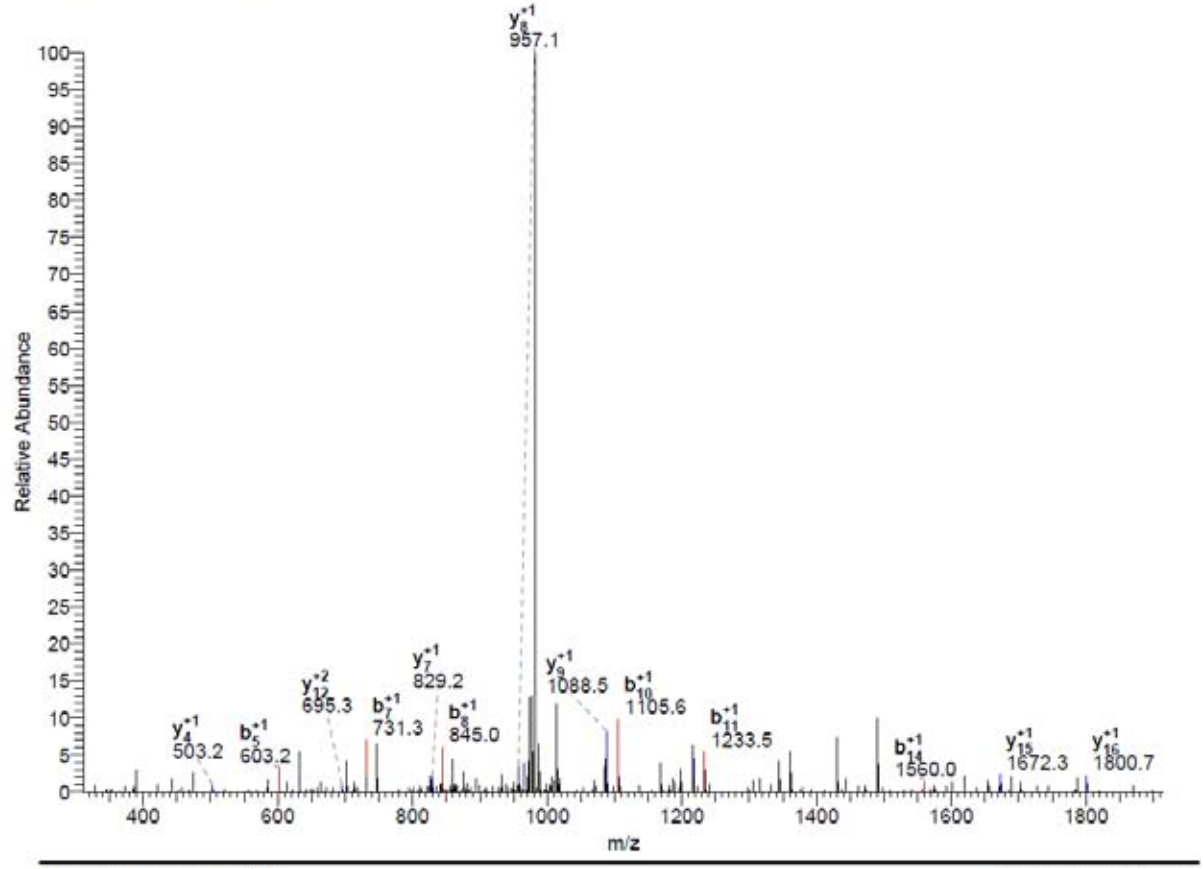
DTA for scan: 8236
Precursor ion: 1031.68
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

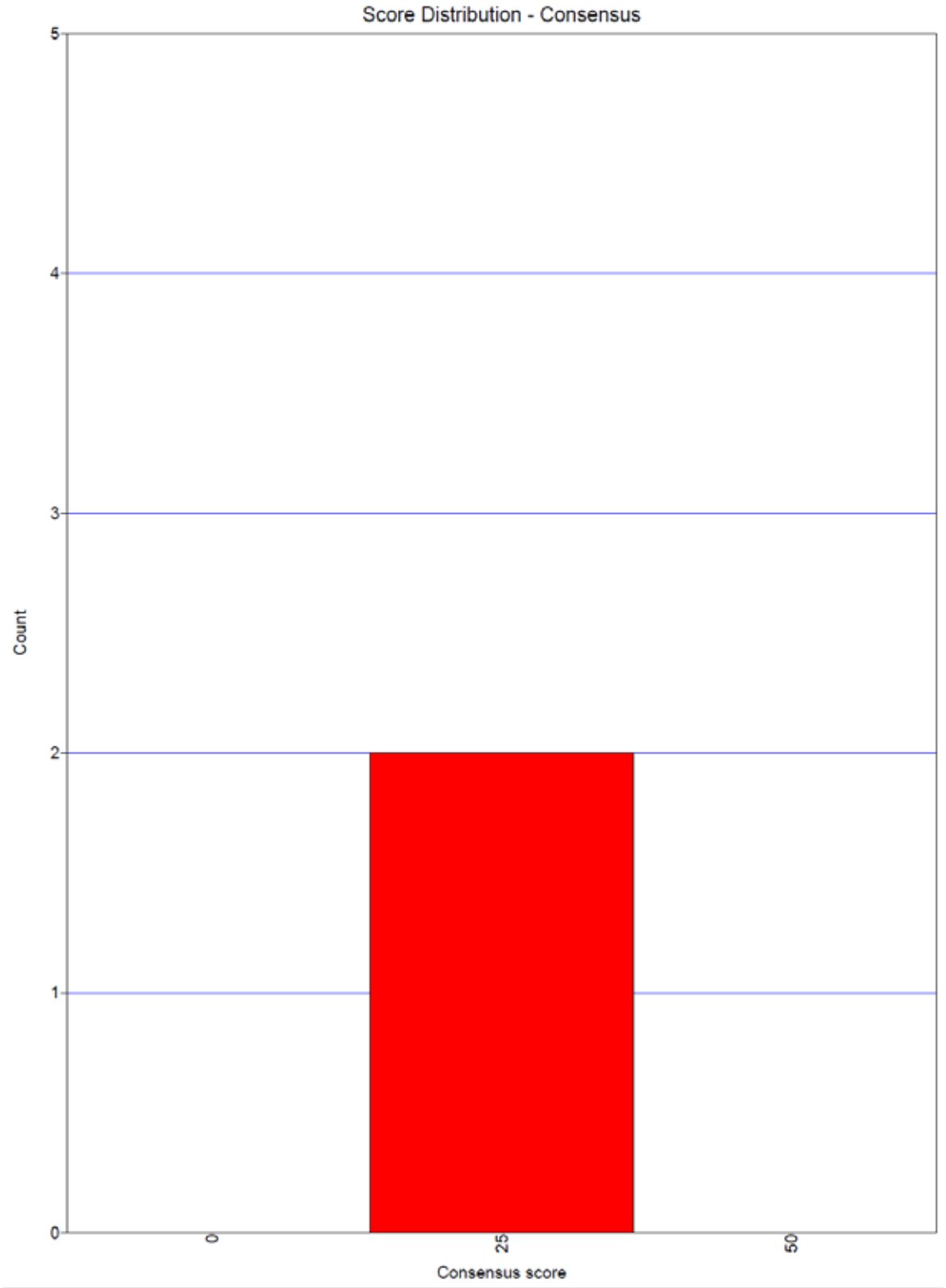
Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
M*		148.04							
N		262.09				1915.01			
E		391.13				1800.96			
I		504.21				1671.92			
V		603.28				1558.84			
A		674.32				1459.77			
G		731.34				1388.73			
N		845.38				1331.71			
E		974.42				1217.67			
M		1105.47				1088.62			
Q		1233.52				957.58			
R		1389.62				829.53			
L		1502.71				673.42			
G		1559.73				560.34			
V		1658.80				503.32			
E		1787.84				404.25			
K		1915.94				275.21			
K						147.11			



#8236-8236 RT:46.28-46.28 NL: 5.33E2




Score Distribution - Consensus
5
4
3
2
1
0
Count
0
25
50
Consensus score

```
[SEQUEST]
first_database_name = C:\Xcalibur\database\rice_all.pep.fasta.hdr
second_database_name =
peptide_mass_tolerance = 1.0000
peptide_mass_units = 0                  ; 0=amu, 1=mmu, 2=ppm
ion_series = 0 1 1 0.0 1.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0
fragment_ion_tolerance = 0.5000           ; for trap data leave at 1.0, for accurate mass data use values < 1.0
fragment_ion_units = 0                  ; 0=amu, 1=mmu
num_output_lines = 10                    ; # peptide results to show
num_results = 250                         ; # results to store
num_description_lines = 5               ; # full protein descriptions to show for top N peptides
show_fragment_ions = 0                  ; 0=no, 1=yes
print_duplicate_references = 0          ; number of duplicate references reported
enzyme_info = Trypsin(KR) 1 1 KR -
max_num_differential_per_peptide = 3            ; max # of diff. mod in a peptide
diff_search_options = 15.994915 M 79.966331 STY 0.000000 M 0.000000 X 0.000000 T 0.000000 Y
term_diff_search_options = 0.000000 0.000000
nucleotide_reading_frame = 0            ; 0=protein db, 1-6, 7 = forward three, 8-reverse three, 9=all six
mass_type_parent = 1                    ; 0=average masses, 1=monoisotopic masses
mass_type_fragment = 1                  ; 0=average masses, 1=monoisotopic masses
normalize_xcorr = 0                     ; use normalized xcorr values in the out file
remove_precursor_peak = 0               ; 0=no, 1=yes
ion_cutoff_percentage = 0.0000            ; prelim. score cutoff % as a decimal number i.e. 0.30 for 30%
max_num_internal_cleavage_sites = 2             ; maximum value is 12
protein_mass_filter = 0 0               ; enter protein mass min & max value ( 0 for both = unused)
match_peak_count = 0                    ; number of auto-detected peaks to try matching (max 5)
match_peak_allowed_error = 1            ;  number of allowed errors in matching auto-detected peaks
match_peak_tolerance = 1.0000             ; mass tolerance for matching auto-detected peaks
partial_sequence =
sequence_header_filter =
digest_mass_range = 600.0000 3500.0000

add_Cterm_peptide = 0.0000              ; added to each peptide C-terminus
add_Cterm_protein = 0.0000              ; added to each protein C-terminus
add_Nterm_peptide = 0.0000              ; added to each peptide N-terminus
add_Nterm_protein = 0.0000              ; added to each protein N-terminus
add_G_Glycine = 0.0000                  ; added to G
add_A_Alanine = 0.0000                  ; added to A
add_S_Serine = 0.0000                   ; added to S
add_P_Proline = 0.0000                  ; added to P
add_V_Valine = 0.0000                   ; added to V
add_T_Threonine = 0.0000                ; added to T
add_C_Cysteine = 0.0000                 ; added to C
add_L_Leucine = 0.0000                  ; added to L
add_I_Isoleucine = 0.0000               ; added to I
add_X_LorI = 0.0000                     ; added to X
add_N_Asparagine = 0.0000               ; added to N
add_O_Ornithine = 0.0000                ; added to O
add_B_avg_NandD = 0.0000                ; added to B
add_D_Aspartic_Acid = 0.0000            ; added to D
add_Q_Glutamine = 0.0000                ; added to Q
add_K_Lysine = 0.0000                   ; added to K
add_Z_avg_QandE = 0.0000                ; added to Z
add_E_Glutamic_Acid = 0.0000            ; added to E
add_M_Methionine = 0.0000               ; added to M
add_H_Histidine = 0.0000                ; added to H
add_F_Phenylalanine = 0.0000            ; added to F
add_R_Arginine = 0.0000                 ; added to R
add_Y_Tyrosine = 0.0000                 ; added to Y
add_W_Tryptophan = 0.0000               ; added to W
add_J_user_amino_acid = 0.0000          ; added to J
add_U_user_amino_acid = 0.0000          ; added to U
```

Output files for KK03 protein spot sequencing

KK09B03.RAW Page

```
SRF File: E:\data\KELVIN\09-07-21\KK09B03.srf
Database... indexed - rice_all.pep.fasta.hdr (7/23/2009)
Filter(s)... xc (± 1,2,3)=1.50,2.00,2.50 ; peptide probability<=1e-003
Mods: (M* +15.99492) (STY# +79.96633)

Reference                                                                  P      Score   Coverage   Accession
  Scan(s)       Sequence                                        MH+     z    P        XC   DeltaCn   Sp   RSp  Ions

protein SLT1 protein, putative, expressed                                  0.0009 10.2    0.0       0
  10600       R.TQQDVEMHS#VVQPSWQHEFSGVM*KKAR.G               3265.49   3  0.0009 3.674  0.144     269.0 5    25/156

  1 of 3 peptide matches reported, 2 removed due to filtering


protein pollen-specific protein SF21, putative, expressed                  0.0009 10.1    0.0       0
  10286       R.QS#MNVWRFIHT#INERHDLT#ESLK.E                  3094.34   3  0.0009 2.822  0.037     125.5 18   24/220

  1 of 1 peptide matches reported, 0 removed due to filtering


protein retrotransposon protein, putative, unclassified                    0.0008 10.1    0.0       0
  11114       R.CQVQDITKAS#VINAAS#AGLLEGELT#R.K               2927.30   3  0.0008 2.822  0.039     102.6 103  22/250

  1 of 4 peptide matches reported, 3 removed due to filtering


protein peptide-N4-asparagine amidase A, putative                          0.0004 10.1    0.0       0
  12931       R.LSPADIAVPDAAAADTPTT#YPEVDRPLR.P               3052.47   3  0.0004 2.634  0.110     173.8 4    21/162

  1 of 3 peptide matches reported, 2 removed due to filtering


protein retrotransposon protein, putative, Ty3-gypsy subclass              0.0002 10.1    0.0       0
  13239       R.QSLAWTCFVET#PRVEPVPVM*VTTRR.N                 2998.47   3  0.0002 2.583  0.002     133.3 150  18/144

  1 of 1 peptide matches reported, 0 removed due to filtering


protein integral membrane protein, putative, expressed                     0.0003 10.1    0.0       0
  9347        K.ALPSPAQAAAASAM*AFSVGAVVPLLAAGFIVNYR           3317.77   3  0.0003 2.570  0.018     140.6 20   16/132

  1 of 1 peptide matches reported, 0 removed due to filtering


protein retrotransposon protein, putative, unclassified                    0.0006 10.1    0.0       0
  11248       K.NKLPHMHILYGENVK.N                             1906.04   2  0.0006 2.225  0.039     241.9 1    10/30

  1 of 2 peptide matches reported, 1 removed due to filtering


protein retrotransposon protein, putative, unclassified                    3e-005 8.1     0.0       0
  15687       K.GEAT#GGKQGDAGAPS#GERS#ADK.W                   2285.85   3  3e-005 2.936  0.068     122.6 3    31/210

  1 of 1 peptide matches reported, 0 removed due to filtering


protein hypothetical protein                                               4e-005 8.1     0.0       0
  11370       K.LELRAT#ANKGSVS#K.D                            1633.77   2  4e-005 2.092  0.050     126.3 26   10/52

  1 of 1 peptide matches reported, 0 removed due to filtering


protein pentatricopeptide, putative, expressed                             0.0005 2.1     0.0       0
  9092        R.RAHGVAVRS#GLGPELAVGNALVHM*Y#GK.C              2985.43   3  0.0005 2.625  0.026     152.8 13   29/208

  1 of 2 peptide matches reported, 1 removed due to filtering
```

SRF File: E:\data\KELVIN\09-07-21\KK09B03.srf
Database... indexed - rice_all.pep.fasta.hdr (7/23/2009)
Filter(s)... xc (± 1,2,3)=1.50,2.00,2.50 ; peptide probability<=1e-003
Mods: (M* +15.99492) (STY# +79.96633)

Reference Scan(s)	Sequence	MH+	z	P P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein SLT1 protein, putative, expressed				0.0009	10.2	0.0	0			
10600	R.TQQDVEMHS#VVQPSWQHEPSGVM*KKAR.G	3265.49	3	0.0009	3.674	0.144	269.0	5	25/156	

1 of 3 peptide matches reported, 2 removed due to filtering

Reference: LOC_Os01g62300.1|13101.m06666|protein SLT1 protein, putative, exp
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 521 Monoisotopic MW: 58263.9 pI: 6.47

Protein:

MGEDLVTTLS MENGGGGGGG GNHSLCTLLS MDPAGHLDDR AVGVMVQPRI
GDGAGRAHAV SLSGAHPPDI NQPWQTDLCD MLDVSLGPQI QVYDARALAV
LSSAPKAGNR KAAKRGDSIW GAWPFPTFYF KPLLADKCKG KVTRDANGVS
GFDKTDLRLD MFMVQHDMEN MYMWVFKERP ENALGKMQLR SYMNGHSRPG
EPQFPFSVDR GFVRSHRMQR KHYRGLSNPQ CIHGIEVVRS PNLASITEVD
RRRWVELTGR ELNFAIPQEA CDFGTWRTMP NTEIELDRPH PVMKGNVHQH
PKKLLNGSGL NLSSPSNHSG EDGMDLSPVS SKRRKEVPPH AMDEECFLPL
NSCGERTQQD VEMHSVVQPS WQHEPSGVMK KARGPVTAAK SIYEDDQGYL
IMVSLPPVDQ QKVKVSWRNS LTHGIVKILC VSTAQTPYIR RHDRVPKLAD
PMPEHCPHGE FVREIPLATR IPEDAKLEAY FDEAAAVLEI MVPKRGNEPE
EHEVRVSMRP PHLAANDLLL T

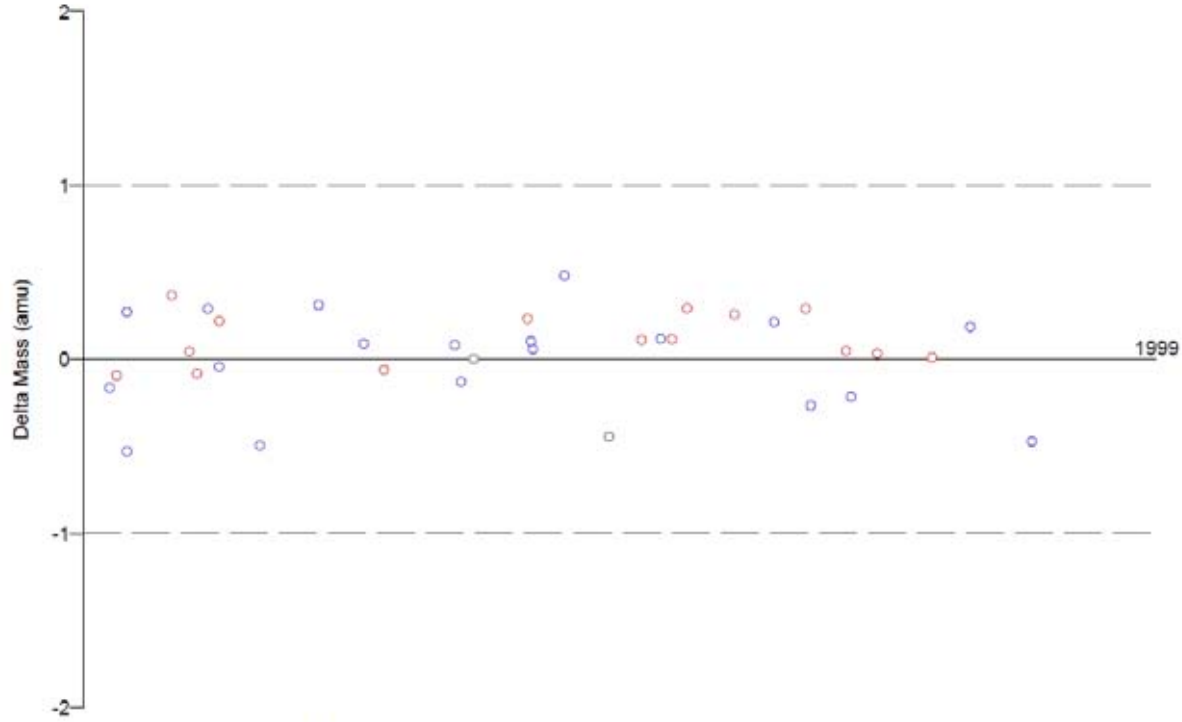
Protein Coverage:

Sequence	MH+	% Mass	AA	% AA
TQQDVEMHSVVQPSWQHEPSGVMKKAR	3169.53	5.44	357 - 383	5.18
VSMRPPHLAANDLLLT	1747.95	3.00	506 - 521	3.07
Totals:	3169.53	5.44	27	5.18

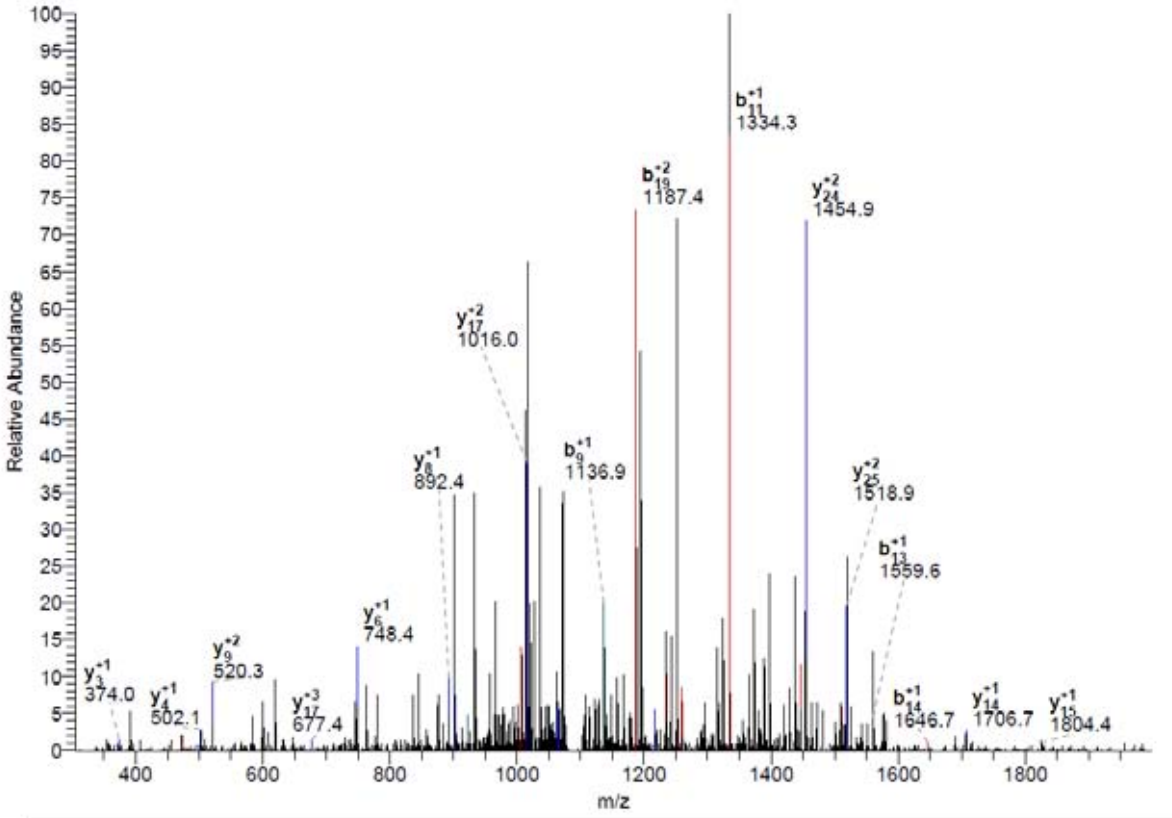
DTA for scan: 10600
Precursor ion: 1089.44
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
T		102.05							
Q		230.11				3164.44			
Q		**358.17**				3036.39			
D		**473.20**				2908.33			
V		572.27				2793.30			
E		701.31				2694.23			
M		832.35				2565.19			
H		969.41				2434.15			
S#		**1136.41**				2297.09			
V		**1235.48**				2130.09			
V		**1334.54**				2031.02			
Q		1462.60				1931.95			
P		**1559.66**				**1803.90**			
S		**1646.69**				**1706.84**			
W		1832.77				1619.81			
Q		1960.83				1433.73			
H		2097.88				1305.67			
E		2226.93				1168.61			
F		2374.00				1039.57			
S		2461.03				**892.50**			
G		2518.05				805.47			
V		2617.12				**748.45**			
M*		2764.15				649.38			
K		2892.25				**502.35**			
K		3020.34				**374.25**			
A		3091.38				246.16			
R						175.12			



#10600-10600 RT:59.72-59.72 NL: 1.91E3



BioworksBrowser rev. 3.3 25/08/2009 13:04:45

Reference Scan(s)	Sequence	MH+	z	P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein pollen-specific protein SP21, putative, expressed				0.0009	10.1	0.0	0			
10286	R.QS#MNVWRPIHT#INERHDLT#ESLK.E	3094.34	3	0.0009	2.822	0.037	125.5	18	24/220	

1 of 1 peptide matches reported, 0 removed due to filtering

Reference: LOC_Os06g36740.1|13106.m03801|protein pollen-specific protein SP2
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 348 Monoisotopic MW: 38547.1 pI: 5.83

Protein:

MGDSGGSVVS VDVERISPGG KEHHIQTNHG SVSVAVYGDH DKPALVTYPD
IALNHMSCFQ GLLFCPEASS LLLHNFCIYH ISPPGHELGA APVSPSSPVA
SVDELADQVS DVLDFPGLGP VMCLGVTAGA YILTLFATKY RERVLGLILV
SPLCRTPSWT EWPHNKVMSN LLYYYGMCNM VKDCLLQRYF SKGVQGCSAV
PESDIVQASR SFLDQRQSMN VWRFIHTINE RHDLTESLKE LQCRTLIPVG
QNSQFHAEAV HMTSKLDERY SALVEVQGCG SVVTEEQPHA MLMPLEYFLM
GYGLYRPSQI SCSPRSPLNP FCISPELLSP ESMGVKLKPI KTRANLEV

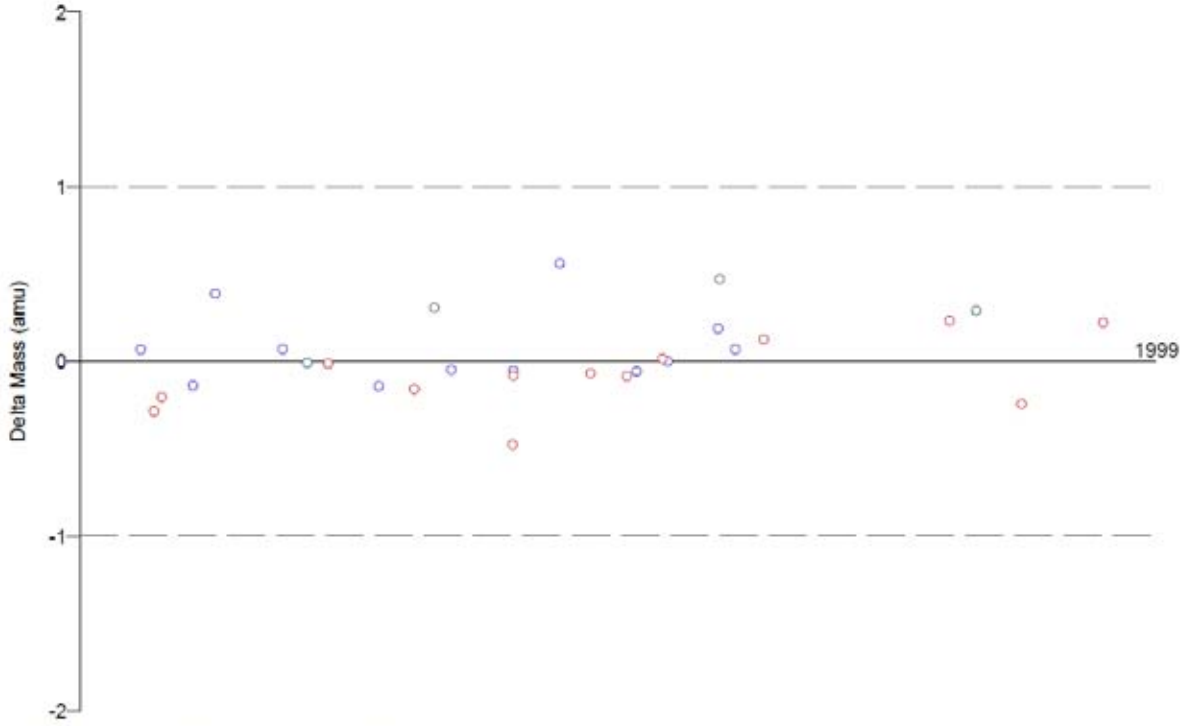
Protein Coverage:

Sequence	MH+	% Mass	AA	% AA
QSMNVWRPIHTINERHDLTESLK	2854.44	7.41	217 - 239	6.61
Totals:	2854.44	7.41	23	6.61

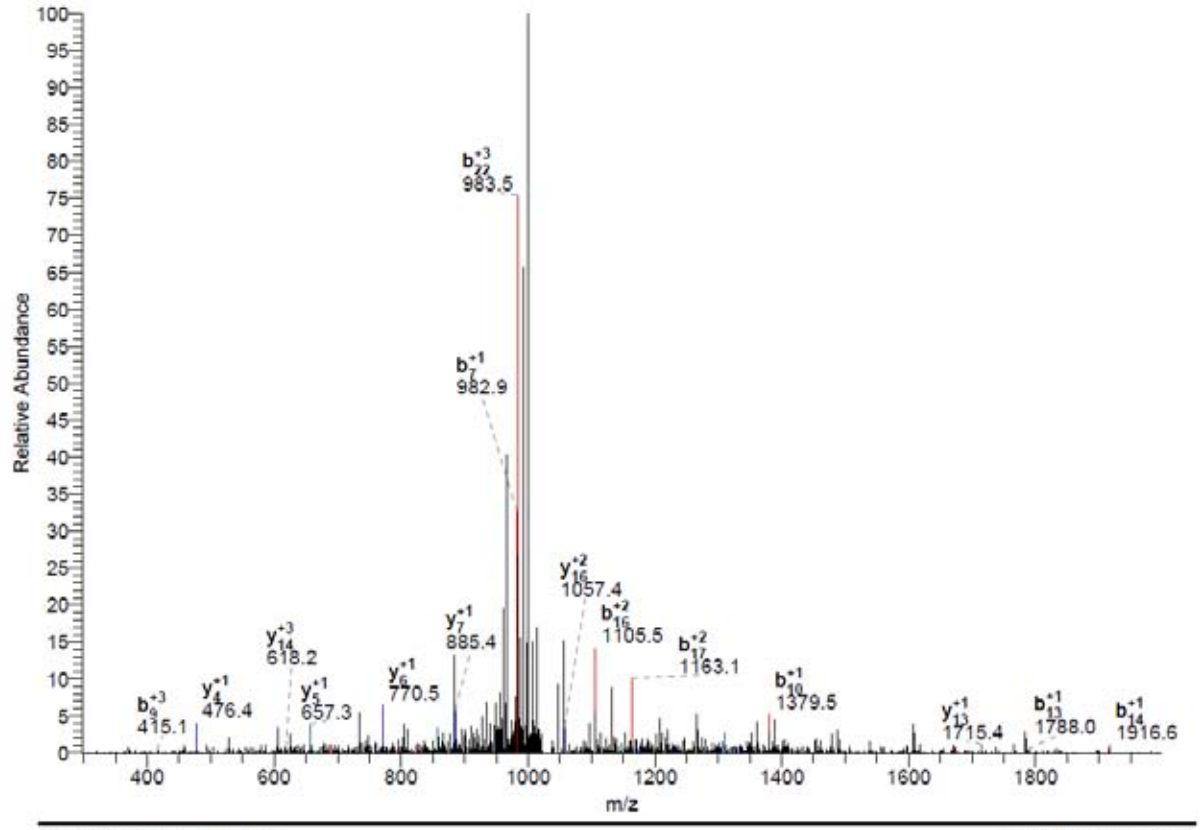
DTA for scan: 10286
Precursor ion: 1032.44
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
Q		129.07							
S#		296.06				2966.28			
M		**427.10**				2799.28			
N		541.15				2668.24			
V		640.22				2554.20			
W		**826.30**				2455.13			
R		**982.40**				2269.05			
F		1129.46				2112.95			
I		1242.55				1965.88			
H		**1379.61**				1852.80			
T#		1560.62				**1715.74**			
I		**1673.71**				1534.73			
N		**1787.75**				1421.64			
E		**1916.79**				**1307.60**			
R		2072.89				**1178.56**			
H		2209.95				1022.46			
D		2324.98				**885.40**			
L		2438.06				**770.37**			
T#		2619.08				**657.29**			
E		2748.12				**476.27**			
S		2835.15				347.23			
L		2948.24				260.20			
K						147.11			



#10286-10286 RT:57.93-57.93 NL: 1.49E3



Reference Scan(s)	Sequence	MH+	z	P P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein retrotransposon protein, putative, unclassified				0.0008	10.1	0.0	0			
11114	R.CQVQDITKAS#VINAAS#AGLLEGELT#R.K	2927.30	3	0.0008	2.822	0.039	102.6	103	22/250	

1 of 4 peptide matches reported, 3 removed due to filtering

Reference: LOC_Os04g11620.1|13104.m01057|protein retrotransposon protein, pu
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 1847 Monoisotopic MW: 204095.5 pI: 9.72

Protein:

MAKAMEQERR EAKSRASALP GQRRLRRRVQ GEDYLPKEDD FALASSPPQR
PKKPFRPIKP RVNRVDDRQT AHQGPMSPYI IMYPCKCPFH KAGAVRGDLR
AKPLRANATE EGREGEGCKA EGKQRRTGHG RQGGSRAFGA SGRRWRRTIC
FHTCAITNAF GPCRLCASAE RGRRGEGGCG GKGTPDQGGD PFDKPTGGAP
SRSIAAGSSL DYARRCAGPK QQQAYEAARR TKVTSAPILQ YSAGYVPPQV
HPQVMTQPFP PQVAQAPMYF AGQQSAAVAA QVHHQPPSQA PQIVAEGASA
LQAQLQAFLQ QLNQPHCISS TTPSAHPEGN TSQGAPNWLP PIQPGLGVSL
WNQGPQFDFV NAAQAPTVRQ QAPTPGFRTN QAPIQVAMTW SQPIFDPSMA
AQQQGAVNRA GGEKGLPLSG GIKTRPIPPQ PKFPPVPRYS GETDPKEFLS
IYESAIEAAH GDENTKAKVI HLALDGIARS WYFNLPANSI YSWEQLRDVF
VLNFRGTYKE PKTQQHLLGI RQKQGESIRE YMRRFSQARC QVQDITKASV
INAASAGLLE GELTRKIANK EPQTLEMLLR IIDGFARGEE DSKRRQAIQV
EYDKASVAAA QAQAQVQVAE PPPLFVRQSQ PAIQGQPPRQ GQAPITWRKF
RTDRAGKAVM AVVEVQALRK EFDAEKASNH QQPARKRSEK TSTAPFTNAL
RTPRNNAETL GNVATRKTQG RSREQLSKHP AKQFRSKHPQ LNNAKMYSEG
ITSAGKGAPQ YLNQLISFGP EDAEEVMFPH QDPLVISAEI AGFEVRRILV
DGGSSADVIF AEAYAKIGLP TQALTPAPAS LRGFGGEAVQ VLGQALLLIA
PGSGENRREE QILFDVVDIP YNYNAIPGRA TLNKFEAIFH HNYLKLKMPG
PTGVIVVKGL QRSAASKGDL AIINRAVHSV BTEPHERLKH TPKPTPHGKI
TKVQIDDADP TKLVSLGGDM GEEEVESILE VLKKNIDIFA WSPDEVGGVP
ADLIMHHLAV KPDIKPRKQK LRKMSADRQE AAKAEVQKLL RAGVMQEIDH
PEWLANPVLD AFRLEERRSN FRRLVYKVLG KQLGRNVEAY VDDIVVKSRK
AFDHAIDLQE TFDSLRTAGI RLNPEKCVFG VRAGKLLGFL VSERGIEANP
EKIDAIQQMK PPSSVHEVQK LAGRIAALSR FLSKAPERGL PFFKTLRGAG
KFNWTPECQA AFDELKQYLQ STPALISTPP GSELLLYLAA SPVAVSAALV
QETEPGQKPV YPVSEALQGA KTRYIEMEKL AYALVMASRK LKHYFQAHKV
IVPSQYPLGE ILRGKEITGR LSKWAAELSP FNLHFVARSA IKSQVLANFV
AEWTPVLAPD PEPAEQFWVM CSDGSWSHKG AGIAAVLFSP NGVPIWYAAR
LQFDTTNNAA EYEAVLLGLR KAKALGVRRL LIRTDSKLVV GHVNKSFEAK
EEGMKRHLEA IRSMEKCFTG ITVEHLPRDQ NEEADALAKS AACGGPHSPS
IFFEVLHAPS VPMDSSEVMA IDQEKLGEDP YDWRTPFVKH LETGWLFVDE
AEAKRLQLRA MKYKMVSGQL YRSGVLQPLL RCISFAEGEE MAKEIHQGLC
GVHQAARTVA SKQYRSLYGK NIVCRFGVPK EFITDNAKQF DSDKFREMCE
GLNLEIRFAS VAHPQSNGAA ERTNGKMLEA LKKRLEGAAK GKWPEELLSV
LWALRTTPTR PTKFSLFMLL YGDEAMTPAE LGANSPRVMF SGGEEGCEVS
LELLEGVRVE ALEHMHKYTT STSSTYNKKV RPMELMPGHL VLRKKANPVA
VGKLESKWEG PYLIKHKSRT GYFRLSTLKG EEFDHSWNAA SLKCFYV

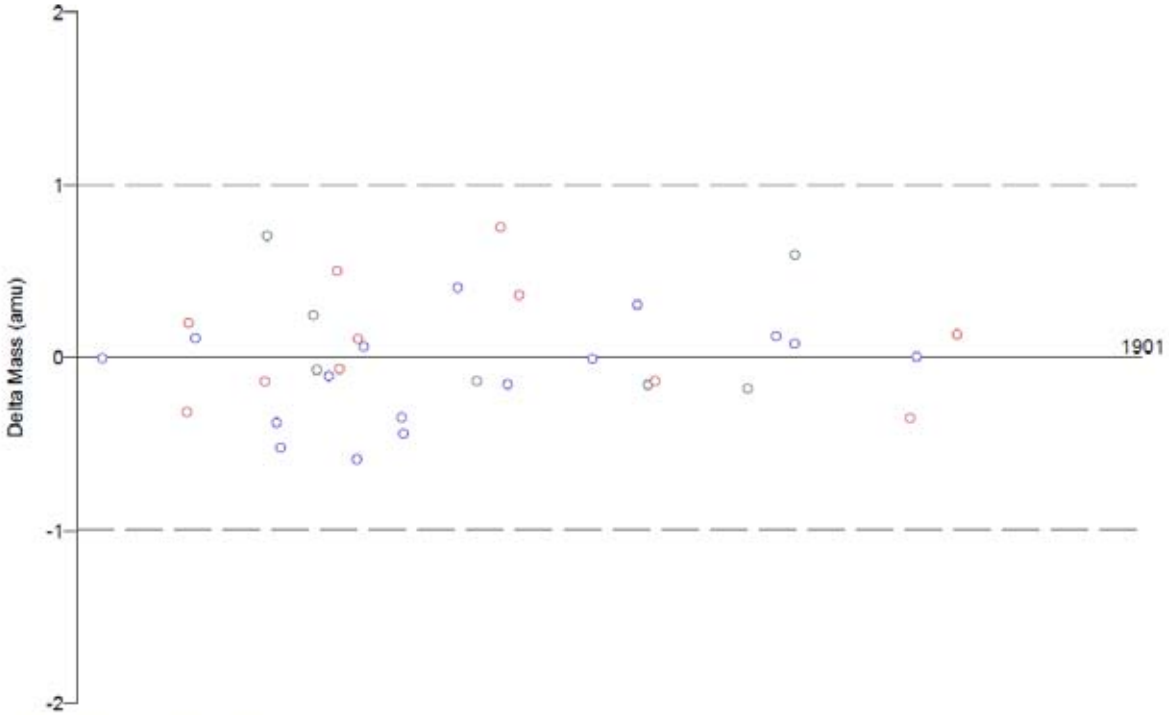
Protein Coverage:

Sequence	MH+	% Mass	AA	% AA
PTGGAPSRSIAAGSSLDYARR	2090.07	1.02	195 - 215	1.14
CQVQDITKASVINAASAGLLEGELTR	2687.40	1.32	540 - 565	1.41
RRQAIQVEYDK	1405.75	0.69	594 - 604	0.60
TSTAPFTNALRTPR	1532.82	0.75	691 - 704	0.76
Totals:	2687.40	1.32	26	1.41

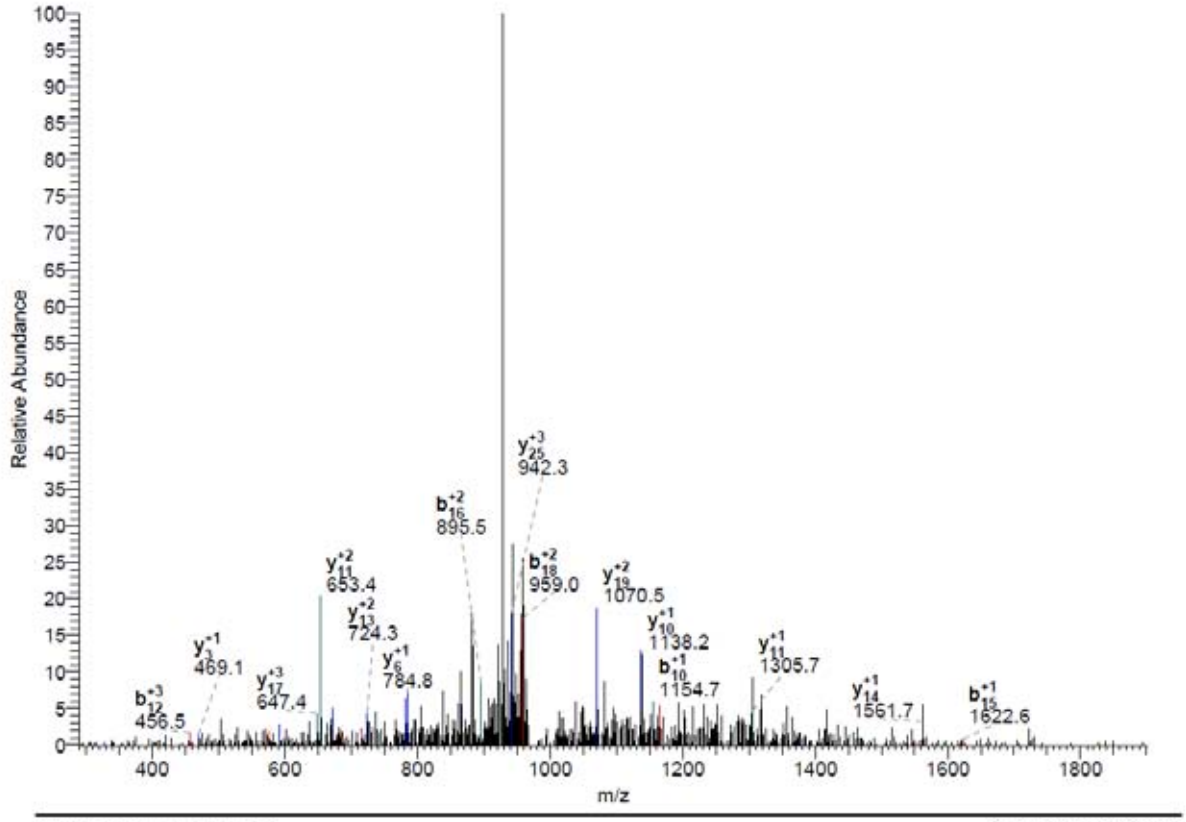
DTA for scan: 11114
Precursor ion: 976.75
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
C		104.02							
Q		232.08				2824.29			
V		331.14				2696.23			
Q		459.20				2597.17			
D		**574.23**				2469.11			
I		**687.31**				2354.08			
T		788.36				2241.00			
K		916.46				2139.95			
A		987.49				2011.85			
S#		**1154.49**				1940.82			
V		1253.56				1773.82			
I		1366.64				1674.75			
N		1480.69				**1561.67**			
A		**1551.72**				1447.62			
A		**1622.76**				**1376.59**			
S#		1789.76				**1305.55**			
A		1860.80				**1138.55**			
G		1917.82				1067.51			
L		2030.90				1010.49			
L		2143.99				897.41			
E		2273.03				**784.32**			
G		2330.05				655.28			
E		2459.09				**598.26**			
L		2572.18				**469.22**			
T#		2753.19				356.13			
R						175.12			



#11114-11114 RT:62.61-62.61 NL: 3.69E2



Reference Scan(s)	Sequence	MH+	z	P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein peptide-N4-asparagine amidase A, putative				0.0004	10.1	0.0	0			
12931	R.LSPADIAVPDAAAADTPTT#YFEVDRPLR.P	3052.47	3	0.0004	2.634	0.110	173.8	4	21/162	

1 of 3 peptide matches reported, 2 removed due to filtering

Reference: LOC_Os01g10960.1|13101.m01183|protein peptide-N4-asparagine amida
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 612 Monoisotopic MW: 65646.9 pI: 4.86

Protein:

MSRVHLLLLL VFYIPAAIAS RRNLRLSPAD IAVPDAAAAD TPTTYFEVDR
PLRPPPGSSG PCSTLLLSSS FGPTYTKPPV TAAYSPPDCL AAAGGGAPAI
SLAVLEWRAT CRGVQFDRIF GVWLGGVELL RSCTAEPRPN GIVWSVSKDV
TRYASLLAAG NSTLAVYLGN LIDDQYTGVY HANITLHLYF GPTPARQPAP
ATAPADIIVP VSRSLPLNDG LWFQIQNATD VESASIVLPS NTYRAVLEVY
VSFHGDDEFW YTHTPDGNGP FREVTVLVDG DLVGAVWPFP VIFTGGINPL
LWRPITGIGS FNLPTYDIEL TPFLAKLLDG KAHELAFAVT NAVDVWYVDG
NLHLWLDPMT TATTGSLVSY DAPRLAAVNT SHTTASRFDG LSERYYYHTT
ASRRISAAGW VESPSHGRIT TNATQTFAFE NTYAFAGDGS AETVNQTTVA
DAAVSATDLA GAVLYSRQAH HDFPLYVDIE AKTSPHAADV TYTVAREYRE
TAIAAGRWLS SGTPPRRYSL RDTQSGAVDV EMRDGNAVSA TWGTRQTYRL
EATDGCYFRN VTSSGYSVAS DESDEVCSDS QEYPAGGAVI GALPPAEAAA
AVTASADELV RK

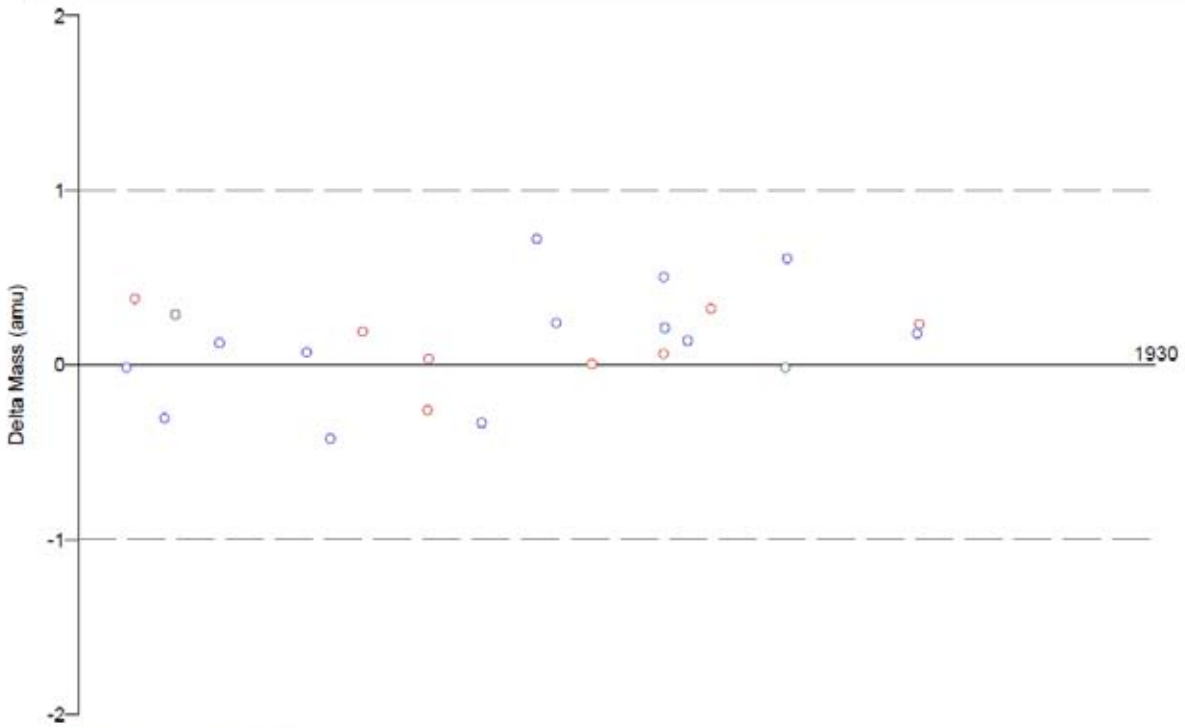
Protein Coverage:

Sequence	MH+	% Mass	AA	% AA
LSPADIAVPDAAAADTPTTYFEVDRPLR	2972.50	4.53	26 - 53	4.58
SCTAEPRPNGIVWSVSK	1830.92	2.79	132 - 148	2.78
Totals:	2972.50	4.53	28	4.58

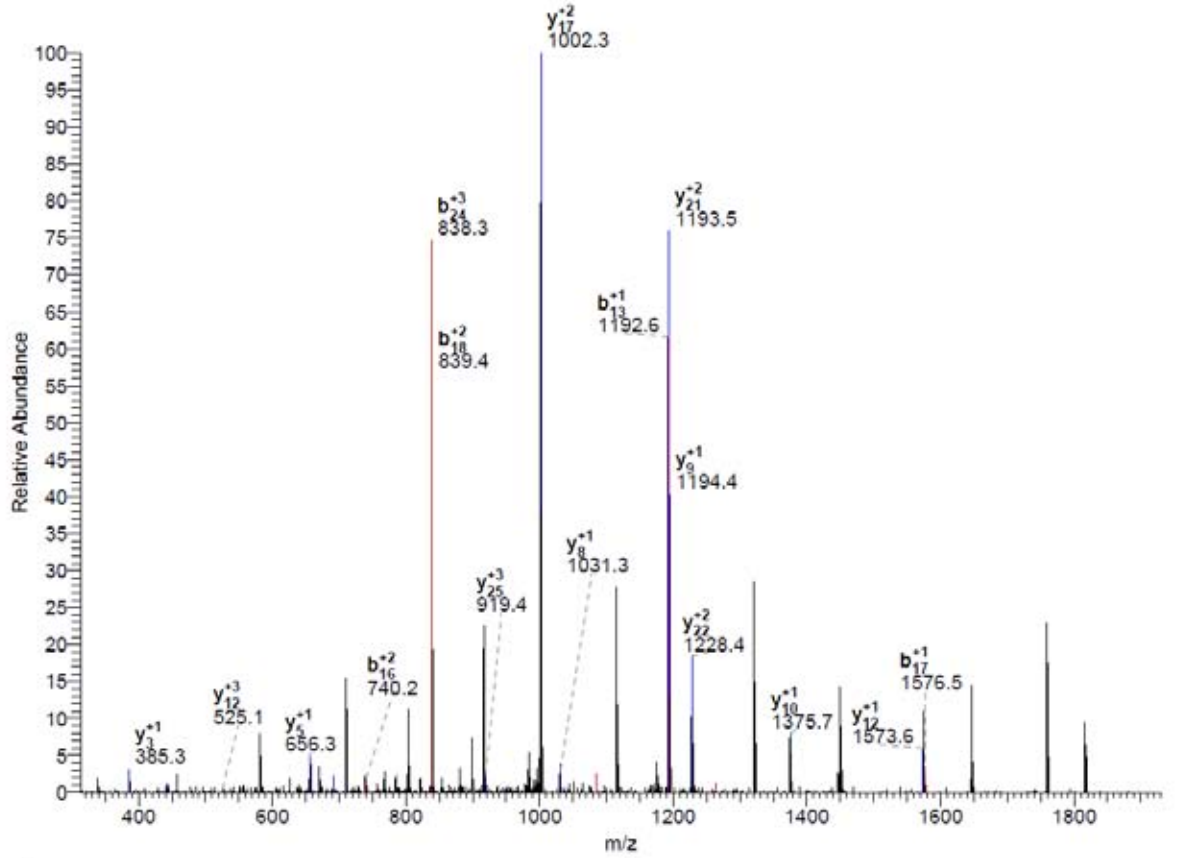
DTA for scan: 12931
Precursor ion: 1017.04
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
L		114.09							
S		201.12				2939.38			
P		298.18				2852.35			
A		369.21				2755.30			
D		484.24				2684.26			
I		597.32				2569.23			
A		668.36				2456.15			
V		767.43				2385.11			
P		864.48				2286.04			
D		979.51				2188.99			
A		1050.55				2073.96			
A		1121.58				2002.93			
A		**1192.62**				1931.89			
A		**1263.66**				1860.85			
D		1378.68				1789.82			
T		1479.73				1674.79			
P		**1576.79**				**1573.74**			
T		1677.83				1476.69			
T#		1858.85				**1375.64**			
Y		2021.91				**1194.63**			
F		2168.98				**1031.56**			
E		2298.02				884.49			
V		2397.09				755.45			
D		2512.12				**656.38**			
R		2668.22				541.36			
P		2765.27				**385.26**			
L		2878.35				288.20			
R						175.12			



#12931-12931 RT:73.05-73.05 NL: 1.61E3



BioworksBrowser rev. 3.3 25/08/2009 13:04:48

Reference Scan(s)	Sequence	MH+	z	P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein retrotransposon protein, putative, Ty3-gypsy subclass				0.0002	10.1	0.0	0			
13239	R.QSLAWTCPVET#PRVEPVPVM*VTTRR.N	2998.47	3	0.0002	2.583	0.002	133.3	150	18/144	

1 of 1 peptide matches reported, 0 removed due to filtering

Reference: LOC_Os04g04574.1|13104.m00381|protein retrotransposon protein, pu
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 747 Monoisotopic MW: 83417.5 pI: 9.78

Protein:

MSWQRVGALL LGAQSCLPVP GVPAVGGIGS VLLSMARMGW KLCHVFRPYT
VVMAGSVTRR GLDMTGFVLE LRGMCVVLGY PHGVDYQARP LPEQRGEDAE
PHAGWEVTAV ILSGSPGRTS LAVTAGGDSF PAACQNAALL AIGTLHQRYP
DELQHSPYRY HPRRGGARDF ATFRDASSED DATIVHLARM VEAYDAARID
FHQMVRRGMV ENNMKILELR QENLQLKKDL DAVEAQLHQL KIAQGEDCRP
KRRRVCRSQK ITARKSTSRP ELVRQSLAWT CPVETPRVEP VPVMVTTRRN
VNTGDGNQPE GSNHNHQGNP PPPPPPPPPP PPDTNAILTQ ILAQQANMMT
AFLHHLQNPS QQNAPPPPPQ HSKLAEPLRI RPPTPSSSNN PVDALDWLHA
VGKKLDTVQC SDEEKVIFAA HQLQGPASLW WDHFQATQPE GQPITWARPT
AAFRRTHVPA GVVALKKREF RELKQGNRSV MEYLHEFNNL ARYAPEDVRE
DEEKQEKFLA GMDPELSVRL VSGDYPDFQR LVDKSIRLEA KHKELESHKR
RLASFRNQQG ANQRVRYTNP YPGGSSSQQQ QQQQQPRSAP RPQPVVRVPQ
PQQQQNQQGT RAPRPPTPAV QPVQGRRDAQ GPQRLCFNCP EPGHFADKCP
KPRRQQGQAP PRPNNGGKDV IRGRVNHVTA EDVLTTPNVI VGTPLIHSIP
ATILFDSGAS HSFISVPFVG KNQLGVERLR KPSAHHHPWR GYDSKVL

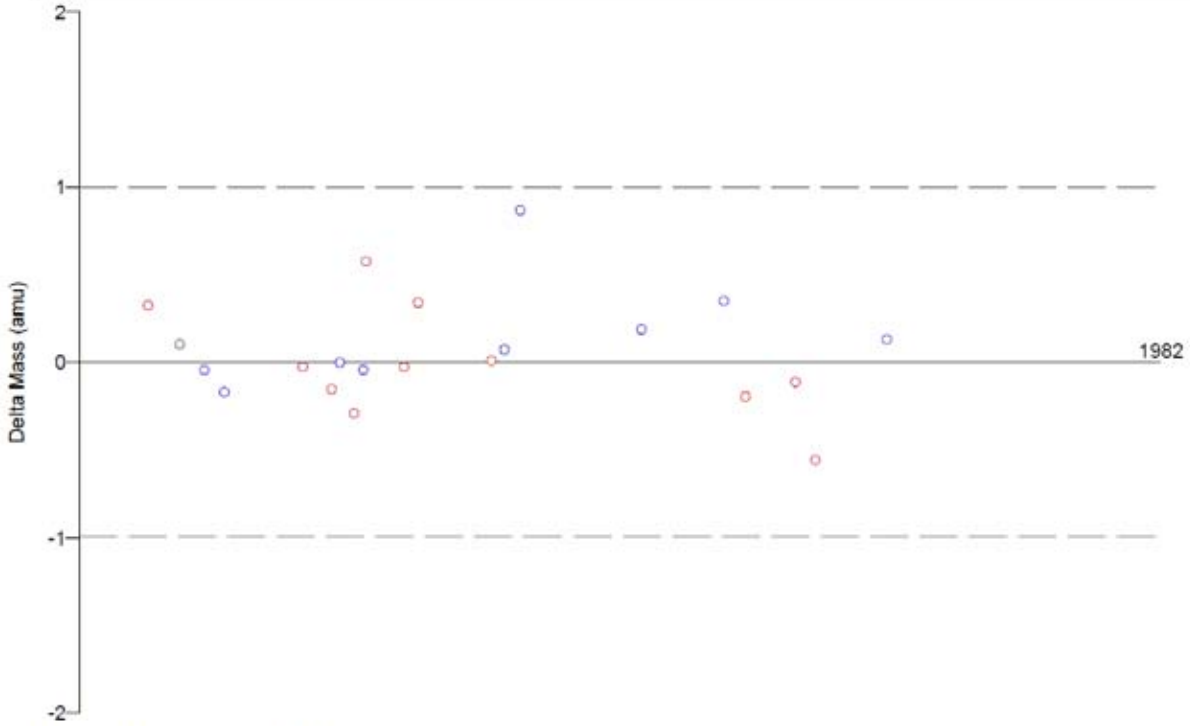
Protein Coverage:

Sequence	MH+	% Mass	AA	% AA
QSLAWTCPVETPRVEPVPVMVTTRR	2902.51	3.48	275 - 299	3.35
Totals:	2902.51	3.48	25	3.35

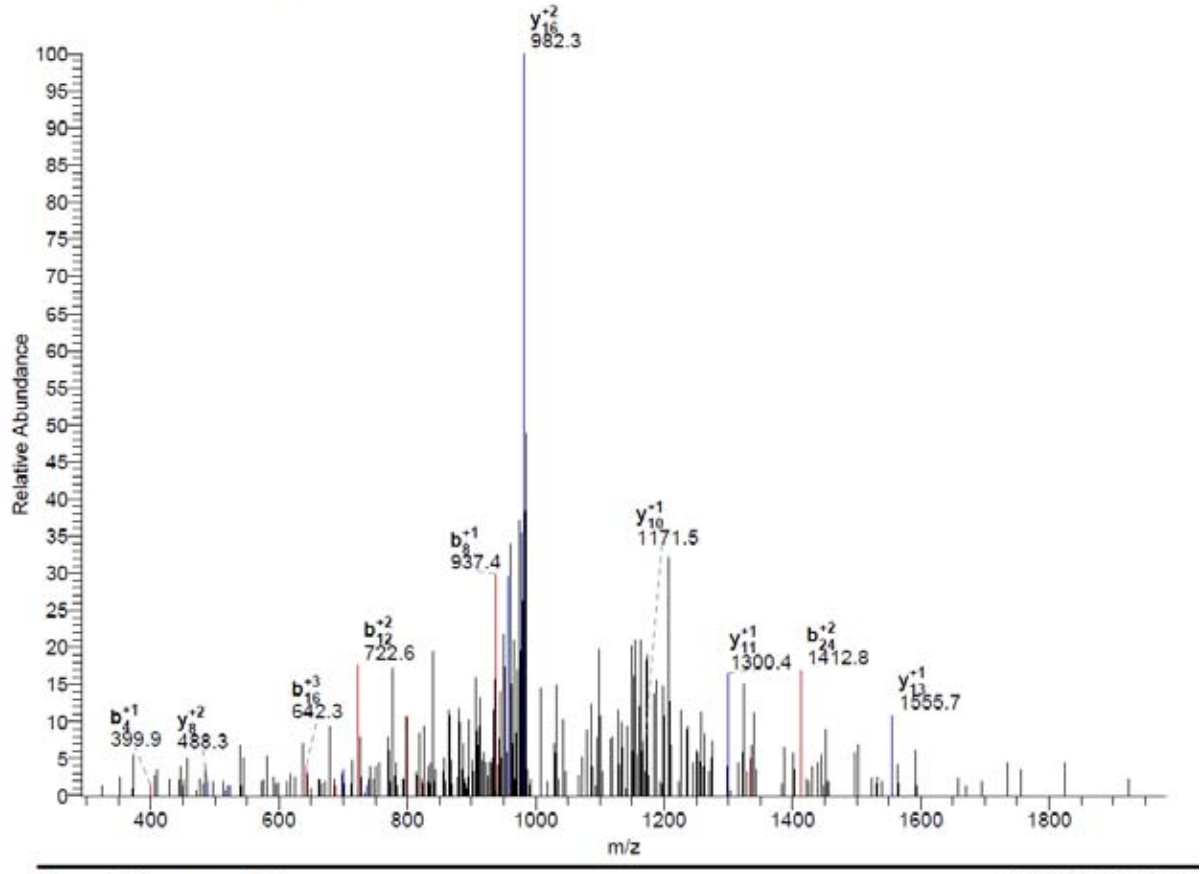
DTA for scan: 13239
Precursor ion: 1000.48
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
Q		129.07							
S		216.10				2870.41			
L		329.18				2783.38			
A		400.22				2670.29			
W		586.30				2599.26			
T		687.35				2413.18			
C		790.36				2312.13			
F		937.42				2209.12			
V		1036.49				2062.05			
E		1165.53				1962.98			
T#		1346.55				1833.94			
P		1443.60				1652.93			
R		1599.70				1555.87			
V		1698.77				1399.77			
E		1827.81				1300.70			
P		1924.87				1171.66			
V		2023.93				1074.61			
P		2120.99				975.54			
V		2220.06				878.49			
M*		2367.09				779.42			
V		2466.16				632.38			
T		2567.21				533.32			
T		2668.26				432.27			
R		2824.36				331.22			
R						175.12			



#13239-13239 RT:74.79-74.79 NL: 6.98E1



Reference Scan(s)	Sequence	MH+	z	P P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein integral membrane protein, putative, expressed				0.0003	10.1	0.0	0			
9347	K.ALPSPAQAAAASAM*AFSVGAVVPLLAAGFIVNYR	3317.77	3	0.0003	2.570	0.018	140.6	20	16/132	

1 of 1 peptide matches reported, 0 removed due to filtering

Reference: LOC_Os04g59020.1|13104.m06217|protein integral membrane protein,
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 199 Monoisotopic MW: 19984.3 pI: 6.60

Protein:

MAAMMNNERS SSNKLQVDAE NPAAVGDELD LAARANWLRA AVLGANDGLV
STASLMLGVG AVKAEAARAMV ISGFAGLLAG ACSMAIGEFV SVCSQRDVEL
AQLERDGKRG GEEEKALPSP AQAAAASAMA FSVGAVVPLL AAGFIVNYRL
RIAVVVAVAS VALAAFGCVG AVLGRAAVAR SSARVVLGGW PPWASPSAS

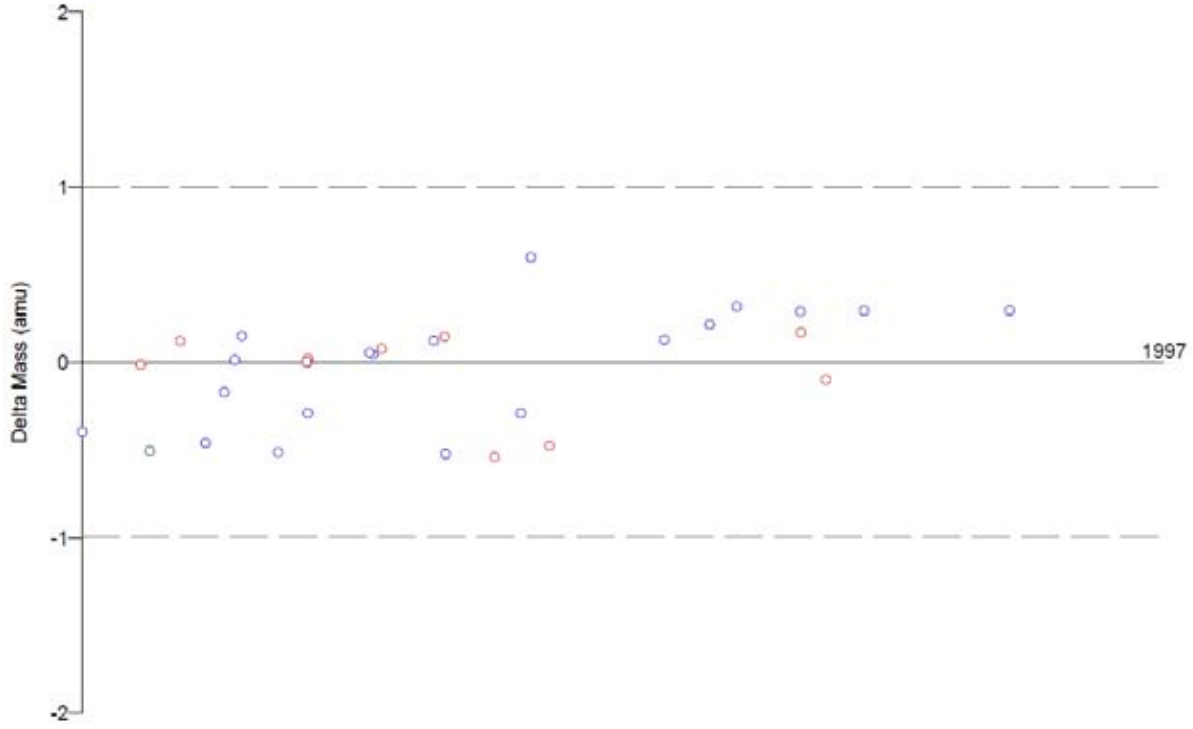
Protein Coverage:

Sequence	MH+	% Mass	AA	% AA
ALPSPAQAAAASAMAFSVGAVVPLLAAGFIVNYR	3301.78	16.52	116 - 149	17.09
Totals:	3301.78	16.52	34	17.09

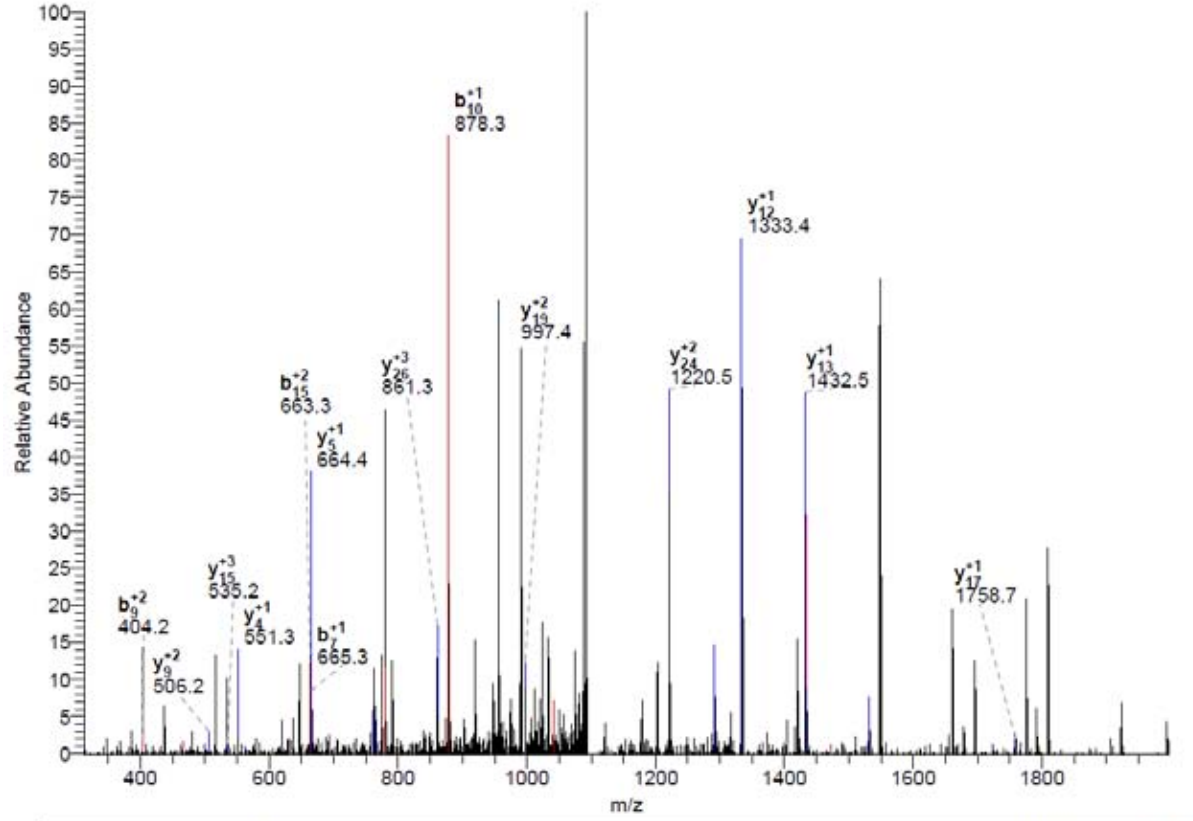
DTA for scan: 9347
Precursor ion: 1106.92
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
A		72.04				3246.73			
L		185.13				3133.65			
P		282.18				3036.60			
S		369.21				2949.57			
P		**466.27**				2852.51			
A		537.30				2781.48			
Q		**665.36**				2653.42			
A		736.40				2582.38			
A		807.44				2511.34			
A		**878.47**				2440.31			
S		949.51				2369.27			
A		1036.54				2282.24			
M*		1107.58				2211.20			
A		1254.61				2064.16			
F		1325.65				1993.13			
S		**1472.72**				1846.06			
V		1559.75				**1759.03**			
G		1658.82				1659.96			
A		1715.84				1602.94			
V		1786.88				**1531.90**			
V		1885.95				**1432.83**			
P		1985.02				**1333.76**			
L		2082.07				1236.71			
L		2195.15				1123.63			
A		2308.24				1010.54			
A		2379.27				939.50			
G		2450.31				868.47			
P		2507.33				811.45			
I		2654.40				**664.38**			
V		2767.49				**551.29**			
N		2866.55				452.23			
Y		2980.60				338.18			
R		3143.66				175.12			



#9347-9347 RT:52.62-52.62 NL: 6.43E2



Reference Scan(s)	Sequence	MH+	z	P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein retrotransposon protein, putative, unclassified				0.0006	10.1	0.0	0			
11248	K.NKLPHMHIILYGENVK.N	1906.04	2	0.0006	2.225	0.039	241.9	1	10/30	

1 of 2 peptide matches reported, 1 removed due to filtering

Reference: LOC_Os11g44014.1|13111.m04391|protein retrotransposon protein, pu
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 1643 MonoIsotopic MW: 184732.8 pI: 6.13

Protein:

MGEADREGGN GAEGGHGSTC VGARAIAHDG EGGGSDDSSR NDDNGYHDSN
PLPSTAWRPR GGRISLPSYN PPPEPILNLL TSASSLSNHF LQHIRRYNSM
FAMTSMGAKI IESINDGHGP YVFKISGQIC HRVGSLIPSQ GRRPEYAQLY
IFDTENEISN RINIASSSRD SFHANEEIVR SLIEMFDTHN PIVKLFRTAR
ERLSENESDH YKIRLFGSTD AHGDIYSAPV AAEVVGLVVG DIGVTDIGRD
IIIQHHSSQL QRIDEKHRKF MAMQYPILFP YGEDGFHESI MYNQTASSSS
LRRNKATMVE FFAYIMHDRA GQFNTPLRCG RKYGCPHLFV TFTSNAYWPE
ILQALAPGQQ PSDRPDIVDR VFKMKLNMLM DDIKKNNFFG PIHAVVYTVE
FQKCGLPHVH IIIWLSKEEP LDAEKVDLRI SAQLPNPTLD PIGFEAVTSF
MIHGPCGPGI SYSPCMSEGR CSKFYPKEFC EHTSILQNGF TQYARPNNQI
VVTKNGVDID NRFIVPHNVD LVVKYQAHIN VESVNHDGMH KYLFKYVTKG
YDCSRAGIRR NSANETINEI DNYLECQCVT PNDAAWHLLQ FDIHHTDPSV
ERLPVHLPLE NNVVYIEDDD LEEVIENPGN QKSKLTAWLE ANSQFPQARE
HTYIEFHEYF TWHASEKYWD IRRGCYNKIG RIAHVDPTKG EQYYLRMLLH
IVKGPKTSEI RNISGQQHPT FRAACEALGL LGDDQEWSHA LNDAAQWALP
YQLRQLFVTI LLPCEVTNPQ RLPTEHAQHM SEDFRYRTNQ NLSQSNSSPT
DSFVGNALLF ELDKLLRNAG YSLSHFNLPL PDDIGSASAD NRLLLDELSY
DITSIASTSA NDINSLNTNQ KEIFNSISNS VINNEGRTFP VYGYGGTGKT
PLWTTLLNFV RTQGKIALAV ASSGIAALLL PGGRTPHSRP KIPLDIRENS
MCSIKKNTHL AELIQQTSLI VWDEAPVNHK YCFEALDRTL RDILSDIRPN
AQHRQFGGIT VAFGGDFRQT LPVIQNATRH QILRASIVNS YLWHQCVVLQ
LTENMRLSSQ NLSPSDKEEL RVFADWLLRV GNGTEPHISI ENETNGYVPS
NITTYFCDRA ILPPINEVVS EINNKIIAQV TAAEMSYYSS DTIDDSCANH
STLEALYPTE FLNTISLNGL PDHVLHLKIG VPIMLLRNLD PSRGLCNATR
LIVTQLTTRI IEGEIMTGKA KGSKAYIPRI ITTSAQSKWP FKLKRRQFPI
RLSYAMTINK SQDLIGVISH VGPYDYAGKT SSKKNRKLKI RNKDEQEQEI
VLWGEYGESF DEAFVLQKST DHKIVVAILA GLTAGTYLGK TEATSSSATQ
IYFDSDITEI AEYQSRYKLP LTIKDESGTL DAVAFYNVAE DLVEVNATQA
TQNLKIDATE HAIALDTAIG KTRLFHIAMN TKYSSHFTIN YVLKKSYPVE
NENTSLMLPT LENTKVAKES ATKQLATDEG LTTMEHCSPA IANTIQVATN
QLYHSQQVDL IKEKQPSTEP SPQQNSKRHK KVTETSPNGE ENQLQQPKIA
DQQPSGHEEQ MDQQPEKNRN HEAQLIHTNY LQAVSKITAS KSTAISITLQ
GKSSTDDLYK LQPTQAHVVN YTKNKLPHMH IILYGENVKN NEI

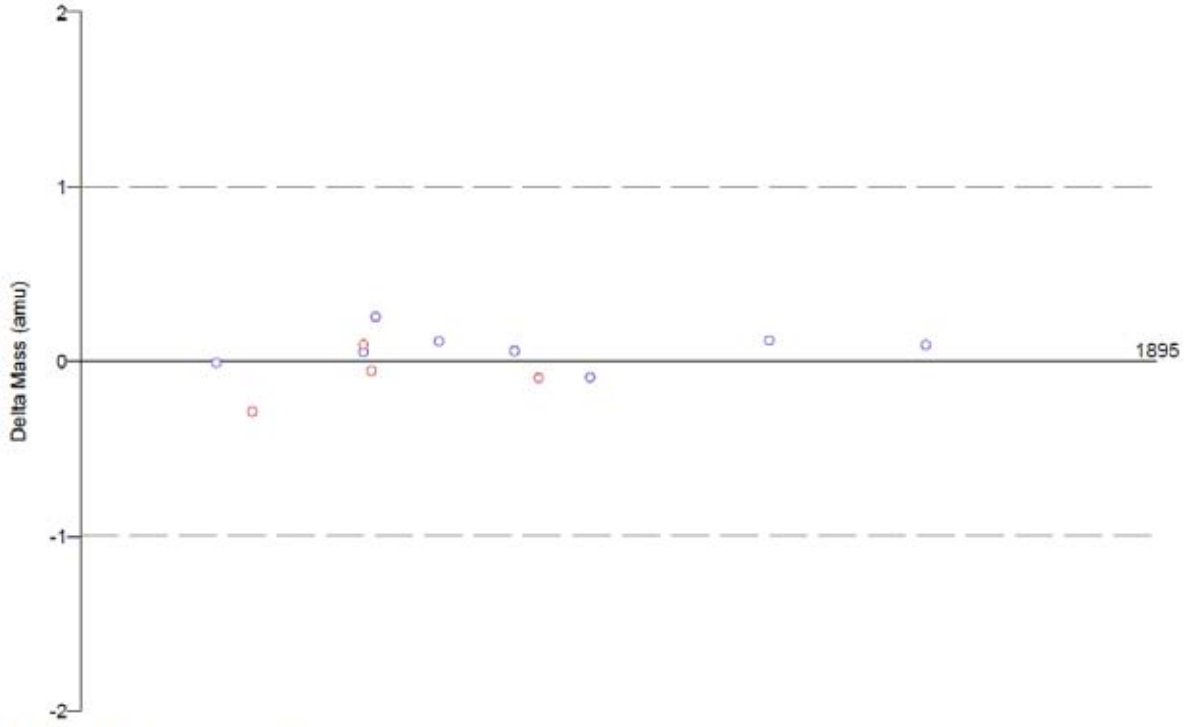
Protein Coverage:

Sequence	MH+	% Mass	AA	% AA
ITASKSTAISITLQGKSSTDDLYK	2528.35	1.37	1587 - 1610	1.46
NKLPHMHIILYGENVK	1906.04	1.03	1624 - 1639	0.97
Totals:	1906.04	1.03	16	0.97

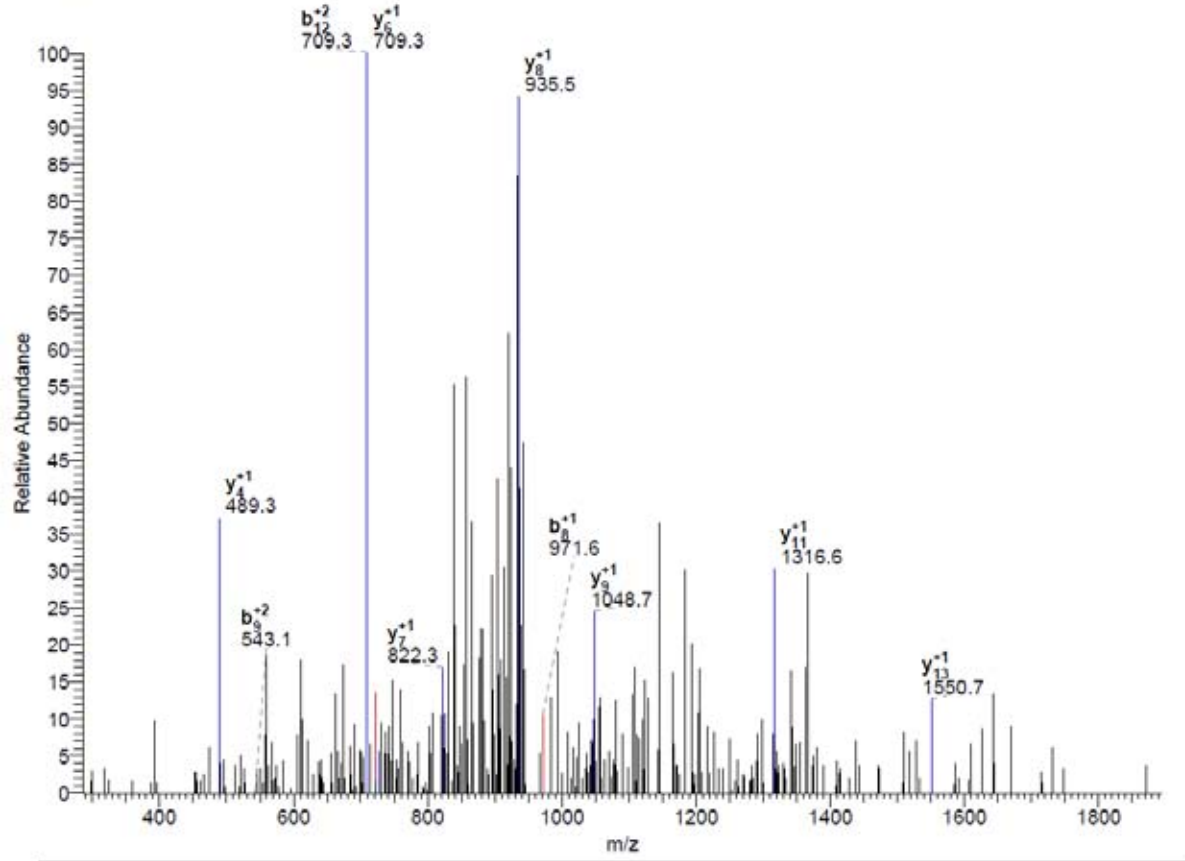
DTA for scan: 11248
Precursor ion: 953.40
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
N		115.05							
K		243.15				1791.99			
L		356.23				1663.90			
P		453.28				**1550.81**			
H		590.34				1453.76			
M		721.38				**1316.70**			
H		858.44				1185.66			
I		971.52				**1048.60**			
I		1084.61				**935.52**			
L		1197.69				**822.44**			
Y		1360.76				**709.35**			
G		1417.78				546.29			
E		1546.82				**489.27**			
N		1660.86				360.22			
V		1759.93				246.18			
K						147.11			



#11248-11248 RT:63.39-63.39 NL: 5.42E1



Reference Scan(s)	Sequence	MH+	z	P P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein retrotransposon protein, putative, unclassified				3e-005	8.1	0.0	0			
15687	K.GEAT#GGKQGDAGAPS#GERS#ADK.W	2285.85	3	3e-005	2.936	0.068	122.6	3	31/210	

1 of 1 peptide matches reported, 0 removed due to filtering

Reference: LOC_Os05g31410.1|13105.m03223|protein retrotransposon protein, pu
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 326 Monoisotopic MW: 35978.9 pI: 8.56

Protein:

MAEEAIVYTP TPIGVEGTAS SRAPRHPRQR ARPSISPLSR ASYHASVASL
RNVRWTPKFR PSLTEKYDGN VNPSEFLQIY TAIIEAAGGD DRVMANFFPM
TLKGQARVWL MNLPPASVHS WEDLCWQFTT NFQGTYLGSG EEADLHAVQR
LDDESLQQYI QFCQVSNTIL HPCSCSDLRA QGGVRHNHML EKVASKEPQT
TAQLFKLAGH VACKEEAWTW NSSHFGAVAA DALEPAARSG QREKRKKKRS
ARSNDEGHVL TAEGALRPPR KGEATGGKQG DAGAPSGERS ADKWCSVLNT
SNHSLADCRS VKNMAELVKK FEQERN

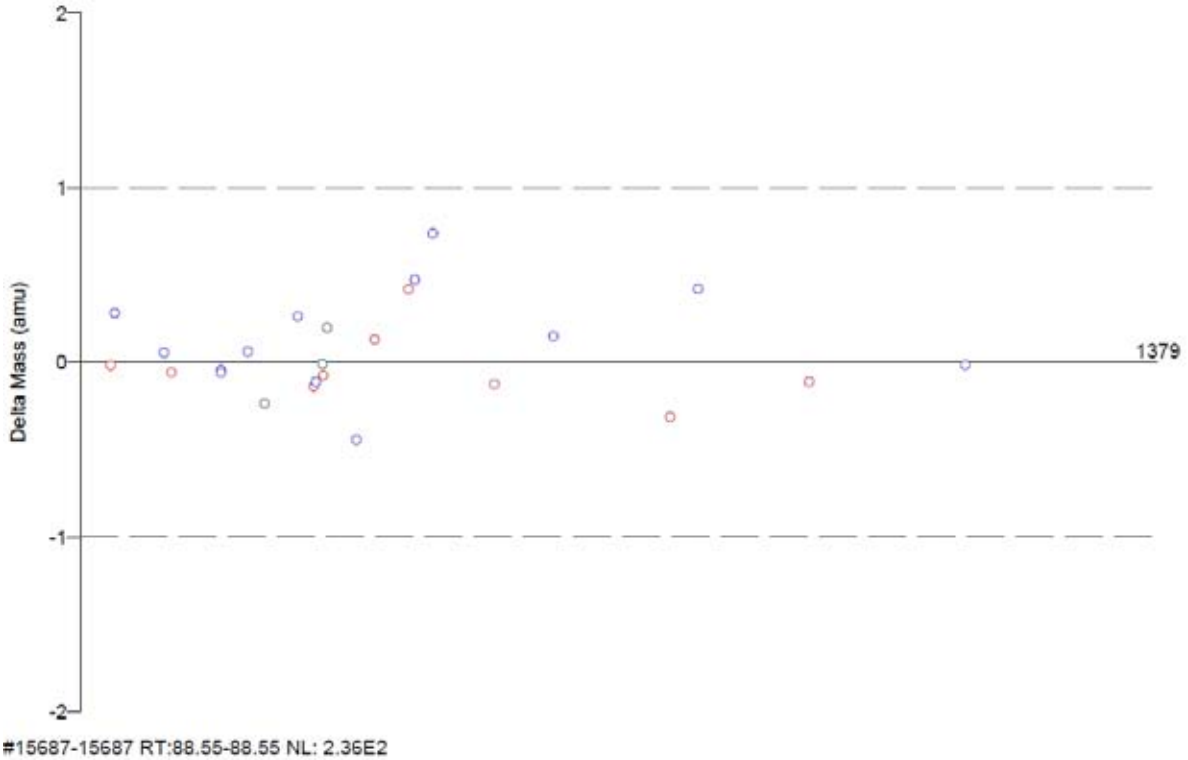
Protein Coverage:

Sequence		MH+	% Mass	AA	% AA
GEATGGKQGDAGAPSGERSADK		2045.95	5.69	272 - 293	6.75
	Totals:	2045.95	5.69	22	6.75

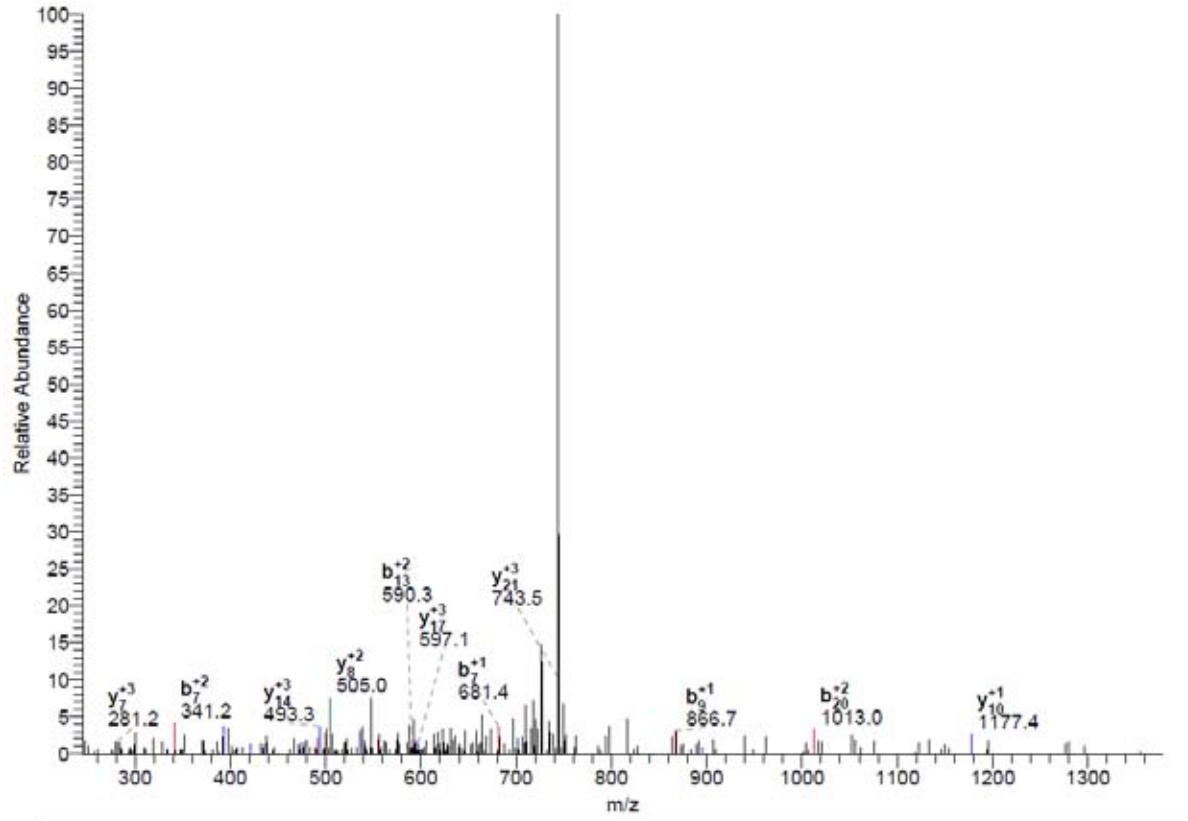
DTA for scan: 15687
Precursor ion: 762.54
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
G		58.03				2228.83			
E		187.07				2099.78			
A		258.11				2028.75			
T#		439.12				1847.73			
G		496.14				1790.71			
G		553.17				1733.69			
K		681.26				1605.59			
Q		809.32				1477.54			
G		866.34				1420.51			
D		981.37				1305.49			
A		1052.40				1234.45			
G		1109.43				1177.43			
A		1180.46				1106.39			
P		1277.52				1009.34			
S#		1444.51				842.34			
G		1501.54				785.32			
E		1630.58				656.28			
R		1786.68				500.18			
S#		1953.68				333.18			
A		2024.71				262.14			
D		2139.74				147.11			
K									



#15687-15687 RT:88.55-88.55 NL: 2.36E2



Reference Scan(s)	Sequence	MH+	z	P P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein hypothetical protein				4e-005	8.1	0.0	0			
11370	K.LELRAT#ANKGSVS#K.D	1633.77	2	4e-005	2.092	0.050	126.3	26	10/52	

1 of 1 peptide matches reported, 0 removed due to filtering

Reference: LOC_Os02g29440.1|13102.m03207|protein hypothetical protein
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 104 Monoisotopic MW: 11328.5 pI: 4.33

Protein:

MWRTTLGDLE AKLELRATAN KGSVSKDSPV IRSPKQLVDV YVEEPEIIDL
CDGSDDDSDY EAYMELKMDS EEECAKKDVE VDVSKGKGVT KAAGSGCKGK
AMVE

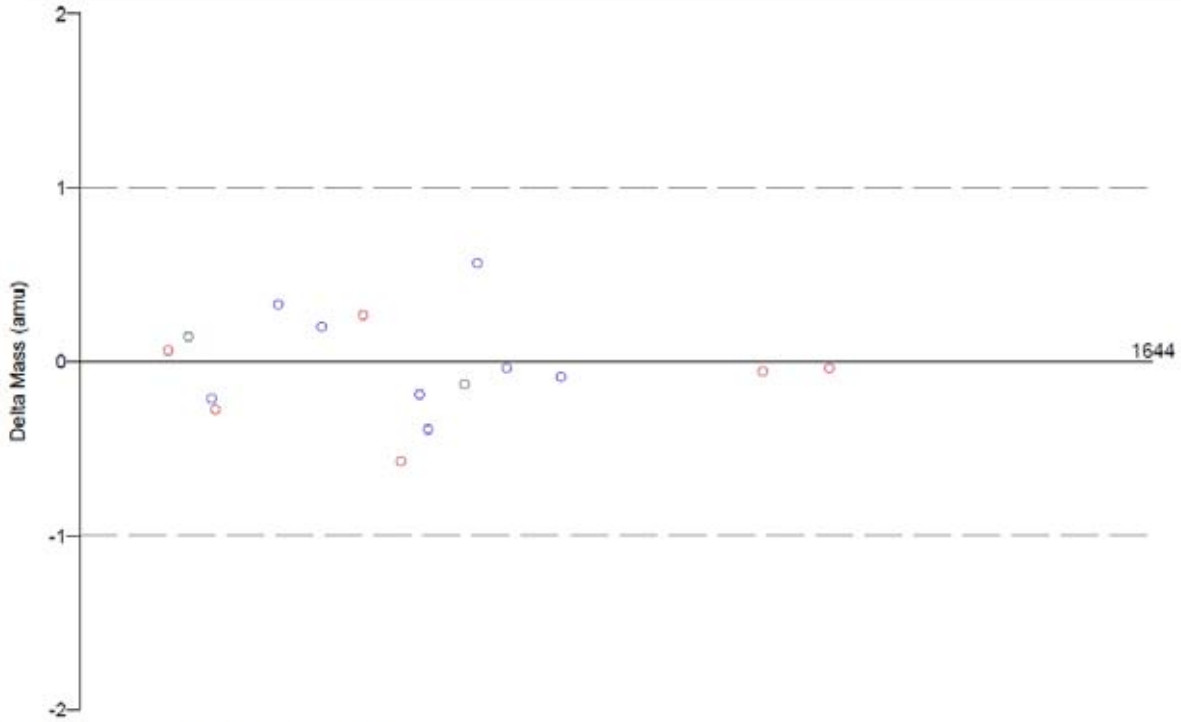
Protein Coverage:

Sequence		MH+	% Mass	AA	% AA
LELRATANKGSVSK		1473.84	13.01	13 - 26	13.46
	Totals:	1473.84	13.01	14	13.46

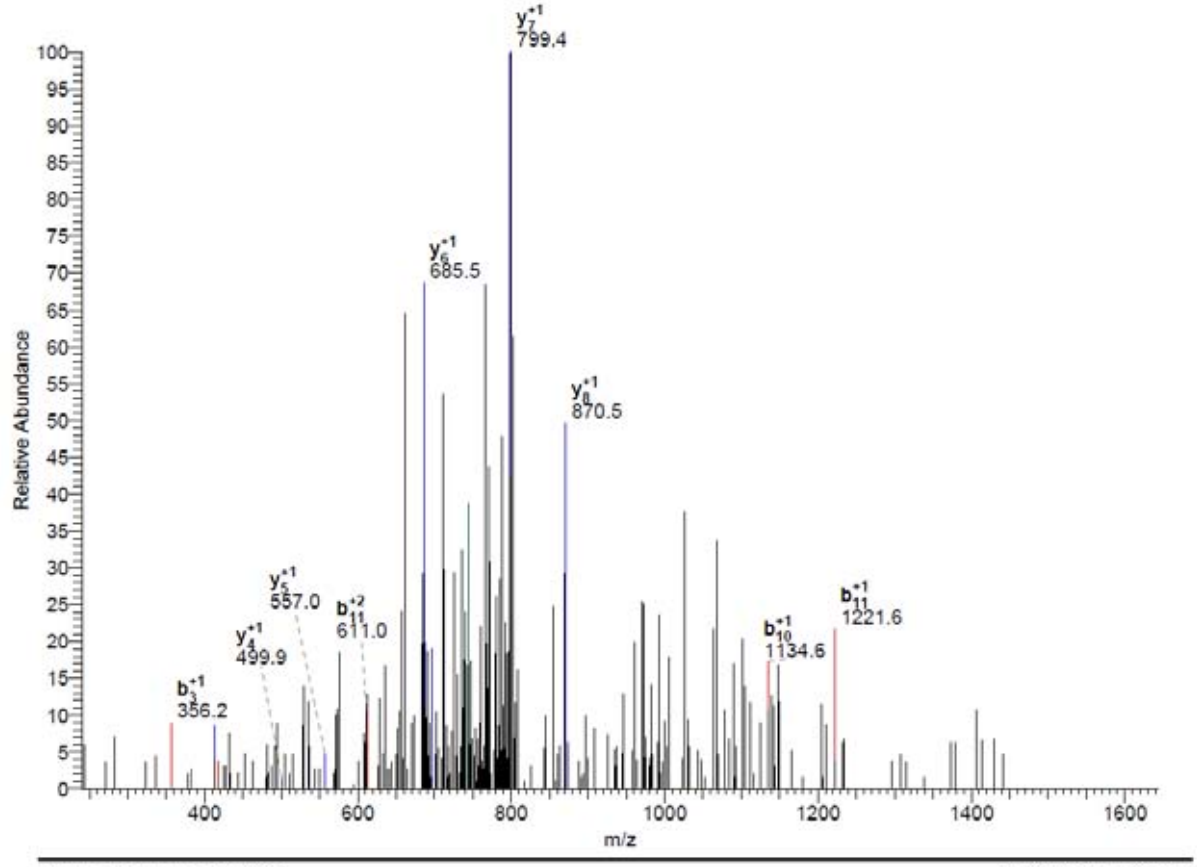
DTA for scan: 11370
Precursor ion: 817.85
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
L		114.09							
E		243.13				1520.69			
L		356.22				1391.64			
R		512.32				1278.56			
A		583.36				1122.46			
T#		764.37				1051.42			
A		835.41				870.41			
N		949.45				799.37			
K		1077.55				685.33			
G		1134.57				557.23			
S		1221.60				500.21			
V		1320.67				413.18			
S#		1487.67				314.11			
K						147.11			



#11370-11370 RT:64.09-64.09 NL: 4.30E1



Reference Scan(s)	Sequence	MH+	z	P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein pentatricopeptide, putative, expressed				0.0005	2.1	0.0	0			
9092	R.RAHGVAVRS#GLGPFELAVGNALVHM*Y#GK.C	2985.43	3	0.0005	2.625	0.026	152.8	13	29/208	

1 of 2 peptide matches reported, 1 removed due to filtering

Reference: LOC_Os02g58620.1|13102.m06790|protein pentatricopeptide, putative
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 1097 Monoisotopic MW: 116328.0 pI: 9.46

Protein:

MATAGVVVEE AVRRYAGGKP AALLPALLGL LSRQQPSSGW ARALASQLHA
DVAKRPLSAA ASNSLLCYYL RSSRLDLALH HLRCRSTPRD SLTYNTLLNH
LPASSSSSTT FRLFRFAMRH AHAAFHPNIA SLLSLLRASS SYSDHFLHMI
HAYLLKTPAS IHTPVANSLL SLYATLGDFA SAAILFGEMP DRDVASWTSM
IGACLGSGYA DQALRLFREM LADGALQPDG VVAVVVLRAC AMLEDVRAGA
SVHAVATRRG LQGDLFVDNS LVDMYAKCLD LRSARKVFDL IAVKNVVSWN
TMLSGLVHAG SYPEALHLLA LQIGVVGDET TLAVLLQLCK KKRLGGQAAR
SVHGAAIRRR LLSMALLNAL LDAYGKCGLV EDVLRLFQGM RERNVITWST
VIAACAHNAR PHAAMACFVA MLVTGERPNS ITVLSLVEAC GSCAEMWASR
RAHGVAVRSG LGPELAVGNA LVHMYGKCGE LGASARVFDT MPVKDVLTWN
SMIGALGMNG RARDALALLH RMEAEGDEVR PNGVTMLAAL WACAHGGLVE
EGIGCLESMA RQSLQPRVEH VSCVVDMLAR AGDLDGAAEI VRRSSGGGSP
AAWSALLSAC RRRGDGGGGE VGRSAAARVL ELEPGKSAGY LMSMGMGLGK
GWAAGMRWAM REKGVKVESG HSVVQHAGGS ERDLRPGGVR NELILLALPA
VLGQAIDPLA QLMETAYIGR LGALELASAG IGVSVFNIVS KIFNIPLLSI
ATSFVAEDIS KNASKHSSSG KLELSSVSSA LVLAAGIGTI EALALFLGSG
LFLKLMGVSP ASPMHKPAKL FLSLRALGAP ANVIMLAVQG IFRGFKDTKT
PVFFIGLGNL SAVVLLPLLI YVFRLGITGA AISTVASQKR KDYRRVDGVG
TVTASIFYRY VDQLVIDRAF IHGPCPFCPL MNGLFRELLF SILFRGGMLL
GRTLSILLTM TIGTSMAARQ GPTAMAAHQI CLQVWLAVSL LADALAVSAQ
AMIASSYAIL DYKRVQKIAM FALQFVCASQ PINALAFIFD GLHYGVSDFD
YVAQATIAVG IMSSLVLLYA PSVFGLAGVW AGLTTLMGLR MASGILR

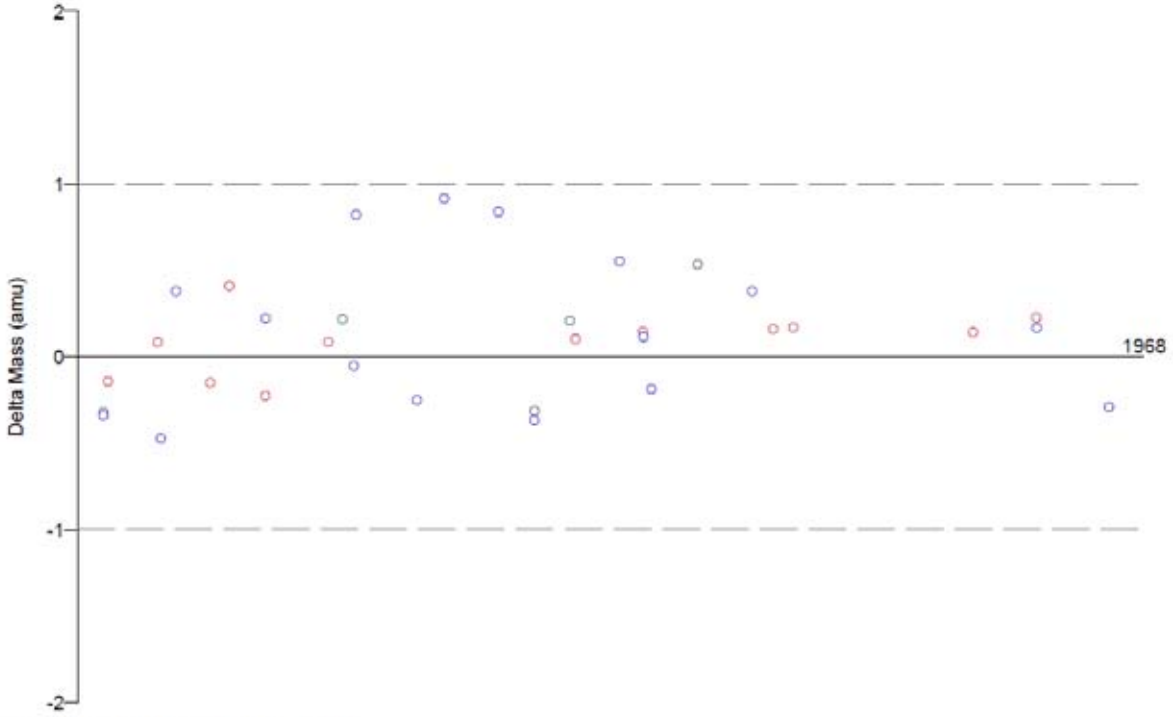
Protein Coverage:

Sequence	MH+	% Mass	AA	% AA
DVASWTSMIGACLGSGYADQALR	2372.10	2.04	193 - 215	2.10
RAHGVAVRSGLGPFELAVGNALVHMYGK	2809.50	2.42	451 - 477	2.46
Totals:	2809.50	2.42	27	2.46

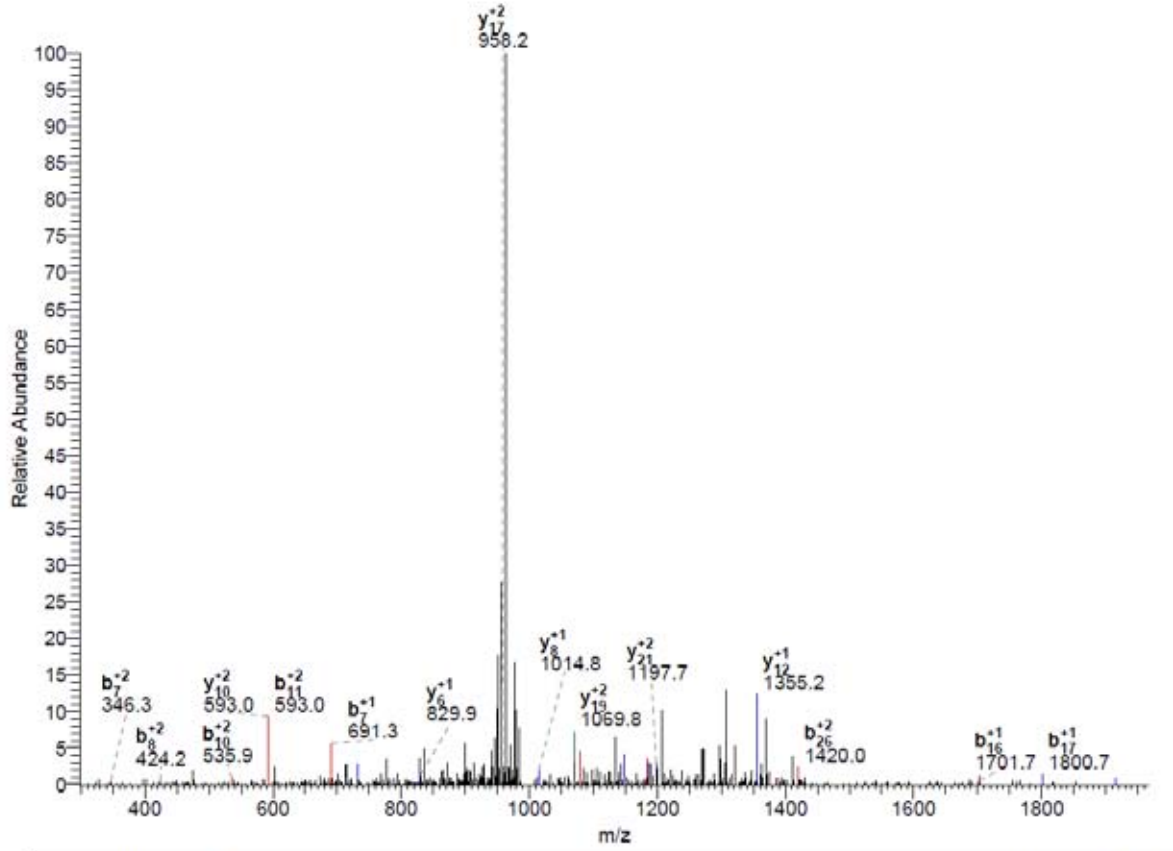
DTA for scan: 9092
Precursor ion: 995.58
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

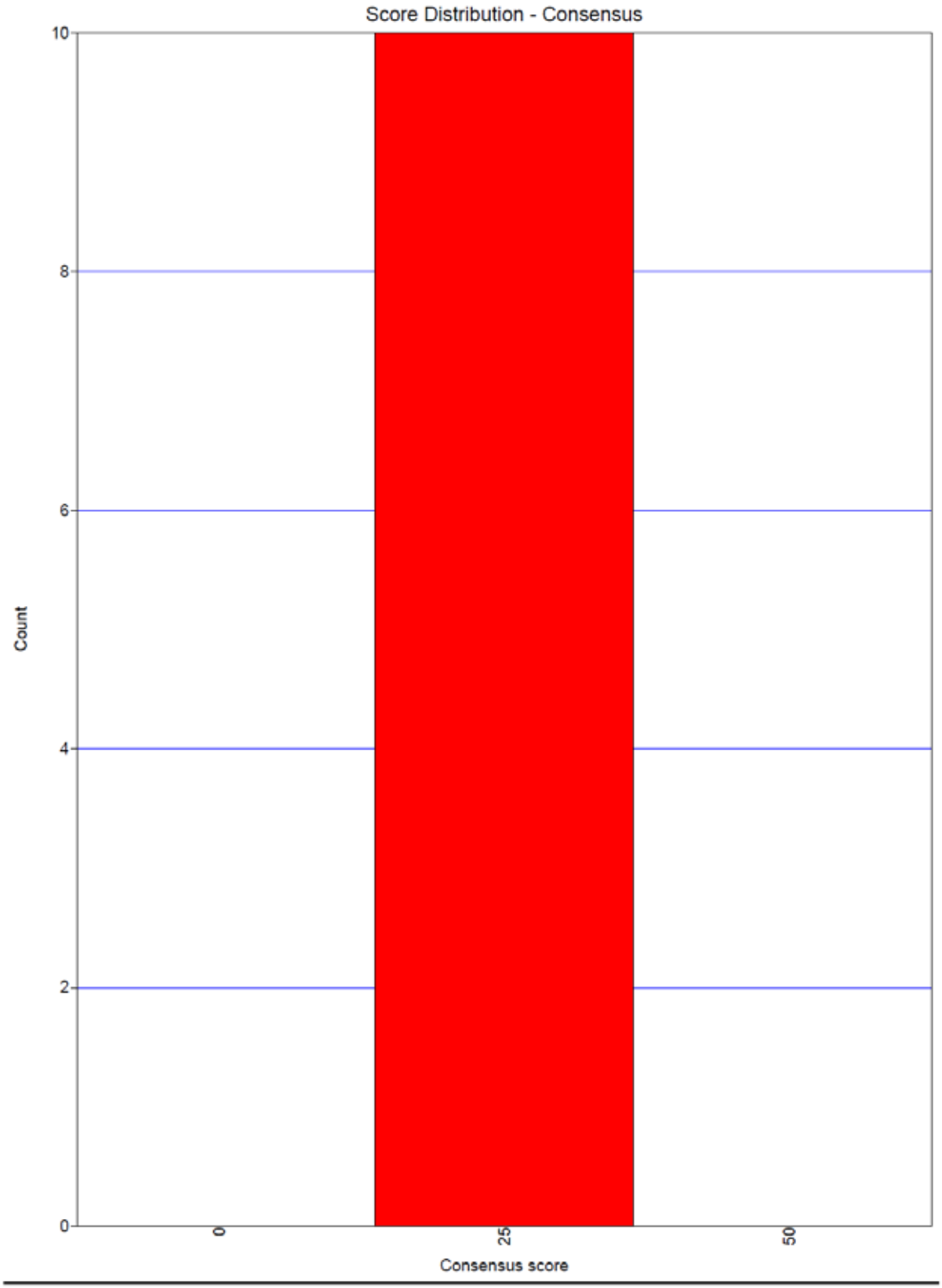
Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
R		157.11							
A		228.15				2829.33			
H		365.20				2758.29			
G		422.23				2621.23			
V		521.29				2564.21			
A		592.33				2465.14			
V		**691.40**				2394.11			
R		847.50				2295.04			
S#		**1014.50**				2138.94			
G		1071.52				1971.94			
L		**1184.60**				**1914.92**			
G		1241.63				**1801.83**			
F		**1388.69**				1744.81			
E		1517.74				1597.74			
L		1630.82				1468.70			
A		**1701.86**				**1355.62**			
V		**1800.93**				1284.58			
G		1857.95				**1185.51**			
N		1971.99				1128.49			
A		2043.03				**1014.45**			
L		2156.11				943.41			
V		2255.18				**830.33**			
H		2392.24				**731.26**			
M*		2539.28				594.20			
Y#		2782.30				447.16			
G		2839.33				204.13			
K						147.11			



#9092-9092 RT:51.19-51.19 NL: 8.75E2




Score Distribution - Consensus
10
8
6
4
2
0
Count
0
25
50
Consensus score

```
[SEQUEST]
first_database_name = C:\Xcalibur\database\rice_all.pep.fasta.hdr
second_database_name =
peptide_mass_tolerance = 1.0000
peptide_mass_units = 0              ; 0=amu, 1=mmu, 2=ppm
ion_series = 0 1 1 0.0 1.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0
fragment_ion_tolerance = 0.5000          ; for trap data leave at 1.0, for accurate mass data use values < 1.0
fragment_ion_units = 0              ; 0=amu, 1=mmu
num_output_lines = 10               ; # peptide results to show
num_results = 250                   ; # results to store
num_description_lines = 5           ; # full protein descriptions to show for top N peptides
show_fragment_ions = 0              ; 0=no, 1=yes
print_duplicate_references = 0      ; number of duplicate references reported
enzyme_info = Trypsin(KR) 1 1 KR -
max_num_differential_per_peptide = 3          ; max # of diff. mod in a peptide
diff_search_options = 15.994915 M 79.966331 STY 0.000000 M 0.000000 X 0.000000 T 0.000000 Y
term_diff_search_options = 0.000000 0.000000
nucleotide_reading_frame = 0        ; 0=protein db, 1-6, 7 = forward three, 8-reverse three, 9=all six
mass_type_parent = 1                ; 0=average masses, 1=monoisotopic masses
mass_type_fragment = 1              ; 0=average masses, 1=monoisotopic masses
normalize_xcorr = 0                 ; use normalized xcorr values in the out file
remove_precursor_peak = 0           ; 0=no, 1=yes
ion_cutoff_percentage = 0.0000          ; prelim. score cutoff % as a decimal number i.e. 0.30 for 30%
max_num_internal_cleavage_sites = 2         ; maximum value is 12
protein_mass_filter = 0 0           ; enter protein mass min & max value ( 0 for both = unused)
match_peak_count = 0                ; number of auto-detected peaks to try matching (max 5)
match_peak_allowed_error = 1        ;  number of allowed errors in matching auto-detected peaks
match_peak_tolerance = 1.0000           ; mass tolerance for matching auto-detected peaks
partial_sequence =
sequence_header_filter =
digest_mass_range = 600.0000 3500.0000

add_Cterm_peptide = 0.0000              ; added to each peptide C-terminus
add_Cterm_protein = 0.0000              ; added to each protein C-terminus
add_Nterm_peptide = 0.0000              ; added to each peptide N-terminus
add_Nterm_protein = 0.0000              ; added to each protein N-terminus
add_G_Glycine = 0.0000                  ; added to G
add_A_Alanine = 0.0000                  ; added to A
add_S_Serine = 0.0000                   ; added to S
add_P_Proline = 0.0000                  ; added to P
add_V_Valine = 0.0000                   ; added to V
add_T_Threonine = 0.0000                ; added to T
add_C_Cysteine = 0.0000                 ; added to C
add_L_Leucine = 0.0000                  ; added to L
add_I_Isoleucine = 0.0000               ; added to I
add_X_LorI = 0.0000                     ; added to X
add_N_Asparagine = 0.0000               ; added to N
add_O_Ornithine = 0.0000                ; added to O
add_B_avg_NandD = 0.0000                ; added to B
add_D_Aspartic_Acid = 0.0000            ; added to D
add_Q_Glutamine = 0.0000                ; added to Q
add_K_Lysine = 0.0000                   ; added to K
add_Z_avg_QandE = 0.0000                ; added to Z
add_E_Glutamic_Acid = 0.0000            ; added to E
add_M_Methionine = 0.0000               ; added to M
add_H_Histidine = 0.0000                ; added to H
add_F_Phenylalanine = 0.0000            ; added to F
add_R_Arginine = 0.0000                 ; added to R
add_Y_Tyrosine = 0.0000                 ; added to Y
add_W_Tryptophan = 0.0000               ; added to W
add_J_user_amino_acid = 0.0000          ; added to J
add_U_user_amino_acid = 0.0000          ; added to U
```

Output files for KK04 protein spot sequencing

KK09B04.RAW Page 1

```
SRF File: E:\data\KELVIN\09-07-21\KK09B04.srf
Database... indexed - rice_all.pep.fasta.hdr (7/23/2009)
Filter(s)... xc (± 1,2,3)=1.50,2.00,2.50 ; peptide probability<=1e-003
Mods: (M* +15.99492) (STY# +79.96633)

Reference                                                                  P      Score    Coverage    Accession
  Scan(s)       Sequence                                    MH+      z     P        XC   DeltaCn    Sp   RSp  Ions

protein CHD3-type chromatin-remodeling factor PICKLE, putative          0.0006  16.1      0.0        0
  10019       K.EYY#KAILTKNYEVLTR.R                       2084.05   2  0.0006  2.429   0.004    157.4  1   11/45
  10019       K.EY#YKAILTKNYEVLTR.R                       2084.05   2  0.0006  2.413   0.088    157.4  1   11/45

  2 of 4 peptide matches reported, 2 removed due to filtering


protein OsWAK5 - OsWAK receptor-like protein kinase, expressed          0.0004  6.1       0.0        0
  9102        K.SQILGHGGHGT#VYK.G                         1533.72   2  0.0004  2.101   0.050    258.0  3   11/39

  1 of 5 peptide matches reported, 4 removed due to filtering


protein transporter family protein, putative, expressed                 0.0002  4.1       0.0        0
  10698       K.MPEIMGSM*RSTLFPNPGSMPSVAEQQQAK.G          3265.51   3  0.0002  2.613   0.323    127.2  161 15/112

  1 of 2 peptide matches reported, 1 removed due to filtering
```

KK09B04.RAW Page 1

SRF File: E:\data\KELVIN\09-07-21\KK09B04.srf
Database... indexed - rice_all.pep.fasta.hdr (7/23/2009)
Filter(s)... xc (± 1,2,3)=1.50,2.00,2.50 ; peptide probability<=1e-003
Mods: (M* +15.99492) (STY# +79.96633)

```
Reference                                                                  P      Score    Coverage    Accession
Scan(s)       Sequence                                    MH+      z     P        XC   DeltaCn    Sp   RSp  Ions  Count

protein CHD3-type chromatin-remodeling factor PICKLE, putative          0.0006  16.1      0.0        0
10019         K.EYY#KAILTKNYEVLTR.R                       2084.05   2  0.0006  2.429   0.004    157.4  1   11/45
10019         K.EY#YKAILTKNYEVLTR.R                       2084.05   2  0.0006  2.413   0.088    157.4  1   11/45

2 of 4 peptide matches reported, 2 removed due to filtering
```

KK09B04.RAW Page 2

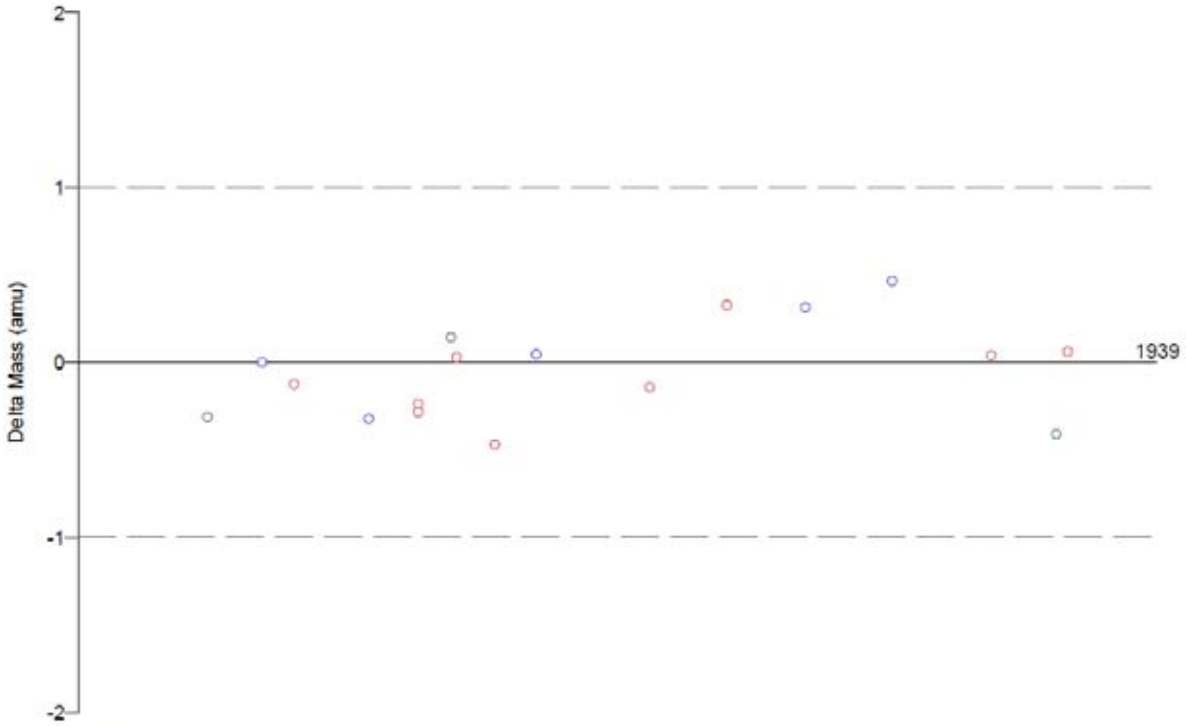
```
Reference: LOC_Os06g08480.1|13106.m00885|protein CHD3-type chromatin-remodel
Database:  C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids:  1250     Monoisotopic MW:  143984.7      pI:  5.77
```

```
Protein:

MEKILDCEET KPDASEETSS SESGSKKKPV KRYLIKWKGI SHLHCTWVSE
SEYLETAKIY PRLKTRLNNF HKQMDSTDKS DDDYSAIRPE WTTVDRILAT
RKSSTGERKY YVKWKELTYD ECTWENDSDI AVFQPQIERP NEIQSRRKKS
TDKCKSVTRE IRQYKESPKF LSGGTLHPYQ LEGLNFLRYS WYHNKRVILG
DEMGLGKTIQ SIAFLGSLFV DKLGPHLVVA PLSTLRNWER EFATWAPQMN
VVMYFGSAAS REIIRKYEFY YPKEKPKKLK KKKSSPSNED KKQSRIKFDV
LLTSYEMINM DSTVLKTIEW ECMIVDEGHR LKNKDSKLFG QLKEYHTKHR
VLLTGTPVQN NLDELFMLMH FLEGDSFGSI ADLQEEFKDI NQDKQVEKLH
GMLKPHLLRR FKKDVMKELP PKKELILRVE LTSKQKEYYK AILTKNYEVL
TRRSGGHVSL INVVMELRKL CCHAFMTDEP EEPANSEEAL RRLLESSGKM
ELLDKMMVKL KEQGHRVLIY SQFQHMLDLL EDYLSYRKWS YERIDGKIGG
AERQIRIDRF NAKNSTRFCF LLSTRAGGLG INLATADTVI IYDSDWNPHA
DLQAMARAHR LGQTSKVMIY RLVSRGTIEE RMMQLTKKKM VLEHLVVGRL
TKGTNIVQEE LDDIIRHGSK ELFDDENDEA GKSCQIHYDD AAIDRLLDRD
QADGEEPVED EEEDEFLKGF KVANFEYIDE AKALAAKEEE ARKKAEAEAA
NSDRANFWDK LLKDRYDVQK VEEHTTMGKG KRSRKQMAAA DEDDITGLMD
MSSEDDDYSY DDDVSDNDTS LQSGLAGRRG PYSKKKQRNV DSLPFMEGEG
RALRVYGFNQ IQRTQFLQTL MRYGFQNYDW KEFTPRLKGK SVEEIQRYAE
LVMIHLLEDI NDSGYYADGV PKEMRTDETL VRLANISLVE EKVAAMEQGK
ITKLFPSYLL YEFPSLVGGR VWKAEQDLLL LKALIKHGYA RWQYISDDRD
NGIFEAARQE LRLPTANELI SSHSNNETNG NLESTQEGQS NPTSMIHYRD
TQRKIVEFIR KRYHLLERCL NLEYAVIKTK TPVPDDLAEQ DFPGGHRPAV
PDYSEMLREL PVLEPISKEV APEGTTDQSQ VSHLYNKMCF VLEDSAVPAL
NSHFGDKAAS SGLANSLHKF EAVCEDVSRI LRSHENGTTP KEEVMLDASS
KETTSPKDPA TEVPSSASKE ATPPVQDPVI EAVKEEPPTV KAEDKMEIDS

Protein Coverage:

Sequence                          MH+      % Mass        AA        % AA
EYYKAILTKNYEVLTR                2004.08     1.39     437 - 452     1.28
SHENGTTPK                        970.46     0.67    1183 - 1191    0.72
                    Totals:     2004.08     1.39            16     1.28
```

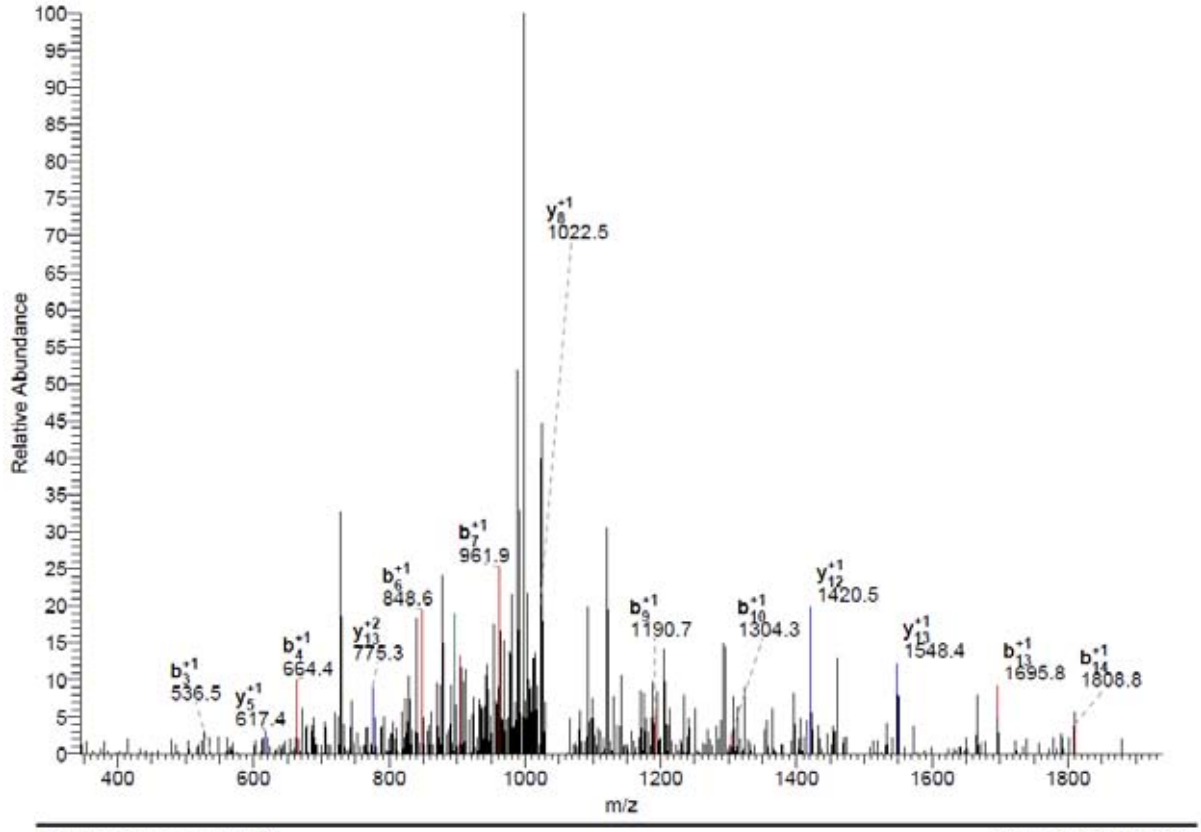
DTA for scan: 10019
Precursor ion: 1042.23
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
E		130.05							
Y		293.11				1955.00			
Y#		**536.14**				**1791.94**			
K		**664.24**				**1548.91**			
A		735.27				**1420.82**			
I		**848.36**				1349.78			
L		**961.44**				1236.69			
T		1062.49				1123.61			
K		**1190.59**				**1022.56**			
N		**1304.63**				894.47			
Y		1467.69				780.43			
E		1596.73				**617.36**			
V		**1695.80**				488.32			
L		**1808.89**				389.25			
T		1909.93				276.17			
R						175.12			



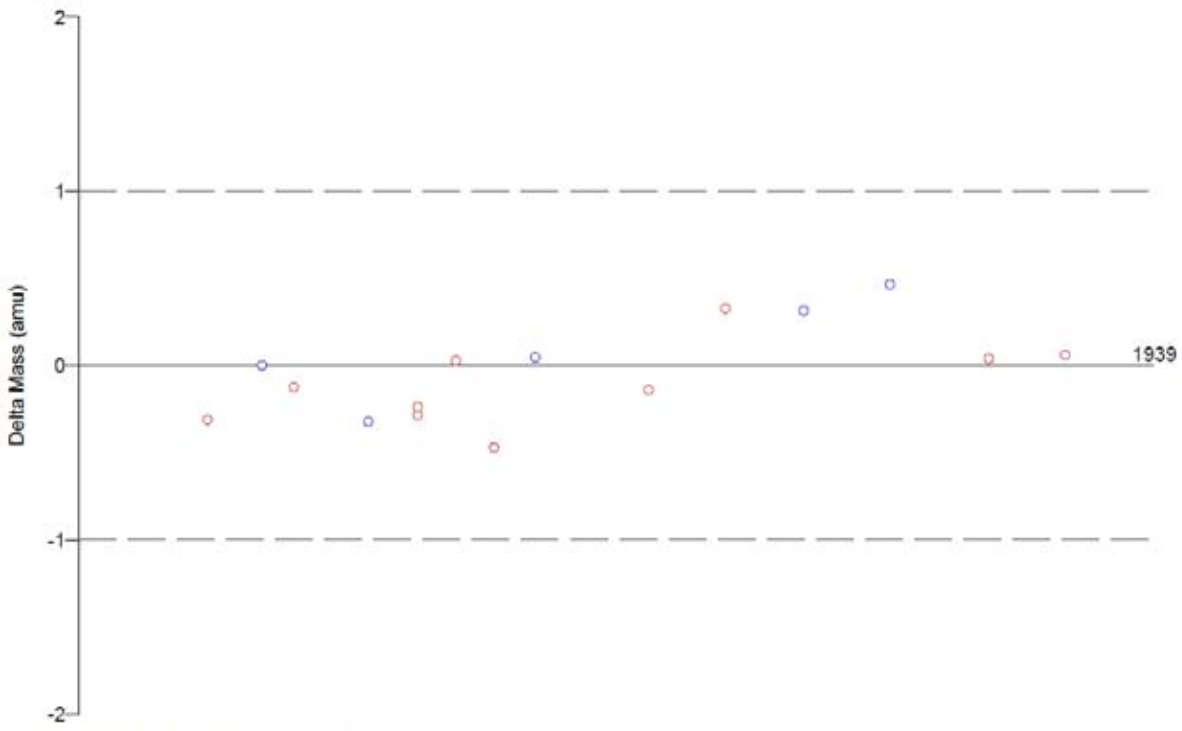
#10019-10019 RT:56.26-56.26 NL: 1.58E2



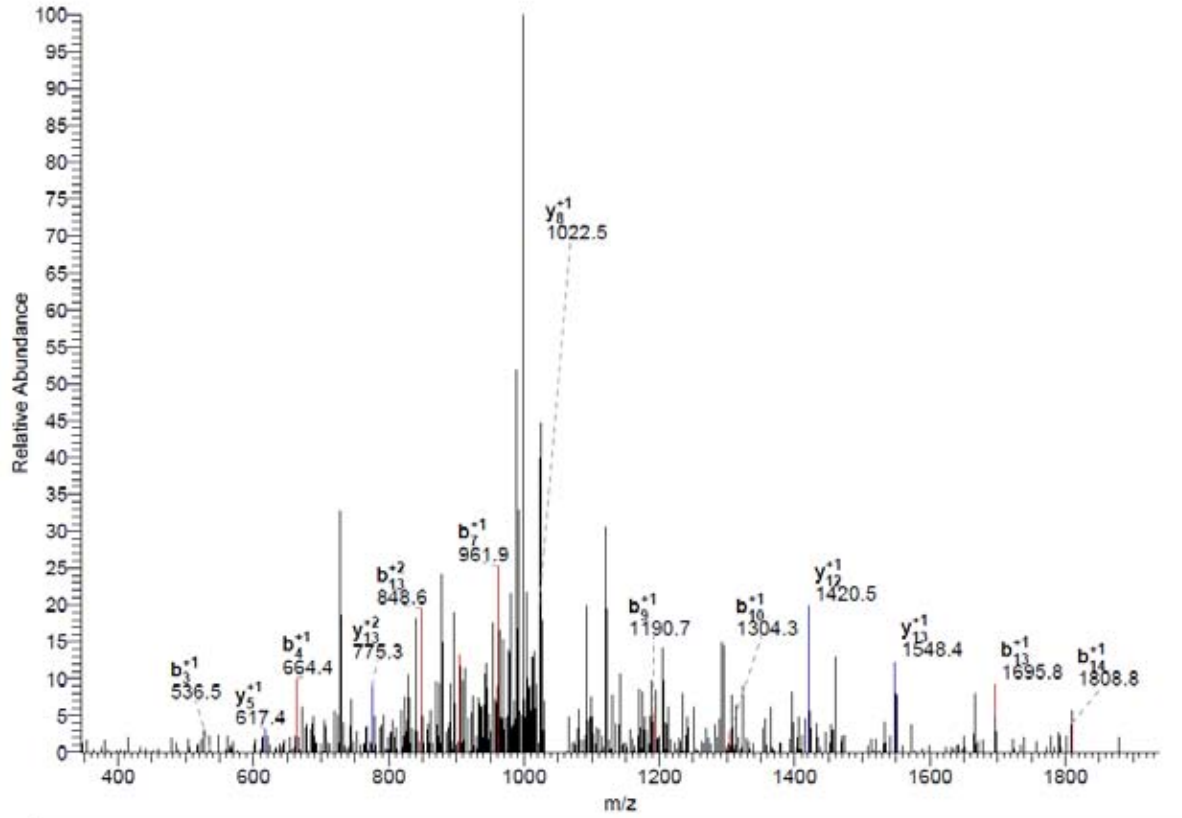
DTA for scan: 10019
Precursor ion: 1042.23
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
E		130.05							
Y#		373.08				1955.00			
Y		**536.14**				1711.97			
K		**664.24**				**1548.91**			
A		735.27				**1420.82**			
I		**848.36**				1349.78			
L		**961.44**				1236.69			
T		1062.49				1123.61			
K		**1190.59**				**1022.56**			
N		**1304.63**				894.47			
Y		1467.69				780.43			
E		1596.73				**617.36**			
V		**1695.80**				488.32			
L		**1808.89**				389.25			
T		1909.93				276.17			
R						175.12			



#10019-10019 RT:56.26-56.26 NL: 1.58E2



Reference Scan(s)	Sequence	MH+	z	P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein OsWAK5 - OsWAK receptor-like protein kinase, expressed				0.0004	6.1	0.0	0			
9102	K.SQILGHGGHGT#VYK.G	1533.72	2	0.0004	2.101	0.050	258.0	3	11/39	1

1 of 5 peptide matches reported, 4 removed due to filtering

Reference: LOC_Os01g26174.1|13101.m02771|protein OsWAK5 - OsWAK receptor-lik
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 725 Monoisotopic MW: 80421.6 pI: 5.60

Protein:

MASAILLSIA IMAQLSSISA QPAPGCQSHC GDMEIPYPFG IGTECAIEPG
FVIYCNKTAD GSMKPFLINV EVLNISLLHG QTRALNALST YCYNDVTKSM
ESSRWSLDPS TWPYRFSNLH NKFVVIGCNT LSYIYNGEYT TACASVCAKA
PTNDSCDGVG CCQNNIAKGL NSYNVTFPTV YNDSSNLQSN PCSYAALVET
DTFRFKTEYV TTMKFNETYN GQQPVVLDWA IGKVGCKEAN MTSYACRSKH
SECVDSINGP GYLCNCTLGY HGNPYITDGC IDVNECEQNQ SPCPKGATCR
NTEGWYHCSC PVGRKLAKET NTCNPDISLI IGVSIGSIVL VIIIFFVRII
FERRKLTDVK KKYIQEHGGL LLFEKMKSDQ GLAFKVFTQA ELEQATNKFE
KSQILGHGGH GTVYKGITKD NITVAIKKCA LIDDRHKKEF GKEMLILSQI
NHKNIVKLLG CCLEVDVPML VYEFIPNGTL FDLIHGKNRT LHIPFSSLLR
IVNRAAEGLA FLHSYANPPI LHGDVKTSNI LLDENYMAKV SDFGASILAP
NDEAQFVTMV QGTCGYLDPE YLQTCQLTEK SDVYSFGVVI LEILTGQMPL
KLEGSELQKS LSSSFLLAMK ENNLEAMLDS QIKGHESMEL LSGLAELAKQ
CLDMCSENRP SMKDVAEEIS RLRKLSKHPW IQRDSETEGY LSGPSTSNFE
IEQSTEYTRK DEQMPINPST SYFIR

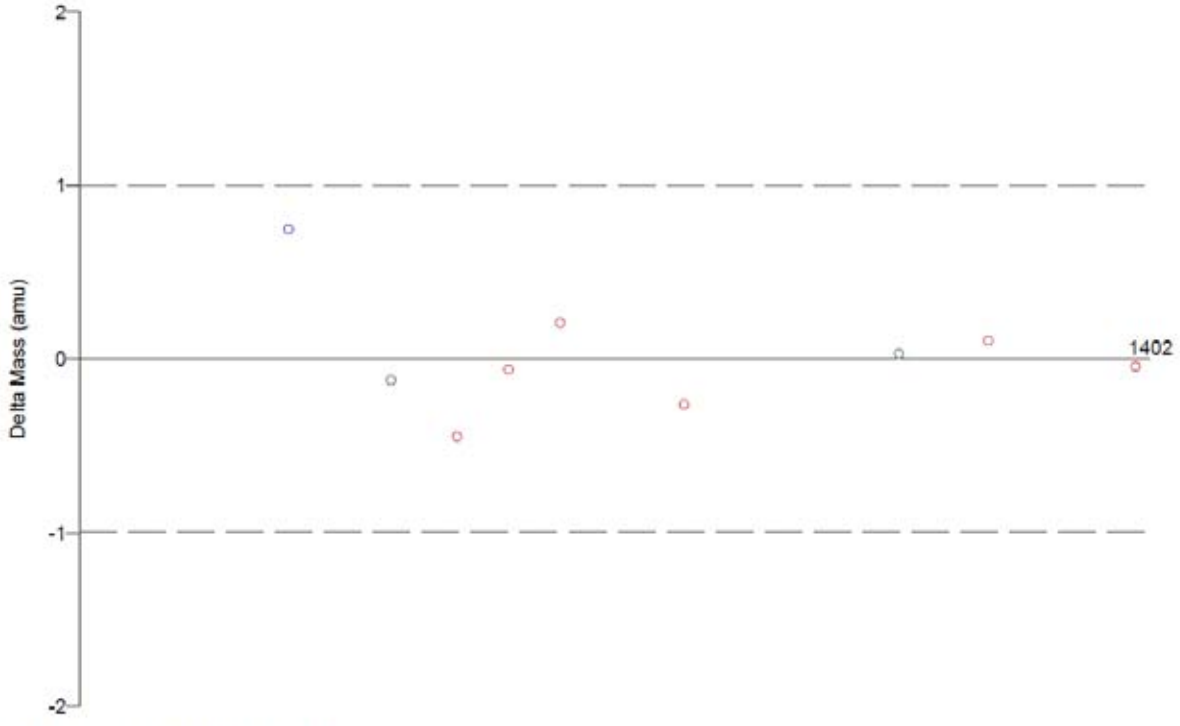
Protein Coverage:

Sequence		MH+	% Mass	AA	% AA
TADGSMK		709.32	0.88	58 - 64	0.97
SQILGHGGHGTVYK		1453.75	1.81	402 - 415	1.93
SQILGHGGHGTVYKGITK		1853.00	2.30	402 - 419	2.48
TLHIPPSSLLR		1283.75	1.60	490 - 500	1.52
	Totals:	1453.75	1.81	14	1.93

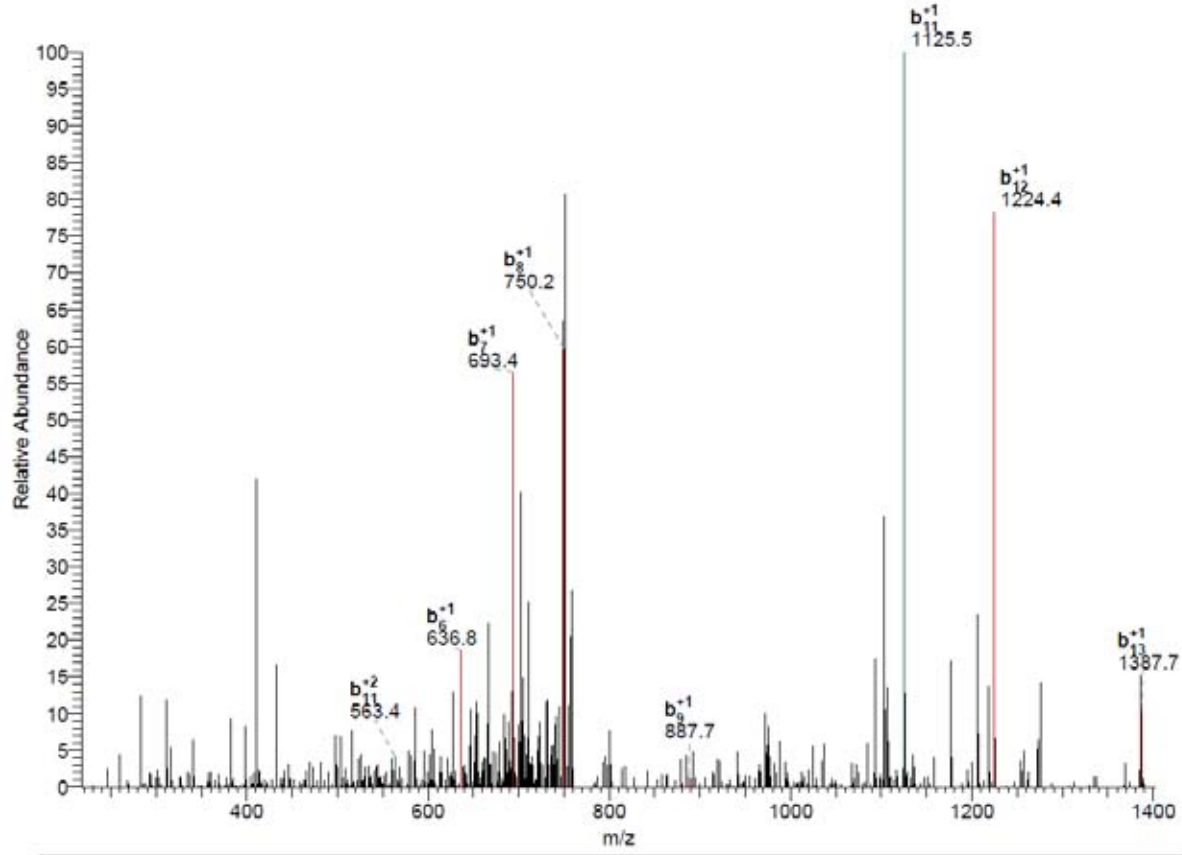
DTA for scan: 9102
Precursor ion: 767.77
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
S		88.04							
Q		216.10				1446.69			
I		329.18				1318.63			
L		442.27				1205.55			
G		499.29				1092.46			
H		**636.35**				1035.44			
G		**693.37**				898.38			
G		**750.39**				841.36			
H		**887.45**				784.34			
G		944.47				647.28			
T#		**1125.48**				590.26			
V		**1224.55**				409.24			
Y		**1387.62**				310.18			
K						147.11			



#9102-9102 RT:51.12-51.12 NL: 1.92E2



Reference Scan(s)	Sequence	MH+	z	P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein transporter family protein, putative, expressed				0.0002	4.1	0.0	0			
10698	K.MPEIMGSM*RSTLPPNPGSMPSVAEQQQAK.G	3265.51	3	0.0002	2.613	0.323	127.2	161	15/112	1

1 of 2 peptide matches reported, 1 removed due to filtering

Reference: LOC_Os10g39440.1|13110.m03612|protein transporter family protein,
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 740 Monoisotopic MW: 78849.1 pI: 4.78

Protein:

MAGAVLVAIA ASIGNLLQGW DNATIAGAVL YIKKEFNLQS EPLIEGLIVA
MSLIGATIIT TFSGAVADSF GRRPMLIASA VLYFVSGLVM LWAPNVYVLL
LARLIDGFGI GLAVTLVPLY ISETAPTDIR GLLNTLPQFS GSGGMFLSYC
MVFGMSLMPQ PDWRIMLGVL SIPSLIYFAL TIFYLPESPR WLVSKGRMAE
AKRVLQQLRG REDVSGEMAL LVEGLGVGKD TKIEEYIIGP DDELADEGLA
PDPEKIKLYG PEEGLSWVAR PVWGQSALGS ALGLISRHGS MVSQGKPLVD
PVVTLPGSVH EKMPEIMGSM RSTLPPNPGS MPSVAEQQQA KGDWDAESQR
EGEDYGSDHG GDDIEDSLQS PLISRQATSV EGKEIAAPHG SIMGAVGRSS
SLMQGGEAVS SMGIGGGWQL AWKWTEREGA DGEKEGGFQR IYLHEEGVTG
DRRGSILSLP GGDVPPGGEF VQAAALVSQP ALYSKELMEQ RLAGPAMVHP
SQAVAKGPKW ADLFEPGVKH ALFVGIGIQI LQQPAGINGV LYYTPQILEQ
AGVGVLLANI GLSSSSASIL ISGLTTLLML PSIGIAMRLM DMSGRRFLLL
ATIPILIVAL AILILVNILD VGTMVHASLS TVSVILYFCF FVMGFGPIPN
ILCAEIFPTT VRGICIAICA LTFWIGDIIV TYTLPVMLNA IGLAGVFGIY
AVVCILAFLF VFMKVPETKG MPLEVITEFF SVGAKQAKED

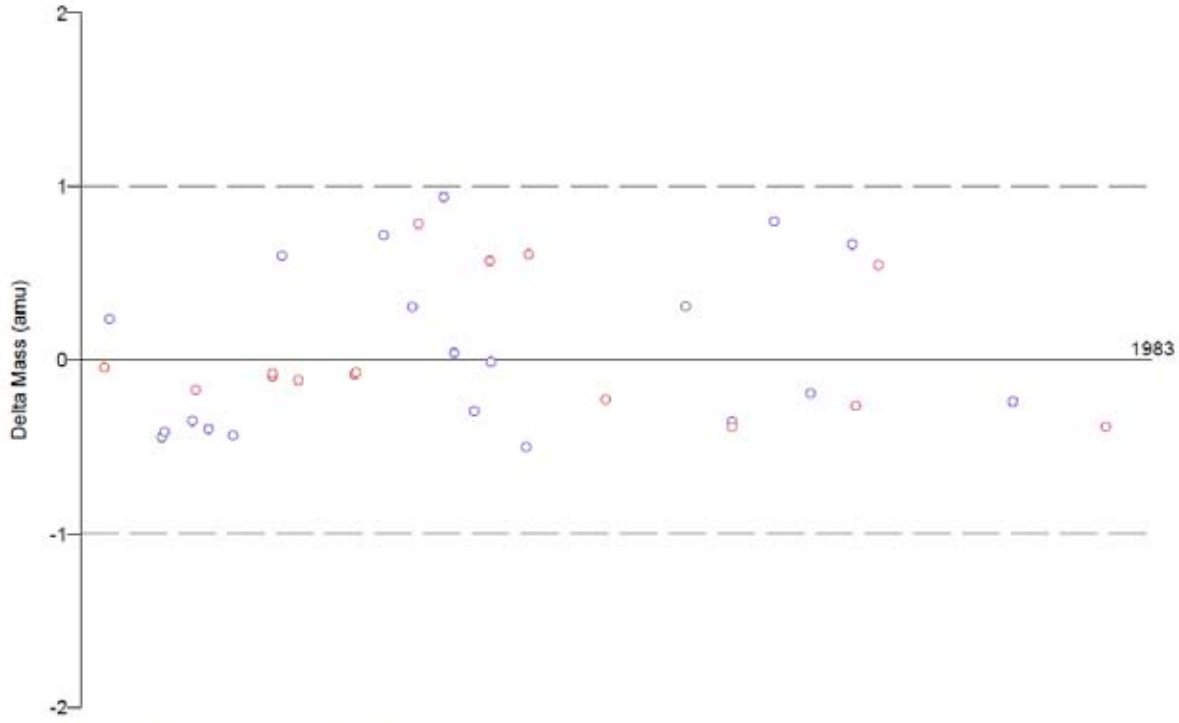
Protein Coverage:

Sequence	MH+	% Mass	AA	% AA
MPEIMGSMRSTLPPNPGSMPSVAEQQQAK	3249.52	4.12	313 - 341	3.92
Totals:	3249.52	4.12	29	3.92

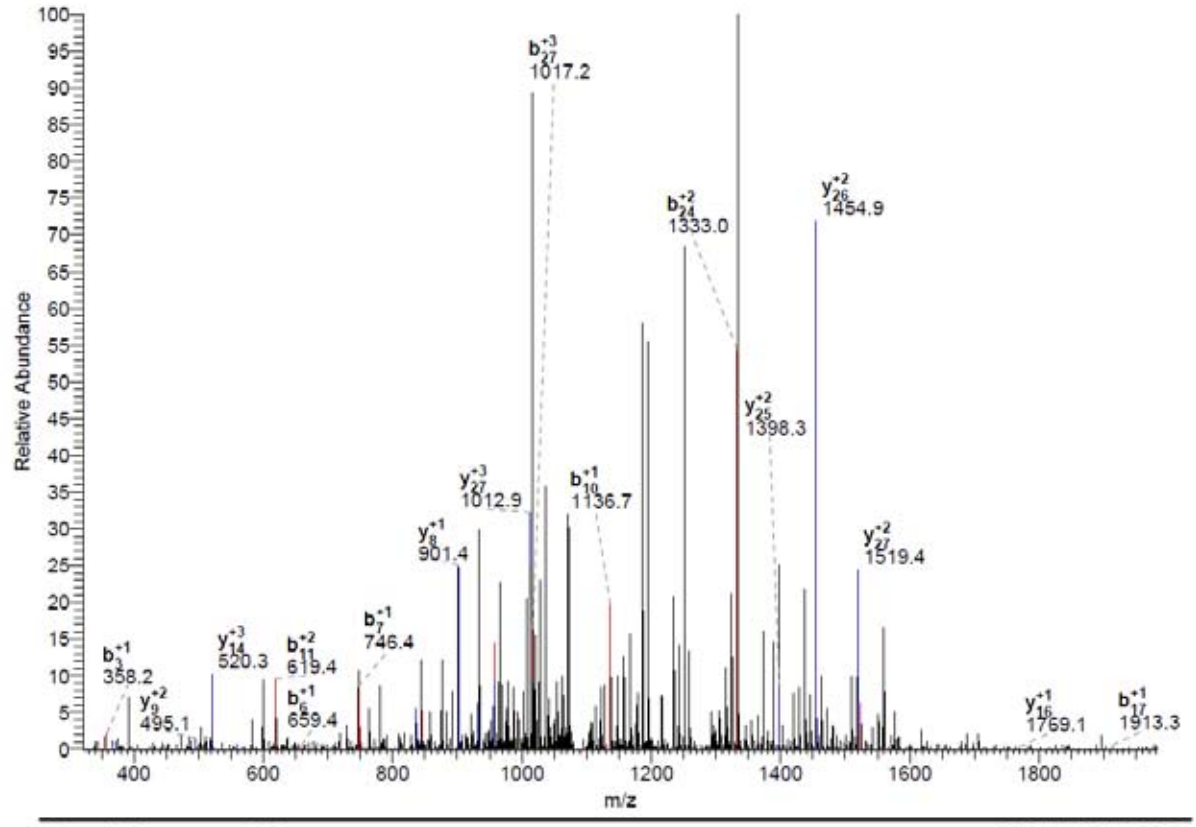
DTA for scan: 10698
Precursor ion: 1089.21
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

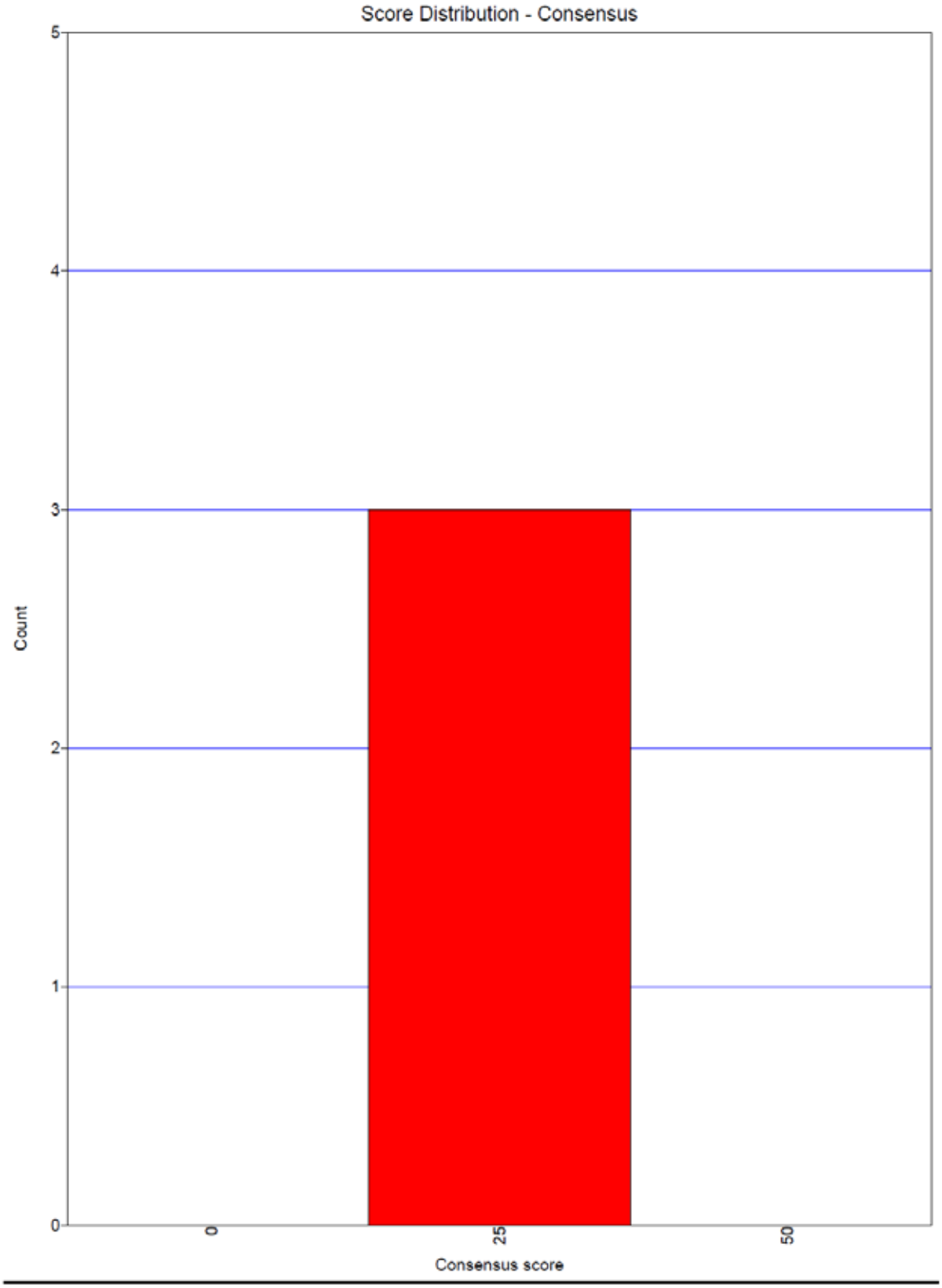
AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
M		132.05							
P		229.10				3134.47			
E		**358.14**				3037.42			
I		471.23				2908.38			
M		602.27				2795.29			
G		**659.29**				2664.25			
S		**746.32**				2607.23			
M*		893.36				2520.20			
R		1049.46				2373.17			
S		**1136.49**				2217.06			
T		1237.54				2130.03			
L		1350.62				2028.98			
F		1497.69				1915.90			
P		1594.74				**1768.83**			
N		1708.79				1671.78			
P		1855.85				1557.74			
G		**1912.88**				1410.67			
S		1999.91				1353.65			
M		2130.95				1266.61			
P		2278.02				1135.57			
S		2365.05				988.51			
V		2464.12				**901.47**			
A		2535.15				802.41			
E		2664.20				731.37			
Q		2792.26				602.33			
Q		2920.31				474.27			
Q		3048.37				346.21			
A		3119.41				218.15			
K						147.11			



#10698-10698 RT:60.08-60.08 NL: 6.52E2



BioworksBrowser rev. 3.3 25/08/2009 13:06:03


Score Distribution - Consensus
5
4
3
2
1
0
Count
0
25
50
Consensus score

```
[SEQUEST]
first_database_name = C:\Xcalibur\database\rice_all.pep.fasta.hdr
second_database_name =
peptide_mass_tolerance = 1.0000
peptide_mass_units = 0              ; 0=amu, 1=mmu, 2=ppm
ion_series = 0 1 1 0.0 1.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0
fragment_ion_tolerance = 0.5000          ; for trap data leave at 1.0, for accurate mass data use values < 1.0
fragment_ion_units = 0              ; 0=amu, 1=mmu
num_output_lines = 10               ; # peptide results to show
num_results = 250                   ; # results to store
num_description_lines = 5           ; # full protein descriptions to show for top N peptides
show_fragment_ions = 0              ; 0=no, 1=yes
print_duplicate_references = 0      ; number of duplicate references reported
enzyme_info = Trypsin(KR) 1 1 KR -
max_num_differential_per_peptide = 3          ; max # of diff. mod in a peptide
diff_search_options = 15.994915 M 79.966331 STY 0.000000 M 0.000000 X 0.000000 T 0.000000 Y
term_diff_search_options = 0.000000 0.000000
nucleotide_reading_frame = 0        ; 0=protein db, 1-6, 7 = forward three, 8-reverse three, 9=all six
mass_type_parent = 1                ; 0=average masses, 1=monoisotopic masses
mass_type_fragment = 1              ; 0=average masses, 1=monoisotopic masses
normalize_xcorr = 0                 ; use normalized xcorr values in the out file
remove_precursor_peak = 0           ; 0=no, 1=yes
ion_cutoff_percentage = 0.0000         ; prelim. score cutoff % as a decimal number i.e. 0.30 for 30%
max_num_internal_cleavage_sites = 2          ; maximum value is 12
protein_mass_filter = 0 0           ; enter protein mass min & max value ( 0 for both = unused)
match_peak_count = 0                ; number of auto-detected peaks to try matching (max 5)
match_peak_allowed_error = 1        ;  number of allowed errors in matching auto-detected peaks
match_peak_tolerance = 1.0000          ; mass tolerance for matching auto-detected peaks
partial_sequence =
sequence_header_filter =
digest_mass_range = 600.0000 3500.0000

add_Cterm_peptide = 0.0000             ; added to each peptide C-terminus
add_Cterm_protein = 0.0000             ; added to each protein C-terminus
add_Nterm_peptide = 0.0000             ; added to each peptide N-terminus
add_Nterm_protein = 0.0000             ; added to each protein N-terminus
add_G_Glycine = 0.0000                 ; added to G
add_A_Alanine = 0.0000                 ; added to A
add_S_Serine = 0.0000                  ; added to S
add_P_Proline = 0.0000                 ; added to P
add_V_Valine = 0.0000                  ; added to V
add_T_Threonine = 0.0000               ; added to T
add_C_Cysteine = 0.0000                ; added to C
add_L_Leucine = 0.0000                 ; added to L
add_I_Isoleucine = 0.0000              ; added to I
add_X_LorI = 0.0000                    ; added to X
add_N_Asparagine = 0.0000              ; added to N
add_O_Ornithine = 0.0000               ; added to O
add_B_avg_NandD = 0.0000               ; added to B
add_D_Aspartic_Acid = 0.0000           ; added to D
add_Q_Glutamine = 0.0000               ; added to Q
add_K_Lysine = 0.0000                  ; added to K
add_Z_avg_QandE = 0.0000               ; added to Z
add_E_Glutamic_Acid = 0.0000           ; added to E
add_M_Methionine = 0.0000              ; added to M
add_H_Histidine = 0.0000               ; added to H
add_F_Phenylalanine = 0.0000           ; added to F
add_R_Arginine = 0.0000                ; added to R
add_Y_Tyrosine = 0.0000                ; added to Y
add_W_Tryptophan = 0.0000              ; added to W
add_J_user_amino_acid = 0.0000         ; added to J
add_U_user_amino_acid = 0.0000         ; added to U
```

Output files for KK05 protein spot sequencing

KK09B05.RAW Page 1

```
SRF File: E:\data\KELVIN\09-07-21\KK09B05.srf
Database... indexed - rice_all.pep.fasta.hdr (7/23/2009)
Filter(s)... xc (± 1,2,3)=1.50,2.00,2.50 ; peptide probability<=1e-003
Mods: (M* +15.99492) (STY# +79.96633)

Reference                                                          P      Score    Coverage   Accession
  Scan(s)       Sequence                                MH+     z    P       XC  DeltaCn   Sp   RSp  Ions

protein OsMan04 - Endo-Beta-Mannanase, expressed                          0.0007  10.1     0.0       0
  9291        R.LPRTLLIRFISLSR.S                     1685.06   3  0.0007  2.586  0.087   300.5  16  17/52

  1 of 1 peptide matches reported, 0 removed due to filtering


protein RNA polymerase IV largest subunit, putative                       0.001   10.1     0.0       0
  9418        K.VCLEENNQITWT#DKPK.A                  1997.90   2  0.001   2.245  0.058   170.1  13  10/45

  1 of 7 peptide matches reported, 6 removed due to filtering


protein AP2 domain containing protein, expressed                          0.0009  10.1     0.0       0
  11275       R.MMEADPT#NEGLWKGDK.D                  1901.78   2  0.0009  2.147  0.035   90.1   19  9/45

  1 of 1 peptide matches reported, 0 removed due to filtering


protein protein kinase, putative, expressed                               0.0009  2.1      0.0       0
  8334        K.SSVAAKNT#EPPKR.I                     1464.72   2  0.0009  2.001  0.126   118.8  4   11/36

  1 of 3 peptide matches reported, 2 removed due to filtering
```

KK09B05.RAW Page 1

```
SRF File: E:\data\KELVIN\09-07-21\KK09B05.srf
Database... indexed - rice_all.pep.fasta.hdr (7/23/2009)
Filter(s)... xc (± 1,2,3)=1.50,2.00,2.50 ; peptide probability<=1e-003
Mods: (M* +15.99492) (STY# +79.96633)

Reference                                                          P      Score    Coverage   Accession
Scan(s)       Sequence                                MH+     z    P       XC  DeltaCn   Sp   RSp  Ions  Count

protein OsMan04 - Endo-Beta-Mannanase, expressed                          0.0007  10.1     0.0       0
9291        R.LPRTLLIRFISLSR.S                     1685.06   3  0.0007  2.586  0.087   300.5  16  17/52

1 of 1 peptide matches reported, 0 removed due to filtering
```

KK09B05.RAW Page 2

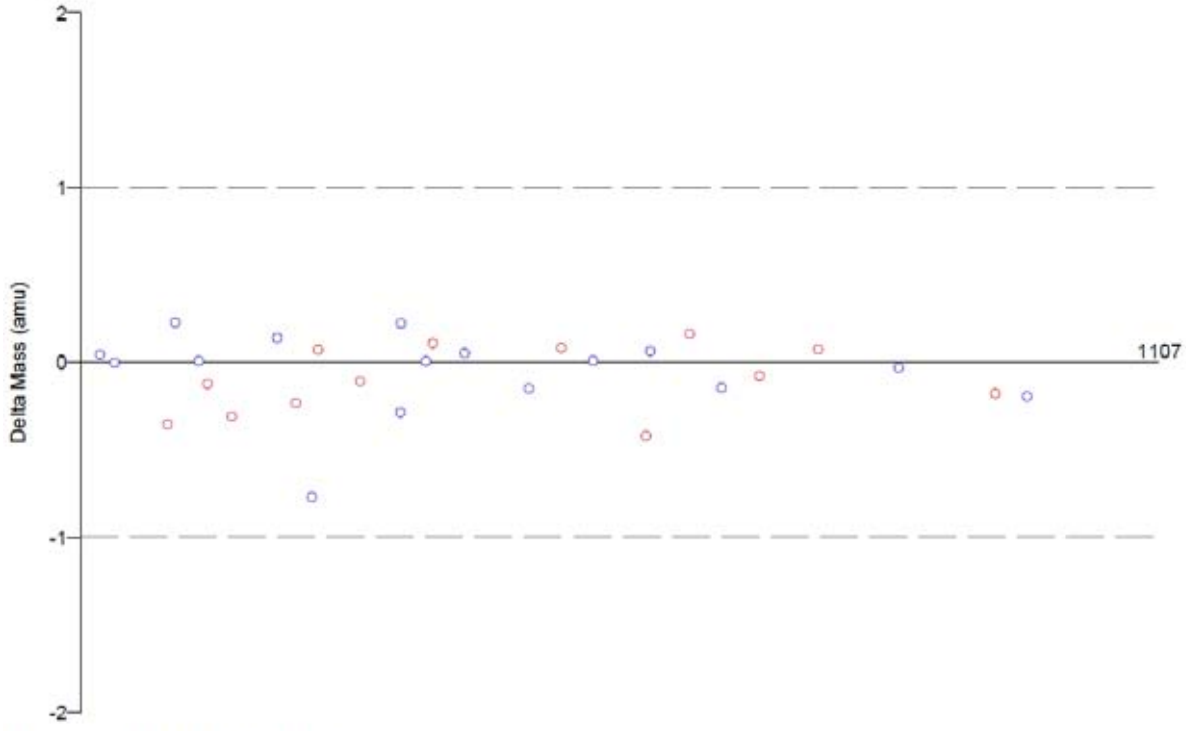
```
Reference: LOC_Os03g61270.1|13103.m06747|protein OsMan04 - Endo-Beta-Mannana
Database:  C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids:  269      Monoisotopic MW:  29852.0       pI:  5.73
```

```
Protein:

MNEPRCDADP TGGMVQAWVE EMAPYVKRVD GGRHLVTAGL EGFYGDGEHE
SKELNPWGIY YGTNYVATHR AAGVDFATIH LYPDVWLWGS TADEQAAPFR
NWTRSHVHDT AAPLGKPLLV TEYGKPLWKG GGANKTQRNY FLDVVLDAIY
ASASRGGPLV GGAFWQLLLD DDVVAGMDDL RDGYEIILAE DSRAASIIGE
HSEQLASLNG QDAEALRRRR RRPASSHRKT RLGSGGDSDA LRLPRTLLIR
FISLSRSISS FIQDNFVLF

Protein Coverage:

Sequence                                MH+       % Mass        AA        % AA
LPRTLLIRFISLSR                       1685.06        5.64   243 - 256       5.20
                        Totals:      1685.06        5.64          14       5.20
```

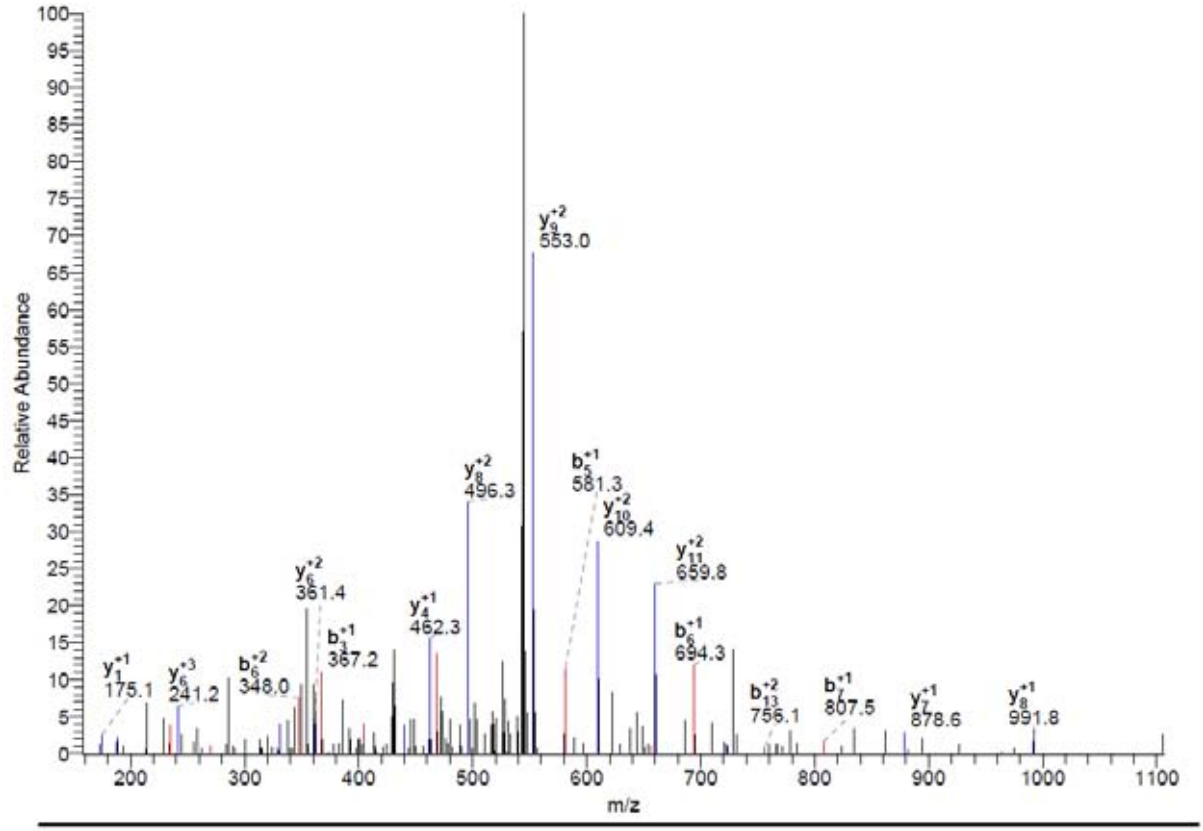
DTA for scan: 9291
Precursor ion: 562.63
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
L		114.09							
P		211.14				1571.97			
R		367.25				1474.92			
T		468.29				1318.82			
L		581.38				1217.77			
L		694.46				1104.69			
I		807.55				991.60			
R		963.65				878.52			
P		1110.71				722.42			
I		1223.80				575.35			
S		1310.83				462.27			
L		1423.91				375.24			
S		1510.95				262.15			
R						175.12			



#9291-9291 RT:52.20-52.20 NL: 1.09E2



BioworksBrowser rev. 3.3 25/08/2009 12:50:13

Reference Scan(s)	Sequence	MH+	z	P P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein RNA polymerase IV largest subunit, putative				0.001	10.1	0.0	0			
9418	K.VCLEENNQITWT#DKPK.A	1997.90	2	0.001	2.245	0.058	170.1	13	10/45	

1 of 7 peptide matches reported, 6 removed due to filtering

Reference: LOC_Os01g73430.1|13101.m08012|protein RNA polymerase IV largest s
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 1570 Monoisotopic MW: 174461.4 pI: 7.60

Protein:

MEGHPDPTSA ATAMIPEASI RRINLSITSN EEILKAQPVN ELEKPIPITH
QSQLLNNPYL GLPLQVGSCQ SCGSNAIEEC EGHFRFIELP MPIFHPSHVT
ELSQILNLIC LRCLKIKNRK KSTLKGSKFT SCSHCQELPP LCVAEVKKSN
GARGLELRAP IKKELEEGFW SFLDQFGSCT RGTSHCRPLL PEEVQNIIKK
IPEETRRWLS VRGYIPQDGF ILSYLCVPPN CLRVSNVLDG NTFSCSGTST
NLLRKALRKI QQIRGSRIGS SNIQVDQVAD DLQVDVANYI NLGGTTKGHG
DDTFTSQPTA MQWKQKMKTL FISKSSSFSS RGVITGDPYI GLNVVGVPEE
VAKRMSVEEK VTDHNIAQLQ DMMNKGLCLT YTDANSITYS LDAGKDNPNK
KHTILKVGEI VNRRVFDGDI VFLNRPPSTD KHSVEAFVVQ VHNDHTIKIN
PLICDPLGAD FDGDCVQIFY PRSLSARAEA KELYTVDKQL VSSHNGKLNF
QFKNDFSLAL KIMCGREYSE REANQITNAM FSSGMYPQKP LIGGPYWTFP
QILSTTKSNA ITLADHLDRE SVGALATGTT ISSILSTKGP REATEPLNLL
QPLLMESLLI DCFSINLGDF TVPSPILEAI QNNPLELNKY REPIMDFITH
SSAIGLLVDP KSDSNMNKVV EQLGFLGPQL QHNGRLYSSR LVEDCLSKSL
HRCCGSTNCC NPLEEYGTVR SSIYHGLNPY EALLHSICER EKIMRASKGL
VEPGSLPKNM MSRLRDVTAC YDGSIRTSSG NLVLQFGSRD ASNCVTPGDP
VGILAATAVA NAAYKAVLAP NQNNIISWDS MKEVLLTRAS TKADANHRKV
ILYLNQCSCE NECMERALTI RACLRRIKLE DCTTEISIKY QQQATQAAHH
LVGHIHLDKK QLNQIETIMD SVLHKCQETF RNNIKKKGSM REILKTVTFI
SSTSLCDQHT DDDKKFQVSC LQFFLPGSIT KNISESTERV IDFMTNAIFP
IILDTVIKGD PRVEEANLVR IEPESTFWVQ SSGAEQKGEA ALEITVEEAA
AAESGNAWGV AMNACIPVMD LIDTTRSMPY DIQQVRQVFG ISSAFEKVTQ
VQLFPFLLLA EIQYLSKSVG MITKSVLQEH LTTVASSMTC TGDLHGFNNS
GYKATCQSLK VQAPFMEATL SRSIQCFEKA AAKAYSDQLG NVVSACSWGN
NAEIGTGSAF EILWNDENMS SSKSILGGYG LYDFLEAVET TGATKDKAIV
PHNYCLYDVD CIPEDKVCLE RNNQITWTDK PKAEFLMESE GRRAGMHSTG
QKHPRKPNWH EGNTKSSPNS TAVEFTGQVF QRRQLKTKSN WNSDATQQDD
KPSWYSSNSA GTQNFTIAGS SRPGEWNRKN NNRGQGGGRE VWKSEGPHRG
GSSSNRNQGG GRAVWKSEAS HRGSGNNRNR GGGRAVWKSE ASRRGGSMRQ
VASCAFTPVE QQIFEQIEPI TKNVKRIIRE SRDGIKLPPD DEKFIVTNVL
MYHPERKKKI AGNGNYITVD RHQVFHGSRC LYVMSSDGSR KDFSYKKCLE
NYIRAQYPDA ADSPCRKYPK

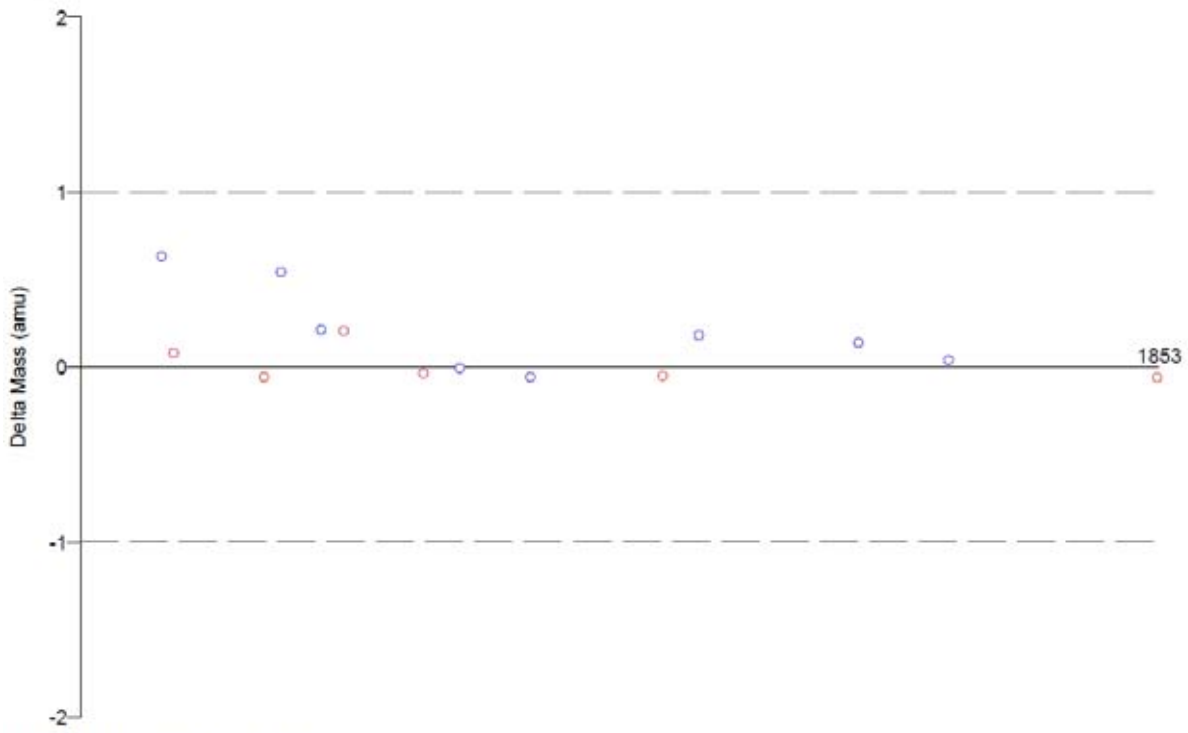
Protein Coverage:

Sequence	MH+	% Mass	AA	% AA
EYSEREANQITNAMFSSGMYPQK	2681.20	1.54	517 - 539	1.46
IMRASKGLVEPGSLFK	1732.98	0.99	743 - 758	1.02
VCLEENNQITWTDKPK	1917.94	1.10	1267 - 1282	1.02
SEGPHRGGSSSNR	1327.61	0.76	1394 - 1406	0.83
CLYVMSSDGSRKDFSYK	1985.91	1.14	1520 - 1546	1.08
Totals:	1917.94	1.10	16	1.02

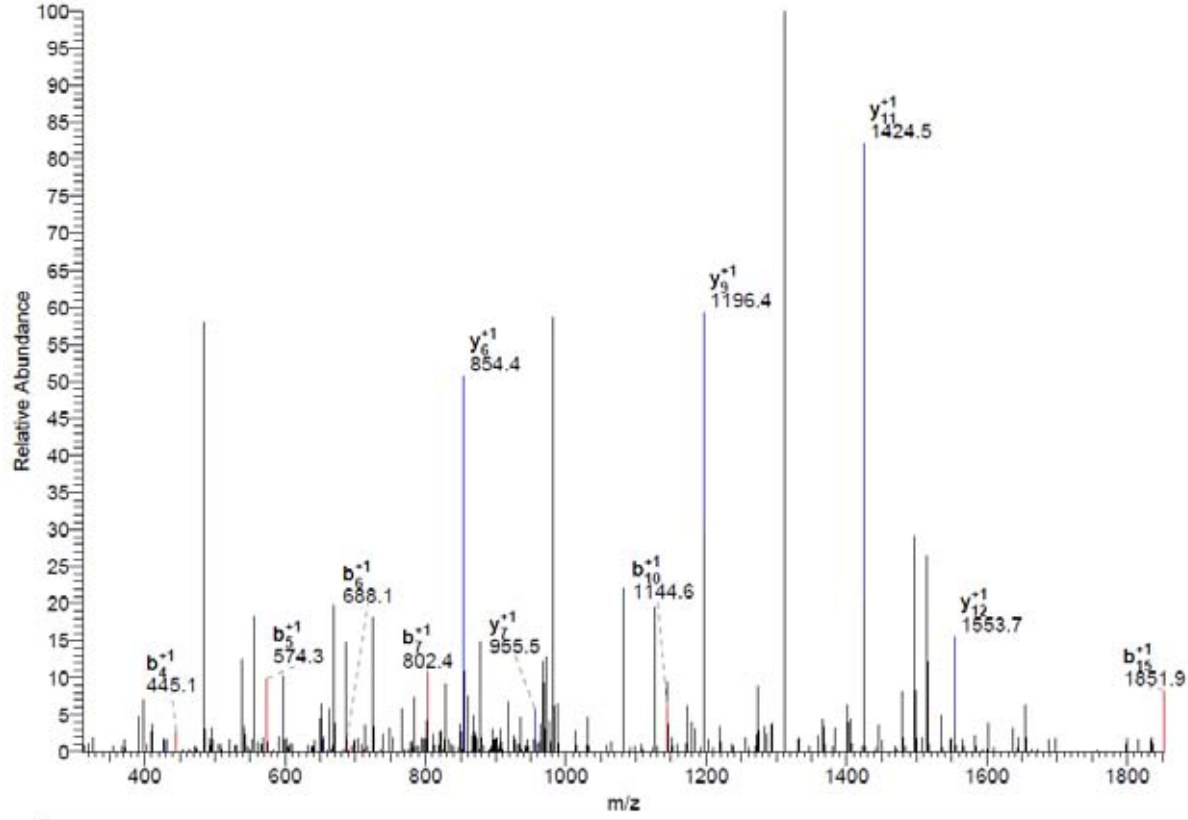
DTA for scan: 9418
Precursor ion: 999.78
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
V		100.08							
C		203.08				1898.84			
L		316.17				1795.83			
E		**445.21**				1682.74			
E		**574.25**				**1553.70**			
N		**688.30**				**1424.66**			
N		**802.34**				1310.61			
Q		930.40				**1196.57**			
I		1043.48				1068.51			
T		**1144.53**				**955.43**			
W		1330.61				**854.38**			
T#		1511.62				668.30			
D		1626.65				487.29			
K		1754.75				372.26			
P		**1851.80**				244.17			
K						147.11			



#9418-9418 RT:52.90-52.90 NL: 2.19E2



Reference Scan(s)	Sequence	MH+	z	P P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein AP2 domain containing protein, expressed				0.0009	10.1	0.0	0			
11275	R.MMEADPT#NEGLWKGDK.D	1901.78	2	0.0009	2.147	0.035	90.1	19	9/45	3

1 of 1 peptide matches reported, 0 removed due to filtering

Reference: LOC_Os05g27930.1|13105.m02872|protein AP2 domain containing prote
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 373 Monoisotopic MW: 40448.7 pI: 4.61

Protein:

MTVDQRTTAK AIMPPVEMPP VQPGRKKRPR RSRDGPTSVA ETIKRWAELN
NQQELDPQGP KKARKAPAKG SKKGCMKGKG GPENTRCDFR GVRQRTWGKW
VAEIREPNQQ SRLWLGTFPT AEAAACAYDE AARAMYGPMA RTNFGQHHAP
AASVQVALAA VKCALPGGGL TASKSRTSTQ GASADVQDVL TGGLSACEST
TTTINNQSDV VSTLHKPEEV SEISSPLRAP PAVLEDGSNE DKAESVTYDE
NIVSQQRAPP EAEASNGRGE EVFEPLEPIA SLPEDQGDYC FDIDEMLRMM
EADPTNEGLW KGDKDGSDAI LELGQDEPFY YEGVDPGMLD NLLRSDEPAW
LLADPAMFIS GGFEDDSQFF EGL

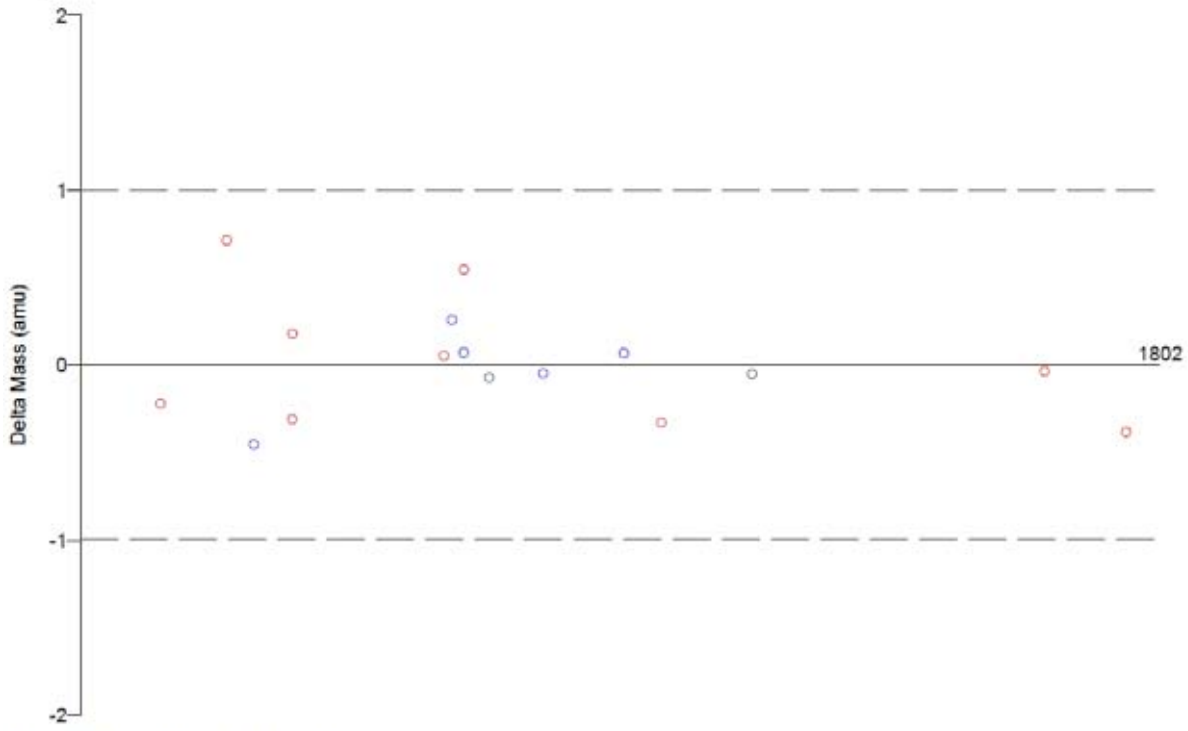
Protein Coverage:

Sequence		MH+	% Mass	AA	% AA
MMEADPTNEGLWKGDK		1821.81	4.50	299 - 314	4.29
	Totals:	1821.81	4.50	16	4.29

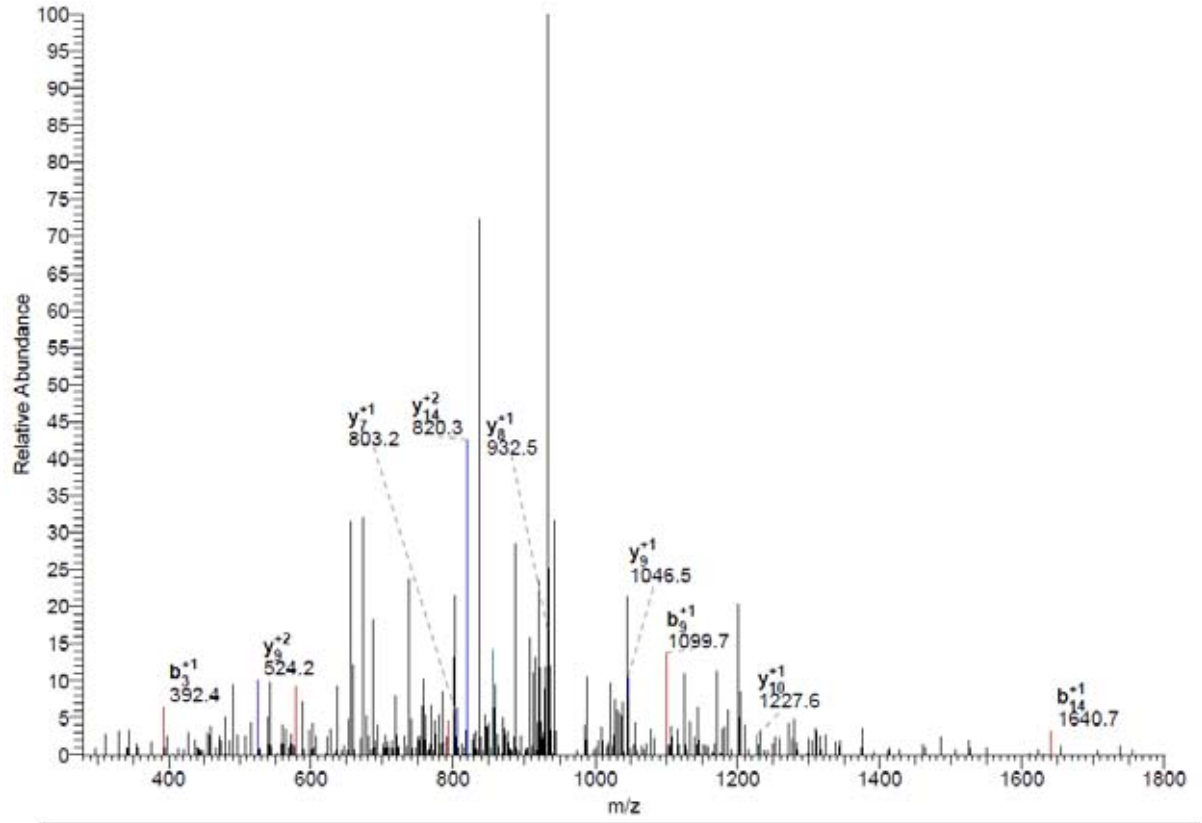
DTA for scan: 11275
Precursor ion: 951.75
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
M		132.05							
M		263.09				1770.74			
E		**392.13**				1639.70			
A		463.17				1510.66			
D		**578.19**				1439.62			
P		675.25				1324.59			
T#		**856.26**				**1227.54**			
N		970.30				**1046.53**			
E		**1099.35**				**932.48**			
G		1156.37				**803.44**			
L		1269.45				746.42			
W		1455.53				633.34			
K		1583.63				447.26			
G		**1640.65**				319.16			
D		**1755.68**				262.14			
K						147.11			



#11275-11275 RT:63.33-63.33 NL: 1.48E2



Reference Scan(s)	Sequence	MH+	z	P P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein protein kinase, putative, expressed				0.0009	2.1	0.0	0			
8334	K.SSVAAKNT#EPPKR.I	1464.72	2	0.0009	2.001	0.126	118.8	4	11/36	

1 of 3 peptide matches reported, 2 removed due to filtering

Reference: LOC_Os05g03460.1|13105.m00331|protein protein kinase, putative, e
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 443 Monoisotopic MW: 49639.1 pI: 8.39

Protein:

MSFFSCFKPD KKMLSKRMEE MPFTVVKKAS SQHGSSLKNS ESDKSPRGNS
NNKKSSVAAK NTEPPKRIPI TAKAERSFTF RELATATNNF HPDCIVGEGG
FGRVYKGQLE DGQVVAVKQM ERNGFQGNRE FLIEVMILGH LNHPNLVNLV
GYCSDGDQRL LAYEYMALGS LADHLLDITP DQEPLSWRTR MKIAHGTAKG
LEHLHEKMSP PVIYRDLKSP NILLDKDYNP KLSDFGLAKL GPFEGDKHVS
TRVMGTFGYC APEYVRTGML STKTDVYSFG VFLLELITGR RAVDTCRPVC
EQILAYWAKP MLHDRRRYHE LVDPLLRGDY PDKDFNQAAA VAAICIEDEA
SVRPYMSDIV VALGFLAEVP AGCEERINAE PQNRKDEDPS VTGNTKKDQS
TFDRQRAVAE AIEWGSVRQK QMAQIKEKKA QPQGIVAPKT NKL

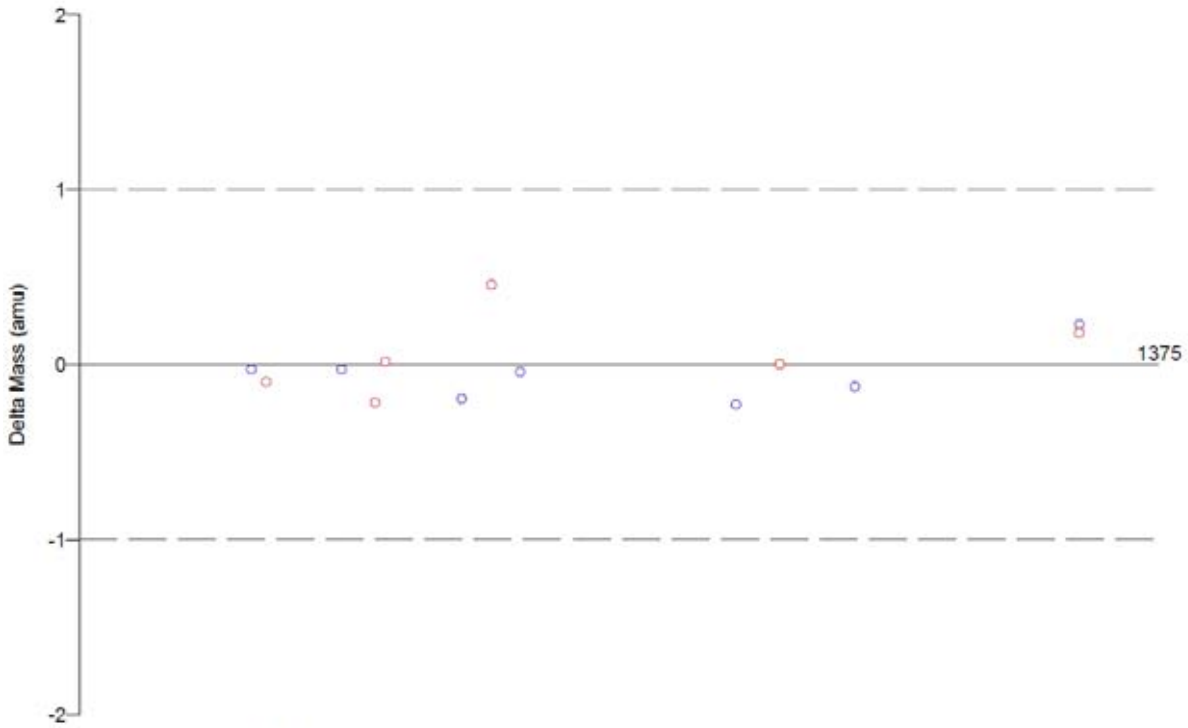
Protein Coverage:

Sequence	MH+	% Mass	AA	% AA
SSVAAKNTEPPKR	1384.75	2.79	55 - 67	2.93
IPITAKAERSFTFR	1636.92	3.30	68 - 81	3.16
MKIAHGTAK	956.53	1.93	191 - 199	2.03
Totals:	1384.75	2.79	13	2.93

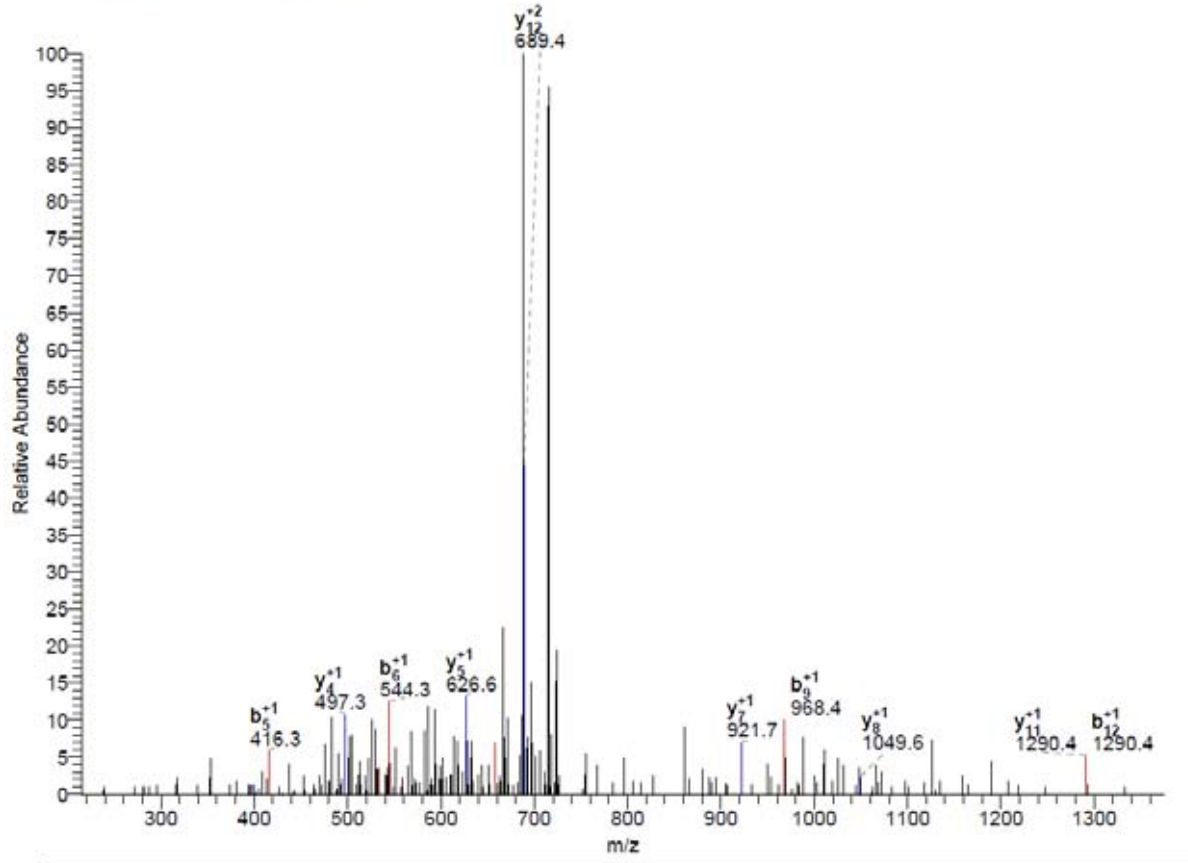
DTA for scan: 8334
Precursor ion: 732.70
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

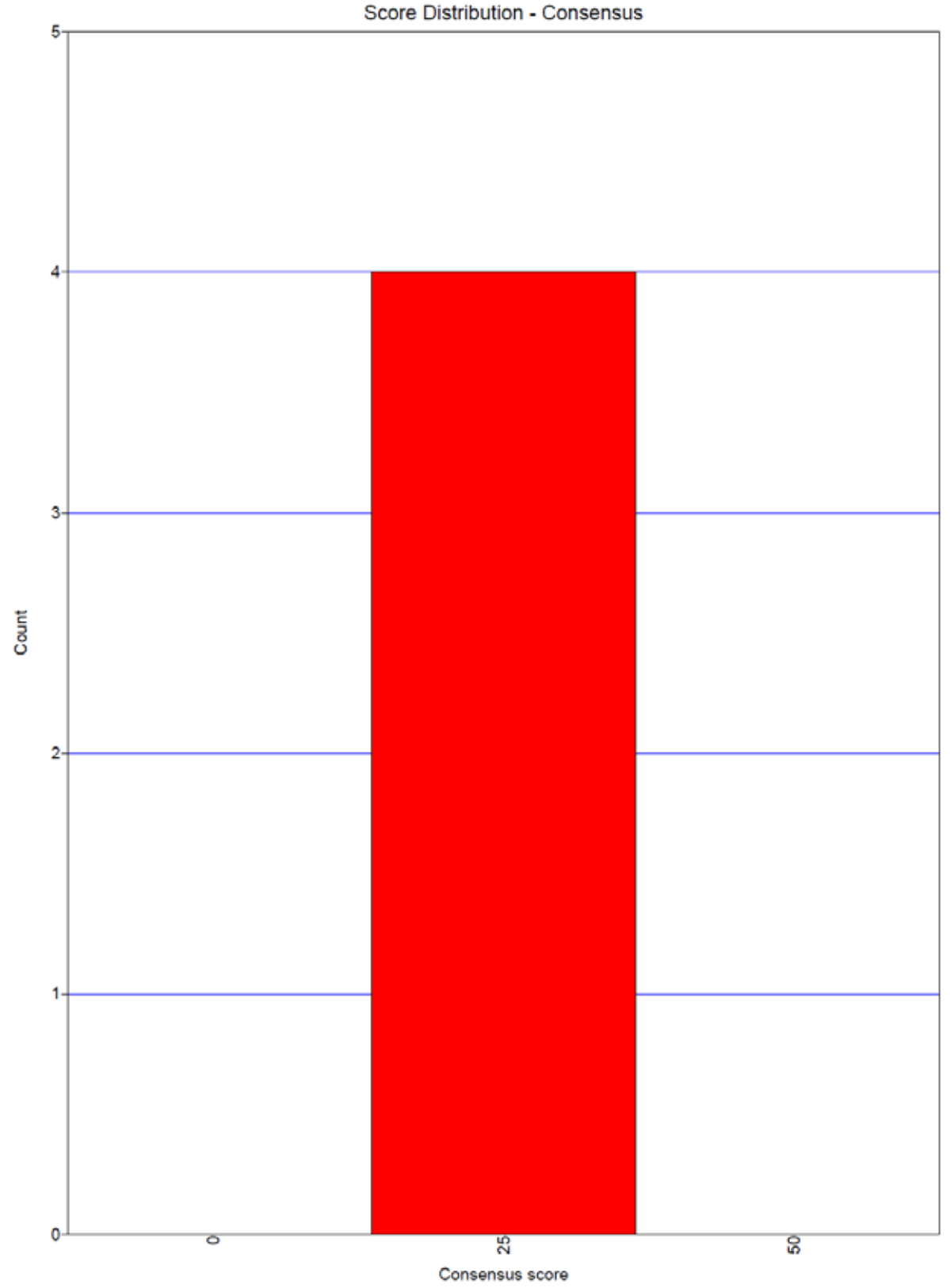
Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
S		88.04							
S		175.07				1377.69			
V		274.14				**1290.66**			
A		345.18				1191.59			
A		**416.21**				1120.55			
K		**544.31**				**1049.51**			
N		**658.35**				**921.42**			
T#		839.37				807.38			
E		**968.41**				**626.36**			
P		1065.46				**497.32**			
P		1162.51				**400.27**			
K		**1290.61**				303.21			
R						175.12			



#8334-8334 RT:46.76-46.76 NL: 9.88E1




Score Distribution - Consensus
5
4
3
2
1
0
Count
0
25
50
Consensus score

```
[SEQUEST]
first_database_name = C:\Xcalibur\database\rice_all.pep.fasta.hdr
second_database_name =
peptide_mass_tolerance = 1.0000
peptide_mass_units = 0                  ; 0=amu, 1=mmu, 2=ppm
ion_series = 0 1 1 0.0 1.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0
fragment_ion_tolerance = 0.5000           ; for trap data leave at 1.0, for accurate mass data use values < 1.0
fragment_ion_units = 0                  ; 0=amu, 1=mmu
num_output_lines = 10                    ; # peptide results to show
num_results = 250                         ; # results to store
num_description_lines = 5               ; # full protein descriptions to show for top N peptides
show_fragment_ions = 0                  ; 0=no, 1=yes
print_duplicate_references = 0          ; number of duplicate references reported
enzyme_info = Trypsin(KR) 1 1 KR -
max_num_differential_per_peptide = 3           ; max # of diff. mod in a peptide
diff_search_options = 15.994915 M 79.966331 STY 0.000000 M 0.000000 X 0.000000 T 0.000000 Y
term_diff_search_options = 0.000000 0.000000
nucleotide_reading_frame = 0            ; 0=protein db, 1-6, 7 = forward three, 8-reverse three, 9=all six
mass_type_parent = 1                    ; 0=average masses, 1=monoisotopic masses
mass_type_fragment = 1                  ; 0=average masses, 1=monoisotopic masses
normalize_xcorr = 0                     ; use normalized xcorr values in the out file
remove_precursor_peak = 0               ; 0=no, 1=yes
ion_cutoff_percentage = 0.0000             ; prelim. score cutoff % as a decimal number i.e. 0.30 for 30%
max_num_internal_cleavage_sites = 2              ; maximum value is 12
protein_mass_filter = 0 0               ; enter protein mass min & max value ( 0 for both = unused)
match_peak_count = 0                    ; number of auto-detected peaks to try matching (max 5)
match_peak_allowed_error = 1            ;  number of allowed errors in matching auto-detected peaks
match_peak_tolerance = 1.0000              ; mass tolerance for matching auto-detected peaks
partial_sequence =
sequence_header_filter =
digest_mass_range = 600.0000 3500.0000

add_Cterm_peptide = 0.0000              ; added to each peptide C-terminus
add_Cterm_protein = 0.0000              ; added to each protein C-terminus
add_Nterm_peptide = 0.0000              ; added to each peptide N-terminus
add_Nterm_protein = 0.0000              ; added to each protein N-terminus
add_G_Glycine = 0.0000                  ; added to G
add_A_Alanine = 0.0000                  ; added to A
add_S_Serine = 0.0000                   ; added to S
add_P_Proline = 0.0000                  ; added to P
add_V_Valine = 0.0000                   ; added to V
add_T_Threonine = 0.0000                ; added to T
add_C_Cysteine = 0.0000                 ; added to C
add_L_Leucine = 0.0000                  ; added to L
add_I_Isoleucine = 0.0000               ; added to I
add_X_LorI = 0.0000                     ; added to X
add_N_Asparagine = 0.0000               ; added to N
add_O_Ornithine = 0.0000                ; added to O
add_B_avg_NandD = 0.0000                ; added to B
add_D_Aspartic_Acid = 0.0000            ; added to D
add_Q_Glutamine = 0.0000                ; added to Q
add_K_Lysine = 0.0000                   ; added to K
add_Z_avg_QandE = 0.0000                ; added to Z
add_E_Glutamic_Acid = 0.0000            ; added to E
add_M_Methionine = 0.0000               ; added to M
add_H_Histidine = 0.0000                ; added to H
add_F_Phenylalanine = 0.0000            ; added to F
add_R_Arginine = 0.0000                 ; added to R
add_Y_Tyrosine = 0.0000                 ; added to Y
add_W_Tryptophan = 0.0000               ; added to W
add_J_user_amino_acid = 0.0000          ; added to J
add_U_user_amino_acid = 0.0000          ; added to U
```

Output files for KK06 protein spot sequencing

KK09B06.RAW Page 1

```
SRF File: E:\data\KELVIN\09-07-21\KK09B06.srf
Database... indexed - rice_all.pep.fasta.hdr (7/23/2009)
Filter(s)... xc (± 1,2,3)=1.50,2.00,2.50 ; peptide probability<=1e-003
Mods: (M* +15.99492) (STY# +79.96633)

Reference                                                                    P      Score   Coverage   Accession
  Scan(s)       Sequence                                    MH+      z       P        XC  DeltaCn    Sp   RSp  Ions

protein NIN, putative, expressed                                          0.0007   16.1     0.0     0
  10313         K.FY#TENQKFAFTEILDVLR.A                   2314.12    2  0.0007   2.786  0.042   99.1    3   11/51
  10313         K.FYT#ENQKFAFTEILDVLR.A                   2314.12    2  0.0007   2.639  0.111   99.1    3   11/51

  2 of 6 peptide matches reported, 4 removed due to filtering


protein dnaJ domain containing protein, expressed                         0.0002   10.2     0.0     0
  9838          K.NKSVGAEGAFKMVQEAWT#VLSDK.T              2575.23    3  0.0002   3.057  0.111   116.7   50  19/132

  1 of 7 peptide matches reported, 6 removed due to filtering


protein retrotransposon protein, putative, unclassified                   2e-005   10.1     0.0     0
  8353          K.M*NEIVAGNEMQRLGVEKK.M                   2062.04    2  2e-005   2.727  0.021   328.1   1   16/34

  1 of 1 peptide matches reported, 0 removed due to filtering


protein expressed protein                                                 0.0004   10.1     0.0     0
  10915         K.SDLDKADEWVK.I                           1305.63    2  0.0004   2.108  0.050   313.6   2   9/20

  1 of 1 peptide matches reported, 0 removed due to filtering


protein inactive receptor kinase At2g26730 precursor, putative            0.0002   8.2      0.0     0
  8246          K.YLNIEDELVQM*LQLAM*ACTSRSPERR.P          3098.51    3  0.0002   3.450  0.158   66.1    30  15/100

  1 of 4 peptide matches reported, 3 removed due to filtering


protein retrotransposon protein, putative, Ty3-gypsy subclass             0.0004   6.2      0.0     0
  9690          R.VQKLLY#GVLIT#VR.K                       1661.88    2  0.0004   2.357  0.078   343.0   2   17/48

  1 of 5 peptide matches reported, 4 removed due to filtering


protein dehydration response related protein, putative, expres            0.0007   6.1      0.0     0
  8375          R.TRTAS#PCGRS#R.W                         1351.53    2  0.0007   2.320  0.195   128.1   5   13/40

  1 of 4 peptide matches reported, 3 removed due to filtering


protein DnaK family protein, putative, expressed                          0.0003   4.1      0.0     0
  8236          R.FDDPQTQKEMKMVPY#K.I                     2064.92    2  0.0003   2.640  0.138   121.6   4   15/45

  1 of 2 peptide matches reported, 1 removed due to filtering


protein expressed protein                                                 0.001    4.1      0.0     0
  11377         R.NFAY#EKEIPAS#S#TCMDK.K                  2173.77    2  0.001    2.002  0.154   136.8   25  12/80

  1 of 3 peptide matches reported, 2 removed due to filtering
```

KK09B06.RAW Page 1


```
SRF File: E:\data\KELVIN\09-07-21\KK09B06.srf
Database... indexed - rice_all.pep.fasta.hdr (7/23/2009)
Filter(s)... xc (± 1,2,3)=1.50,2.00,2.50 ; peptide probability<=1e-003
Mods: (M* +15.99492) (STY# +79.96633)

Reference                                                                    P      Score   Coverage   Accession
Scan(s)         Sequence                                    MH+      z       P        XC  DeltaCn    Sp   RSp  Ions  Count

protein NIN, putative, expressed                                          0.0007   16.1     0.0     0
10313           K.FY#TENQKFAFTEILDVLR.A                   2314.12    2  0.0007   2.786  0.042   99.1    3   11/51
10313           K.FYT#ENQKFAFTEILDVLR.A                   2314.12    2  0.0007   2.639  0.111   99.1    3   11/51

2 of 6 peptide matches reported, 4 removed due to filtering
```

Reference: LOC_Os04g41850.1|13104.m04114|protein NIN, putative, expressed
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 936 MonoIsotopic MW: 101539.6 pI: 5.68

Protein:

MDMPTPSNRA GCNGNTGGTM GPSDDPYGAA AMNLDCYSEI YSPSVADQLF
SLLNDPAAHR MFAMWPSMGS SPCAAGTSED MPLDAYSGLG RAVERPSQIM
SVNPTEAEKT GKSSGELGSD DGAHQGSSMV PRSVVGSSLA DRMLMALSLF
RESLGSGALA QVWMPVEQEG HVVLSTCEQP FLLDQVLAGY REVSRHFVFS
AKEEPGLQPG LPGRVFISGV PEWTSSVLYY NRPEYLRMEH ALHHEIRGSL
AMPIYDPSKD SCCAVFELVT RKEKPDFSAE MDNVCNALQA VNLKATKGSS
NQKFYTENQK FAFTEILDVL RAICHAHMLP LALTWVPTSN GIDGGYVVGK
DGASFSQSGK TIIRIHESAC YVNDGKMQGF LQACARRHLE KGQGIAGRAL
KSNLPFFSPD IREYSIEDYP LAHHARKFSL HAAVAIRLRS TYTGNDDYIL
EFFLPVSCKG SGEQQMLLNN LSSTMQRICK SLRTVYEAEV DNVNAGTAAV
FRKNNESCLP TGHTESSSHG DQSITGASFE DTSLANKPGV MEPELAEQVQ
PSSIGHAEKK RSTAEKNISL DVLRKYFSGS LKDAAKSLGV CPTTLKRICR
HHGISRWPSR KINKVNRSLK KIQTVINSVH GVDRSLQYDP ATGSLVPVVS
LPEKLTFPSC DGLPTPSVGK TVEENSDLKS EEGCSLPDGS QRQSCQLQIS
DVKKSNEDEF HIGSGNSDFY GANATAKSNS EVTQGPLCPT GAPSALHLKG
TDCTNPSSSL RPSSESTRNQ IVGRNSPSIQ QEDLDMLDNH EAEDKDHMHP
STSGMTDSSS GSASSHPTPK QNTRSALKDA ASPALTVKAT YNGDTVRPKP
LPSMGWYHLL EEIAKRFKLP TGAYQLKYKD DEDEWVILAN DSDLQECVDV
LDSIGSRIVK LQVRDLPCIV SSSGSSTCLQ LAAHSS

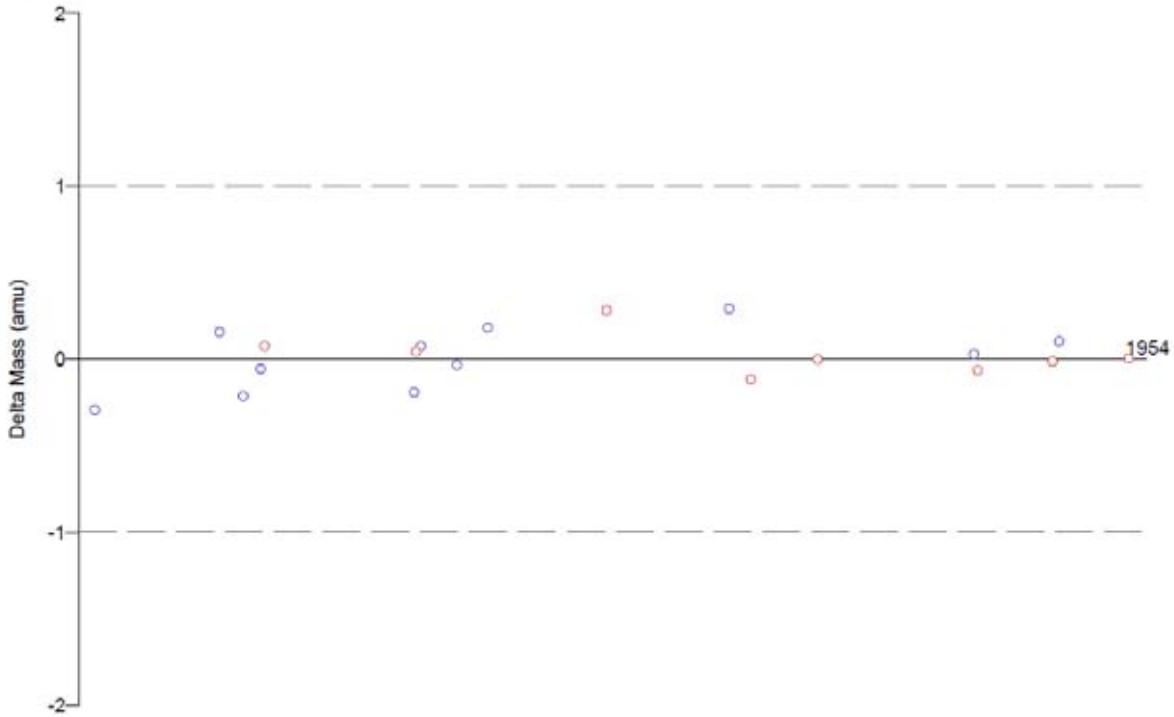
Protein Coverage:

Sequence		MH+	% Mass	AA	% AA
FYTENQKFAFTEILDVLR		2234.15	2.20	304 - 321	1.92
DLPCIVSSSGSSTCLQLAAHSS		2163.01	2.13	915 - 936	2.35
	Totals:	2234.15	2.20	18	1.92

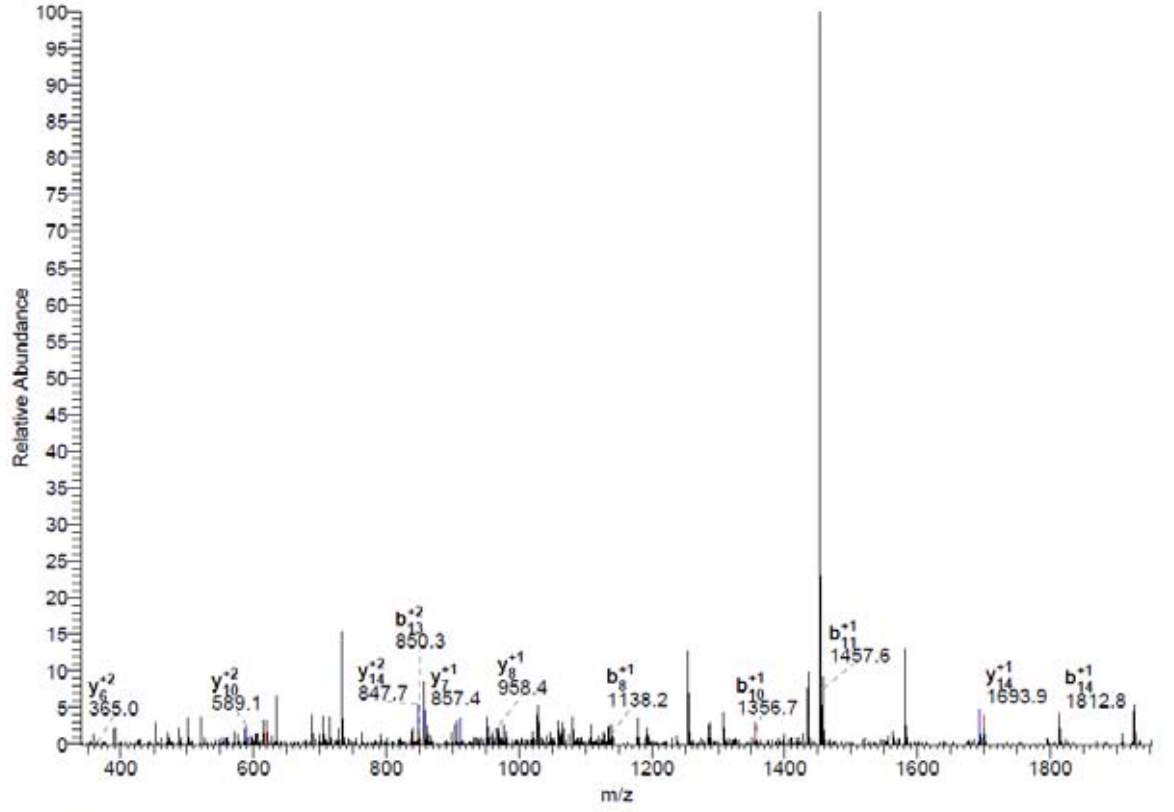
DTA for scan: 10313
Precursor ion: 1157.05
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
F		148.08							
Y#		391.11				2167.05			
T		492.15				1924.02			
E		**621.20**				**1822.97**			
N		735.24				**1693.93**			
Q		863.30				1579.88			
K		991.39				1451.83			
P		**1138.46**				**1323.73**			
A		1209.50				1176.66			
F		**1356.57**				1105.63			
T		**1457.61**				**958.56**			
E		1586.66				**857.51**			
I		**1699.74**				728.47			
L		**1812.82**				**615.38**			
D		**1927.85**				502.30			
V		2026.92				387.27			
L		2140.00				288.20			
R						175.12			



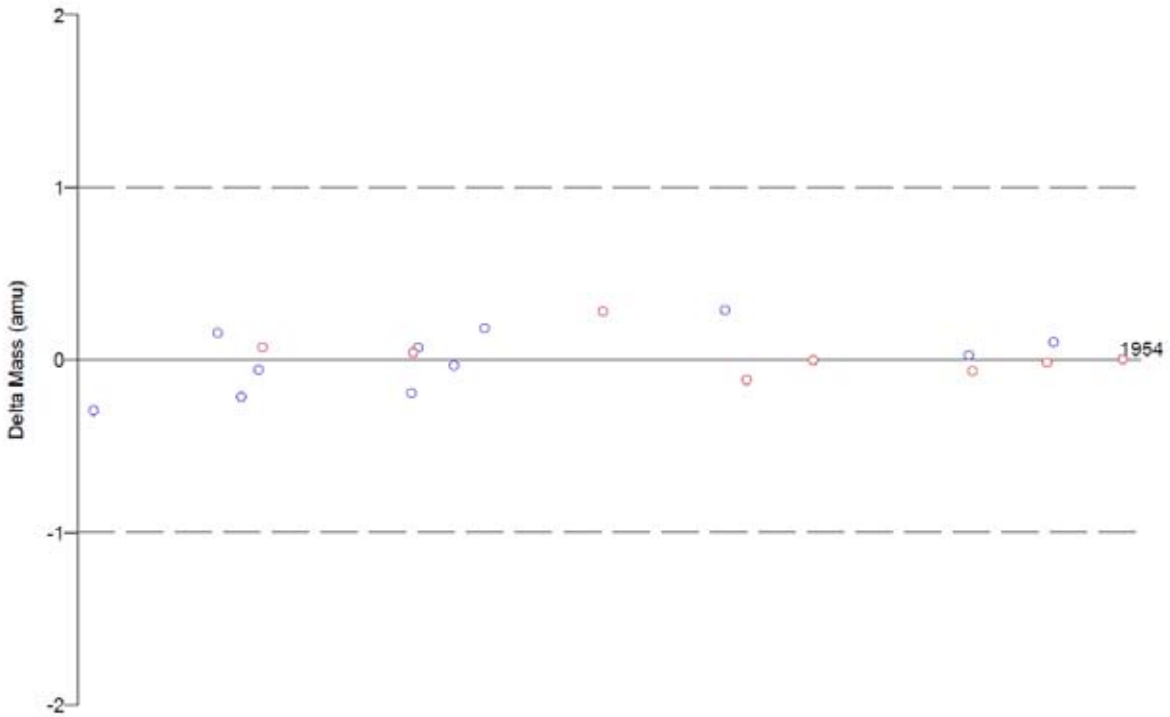
#10313-10313 RT:58.10-58.10 NL: 1.28E3



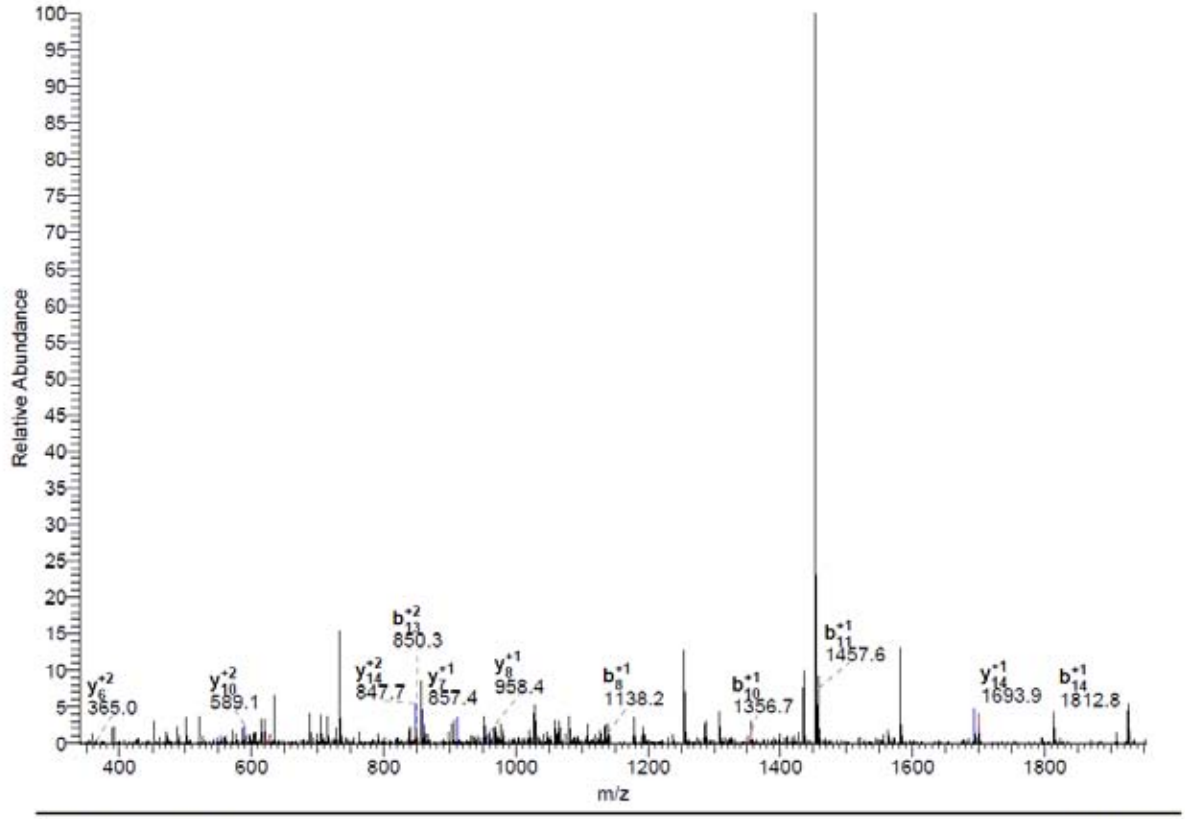
DTA for scan: 10313
Precursor ion: 1157.85
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
F		148.08							
Y		311.14				2167.05			
T#		492.15				2003.98			
E		**621.20**				**1822.97**			
N		735.24				**1693.93**			
Q		863.30				1579.88			
K		991.39				1451.83			
P		**1138.46**				**1323.73**			
A		1209.50				1176.66			
F		**1356.57**				1105.63			
T		**1457.61**				**958.56**			
E		1586.66				**857.51**			
I		**1699.74**				728.47			
L		**1812.82**				**615.38**			
D		**1927.85**				502.30			
V		2026.92				387.27			
L		2140.00				288.20			
R						175.12			



#10313-10313 RT:58.10-58.10 NL: 1.28E3



Reference Scan(s)	Sequence	MH+	z	P P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein dnaJ domain containing protein, expressed				0.0002	10.2	0.0	0			
9838	K.NKSVGAEGAFKMVQEAWT#VLSDK.T	2575.23	3	0.0002	3.057	0.111	116.7	50	19/132	

1 of 7 peptide matches reported, 6 removed due to filtering

Reference: LOC_Os03g28310.1|13103.m03304|protein dnaJ domain containing prot
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 748 Monoisotopic MW: 83354.5 pI: 7.49

Protein:

MECNKDEALR AKEIAERKFE SKDLQGAKKF ALKAQALFPG LEGIVQMITT
LDLYLASEVL ISGEKDWYSI LSVESSADDE TLKKQYRKLV LQLHPDKNKS
VGAEGAFKMV QEAWTVLSDK TKRALYDQKR KLMVLKRNTS QTNKASAAPG
ASNGFYNFAA NAAASKVTRG NKQKAGPATS SVRQRPPPPP PPPRQAPAPP
PAKPPTFWTS CNKCKMNYEY LKVYLNHNLL CPTCREPFLA QEVPMPPTES
VHAVHDPNIS GANQNTNGSR NFQWGPFSRT AGAASATASS AAAAQAANVV
HHTYEKVRRE REEAQAAARR EEALRRKYNP PKRQANISEN LNLGTGGNSS
KKMRTMGNDI GIGSSSILSG SGANYPGVPG GNISFSTNSG AHHPQGVNGG
PSWKPRPPTR ISLVKTFTQP DVRGILMEKA KSDLKDKLKE MQTKRSQVAA
NGKKNKKNMF KESGGDDESL ASDDSTARQA AHVDPEDNAS VNSTDADDEN
DDPLSYNVPD PDFHDFDKDR TEECFQSDQI WATYDDEDGM PRYYAFIQKV
LSLSPFQLKI SPLTSRTNSE PGSLNWVSSG PTKTCGDFRI CRYETCDILN
MFSHQIKWEK GPRGVIKIYP QKGNIWAVYR NWSPDWDEDT PDKVLHAYDV
VEVLDEYDED LGISVIPLVK VAGFRTVFQR NQDLNAIKKI PKEEMFRFSH
EVPFYRMSGE EAPNVPKDSY ELDPAAISKE LLQEITETVE SSKATSEC

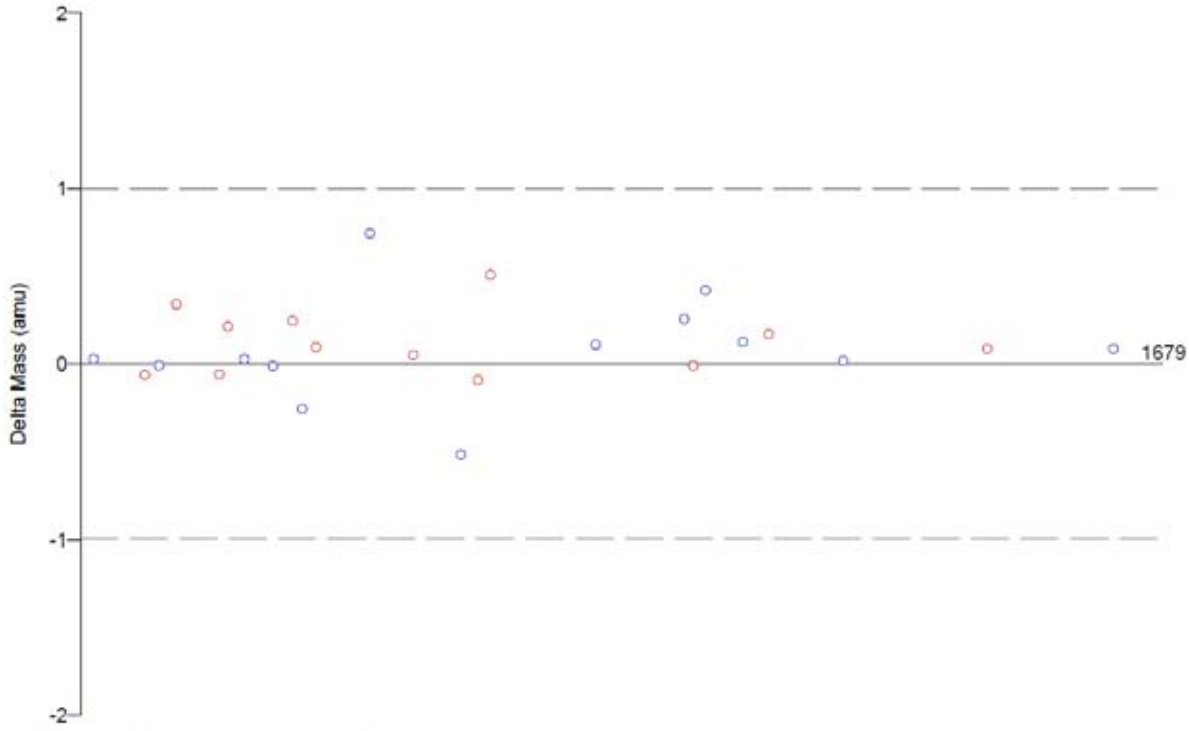
Protein Coverage:

Sequence	MH+	% Mass	AA	% AA
NKSVGAEGAFKMVQEAWTVLSDK	2495.26	2.99	98 - 120	3.07
KNMFKESGGDDESLASDDSTAR	2360.03	2.83	457 - 478	2.94
ISFLTSRTNSEPGSLNWVSSGPTK	2665.33	3.20	560 - 583	3.21
Totals:	2495.26	2.99	23	3.07

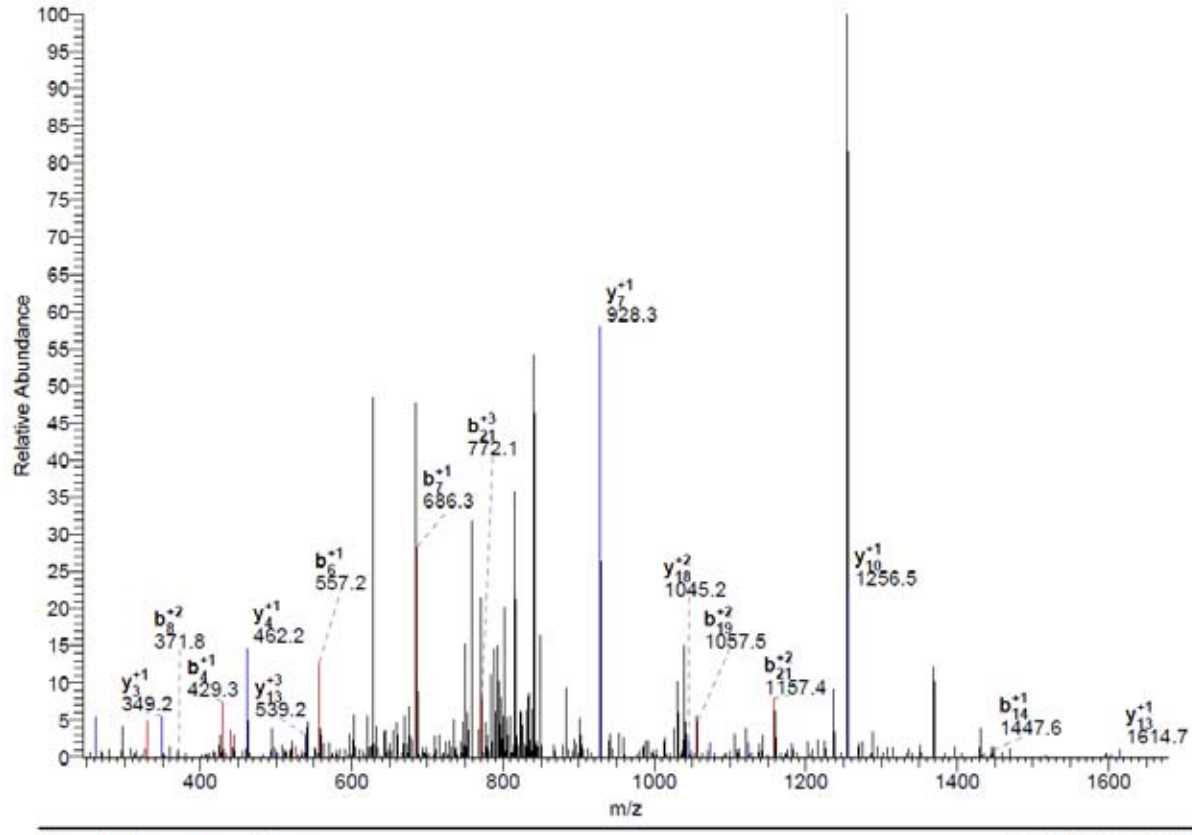
DTA for scan: 9838
Precursor ion: 858.80
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
N		115.05							
K		243.15				2461.18			
S		330.18				2333.09			
V		429.25				2246.06			
G		486.27				2146.99			
A		557.30				2089.97			
E		686.35				2018.93			
G		743.37				1889.89			
A		814.41				1832.87			
P		961.47				1761.83			
K		1089.57				**1614.76**			
M		1220.61				1486.66			
V		1319.68				1355.62			
Q		1447.74				**1256.56**			
E		1576.78				1128.50			
A		1647.82				999.45			
W		1833.90				**928.42**			
T#		2014.91				742.34			
V		2113.98				561.32			
L		2227.06				**462.26**			
S		2314.09				**349.17**			
D		2429.12				**262.14**			
K						147.11			



#9838-9838 RT:55.42-55.42 NL: 2.33E2



Reference Scan(s)	Sequence	MH+	z	P P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein retrotransposon protein, putative, unclassified				2e-005	10.1	0.0	0			
8353	K.M*NEIVAGNEMQRLGVEKK.M	2062.04	2	2e-005	2.727	0.021	328.1	1	16/34	

1 of 1 peptide matches reported, 0 removed due to filtering

Reference: LOC_Os01g19300.1|13101.m02099|protein retrotransposon protein, pu
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 81 Monoisotopic MW: 9367.6 pI: 4.78

Protein:

MMTECRRLFE NTSREVGAQM DRLTARAISS DGRVAQLEME LEVARDDLQK
MNEIVAGNEM QRLGVEKKMN DLQDHIPSIR D

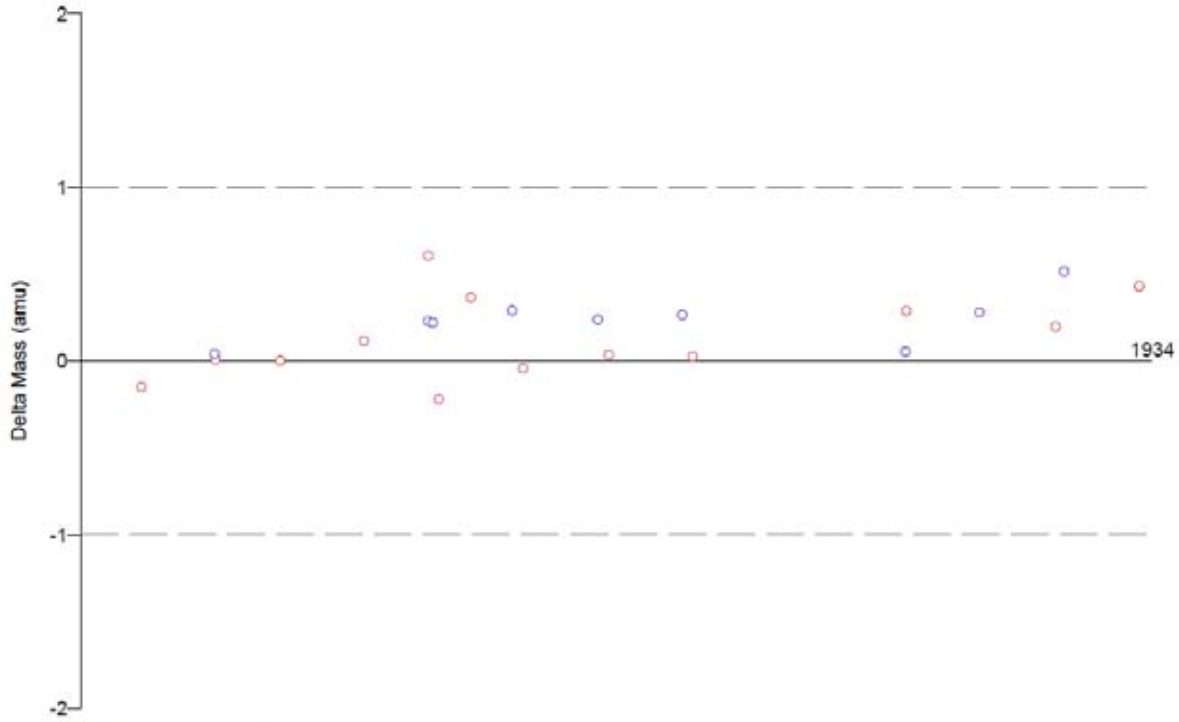
Protein Coverage:

Sequence		MH+	% Mass	AA	% AA
MNEIVAGNEMQRLGVEKK		2046.05	21.84	51 - 68	22.22
	Totals:	2046.05	21.84	18	22.22

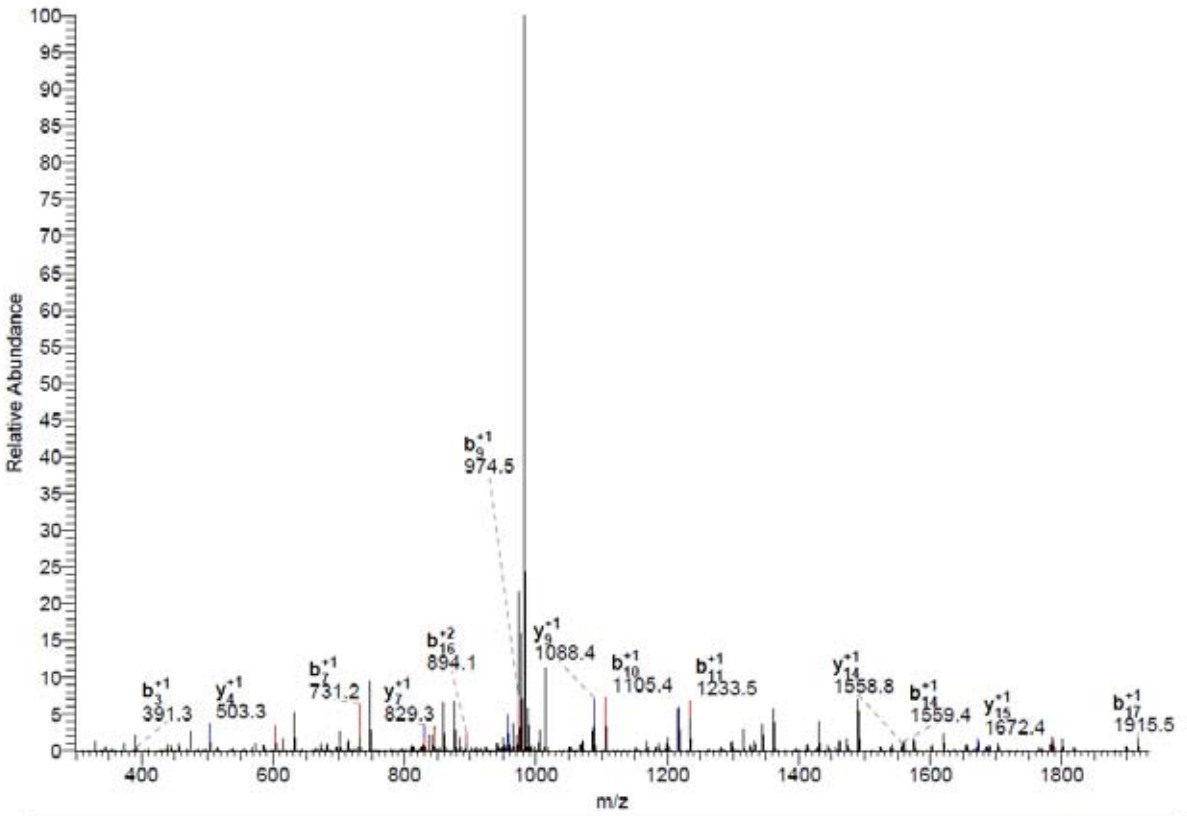
DTA for scan: 8353
Precursor ion: 1032.01
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
M*		148.04							
N		262.09				1915.01			
E		391.13				1800.96			
I		504.21				1671.92			
V		603.28				1558.84			
A		674.32				1459.77			
G		731.34				1388.73			
N		845.38				1331.71			
E		974.42				1217.67			
M		1105.47				1088.62			
Q		1233.52				957.58			
R		1389.62				829.53			
L		1502.71				673.42			
G		1559.73				560.34			
V		1658.80				503.32			
E		1787.84				404.25			
K		1915.94				275.21			
K						147.11			



#8353-8353 RT:47.05-47.05 NL: 1.48E4



Reference Scan(s)	Sequence	MH+	z	P P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein expressed protein				0.0004	10.1	0.0	0			
10915	K.SDLDKADEWVK.I	1305.63	2	0.0004	2.108	0.050	313.6	2	9/20	3

1 of 1 peptide matches reported, 0 removed due to filtering

Reference: LOC_Os01g71540.1|13101.m07788|protein expressed protein
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 532 Monoisotopic MW: 59851.7 pI: 7.14

Protein:

MSESALDKHI VVKIPDHVLP DAFIGESVGK GKEDGATVDN MDPSRPFLVI
KKSDLDKADE WVKIKYAWMA KSLTNDIIIP DPTPQVVRDA FFDISPRLDS
VLRKDSVRCF FRLFAECTAA MAFKCNITSE TLSCIVRHNA LRCAKTVLEG
KAPQLSCMHA NPNCINPYGI FPLHEAAERF SVDMIKLLFC HGASANLHTV
NDAGIPGLLP LHVAVGNTCL HKYLEDNLSP VQYHEDYIYK LIHLLCLPEM
KVFLDTVRLL AEKTDNLADE LWNYMKNDKL VESAVLLLAA RKHIREGKPD
GFDIIAQRIY EDYDSLVCDK GDTAEGQKLL EERRALLKCK CLIVTIISQA
GEVLDNYIQA HSEVPNVEVL ARVSYILKEF GFCPNEEYID TMILCPYNKI
SYSDIVHKDV TKAVAQMSTS LPAAEKKAAR KKALKGWDPT FIKRNFFPYW
RSVLGAQLSV SNGAADEKSM LHRPQFRNSV VQYCCIQNAH EGAKDMSKKV
PSGMLLLITI GLQSIMLRTS FADTLRRLRI YM

Protein Coverage:

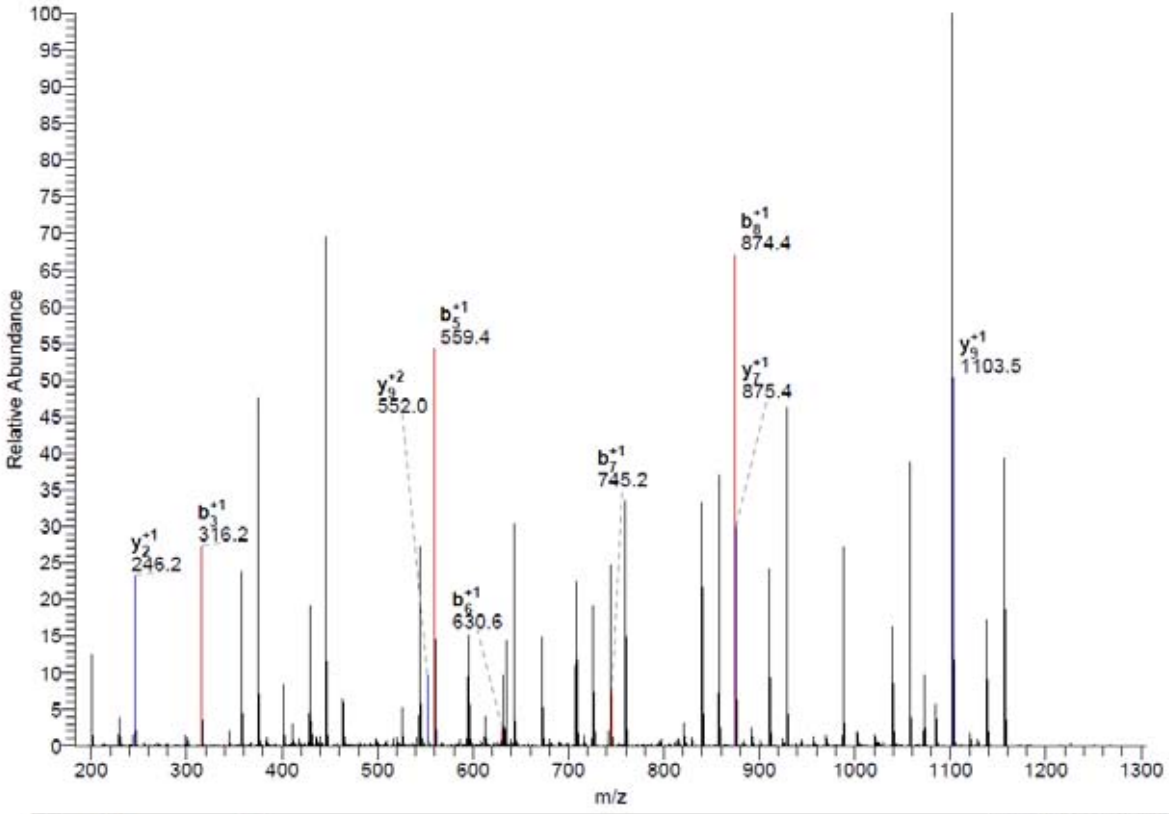
Sequence		MH+	% Mass	AA	% AA
SDLDKADEWVK		1305.63	2.18	53 - 63	2.07
	Totals:	1305.63	2.18	11	2.07

DTA for scan: 10915
Precursor ion: 653.36
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
S		88.04							
D		203.07				1218.60			
L		316.15				1103.57			
D		431.18				990.49			
K		559.27				875.46			
A		630.31				747.37			
D		745.34				676.33			
E		874.38				561.30			
W		1060.46				432.26			
V		1159.53				246.18			
K						147.11			

#10915-10915 RT:61.50-61.50 NL: 1.12E4



Reference Scan(s)	Sequence	MH+	z	P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein inactive receptor kinase At2g26730 precursor, putative				0.0002	8.2	0.0	0			
8246	K.YLNIEDELVQM*LQLAM*ACTSRSPERR.P	3098.51	3	0.0002	3.450	0.158	66.1	30	15/100	2

1 of 4 peptide matches reported, 3 removed due to filtering

Reference: LOC_Os03g21510.1|13103.m02574|protein inactive receptor kinase At
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 634 Monoisotopic MW: 68941.7 pI: 6.88

Protein:

MGYKIPEGSL VLLSLLLFTS LAVADIASEK QALLAPASAV YRGNKLNWDV
NISLCSWHGV TCSPDRSRIS ALRVPAAGLI GAIPPNTLGR LVSLQVLSLR
SNRLIGSIPS DITSLPSLQS IFLQDNELSG DLPSFFSPTL NTIDLSYNSF
AGQIPASLQN LTQLSTLNLS KNSLSGPIPD LKLPSLRQLN LSNNELNGSI
PPFLQIFSNS SFLGNPGLCG PPLAECSLPS PTSSPESSLP PPSALPHRGK
KVGTGSIIAA AVGGFAVFLL AAAIFVVCFS KRKEKKDDGL DNNGKGTDNA
RIEKRKEQVS SGVQMAEKNK LVFLDGCSYN FDLEDLLRAS AEVLGKGSYG
TAYKAILEDG TIVVVKRLKD VVAGKKEFEQ QMEQIGRVGK HANLVPLRAY
YYSKDEKLVV YEYVATGSFS AMLHGIKGIA EKTPLDWNTR MKIILGTARG
IAHIHAEGGS KLAHGNIKAT NVLLDQDHNP YVSDYGLSAL MSFPISTSRV
VVGYRAPETF ESRKFTHKSD VYSFGVLLME MLTGKAPLQS QGQDDVVDLP
RWVHSVVREE WTAEVFDVEL MKYLNIEDEL VQMLQLAMAC TSRSPERRPT
MAEVIRMIEE LRQSASESRD SSNENARESN PPSA

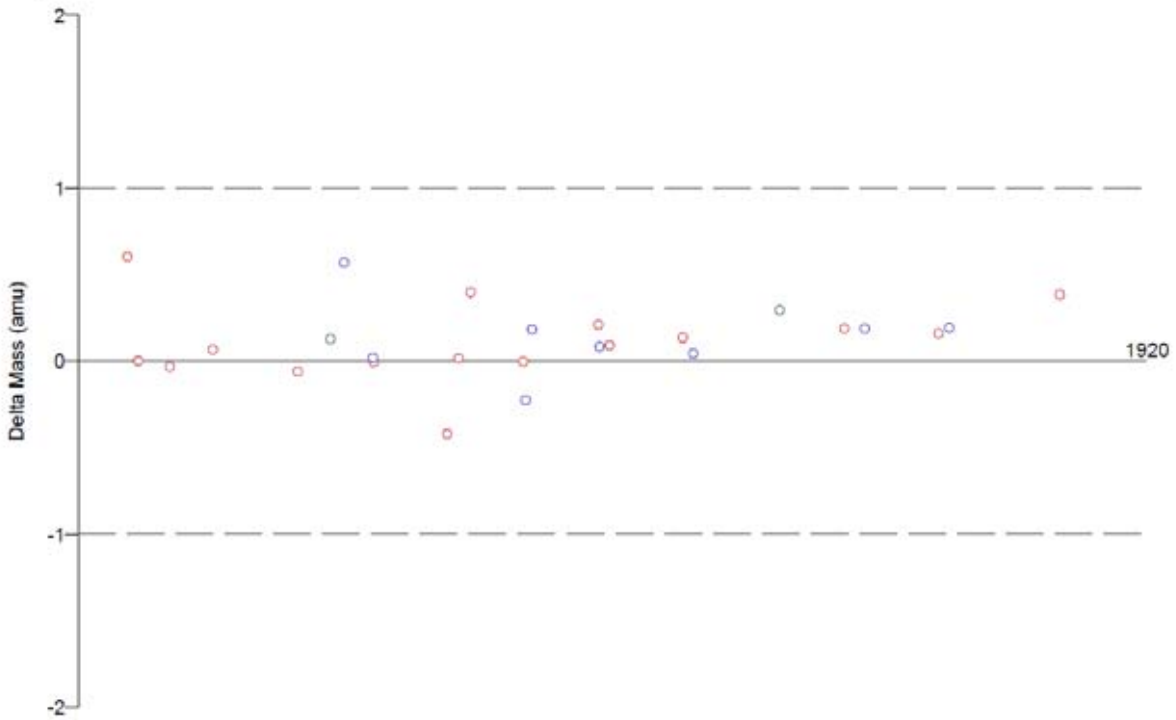
Protein Coverage:

Sequence	MH+	% Mass	AA	% AA
ISALRVPAAGLIGAIPPNTLGR	2157.29	3.13	69 - 90	3.47
YLNIEDELVQMLQLAMACTSRSPERR	3066.52	4.45	573 - 598	4.10
Totals:	3066.52	4.45	26	4.10

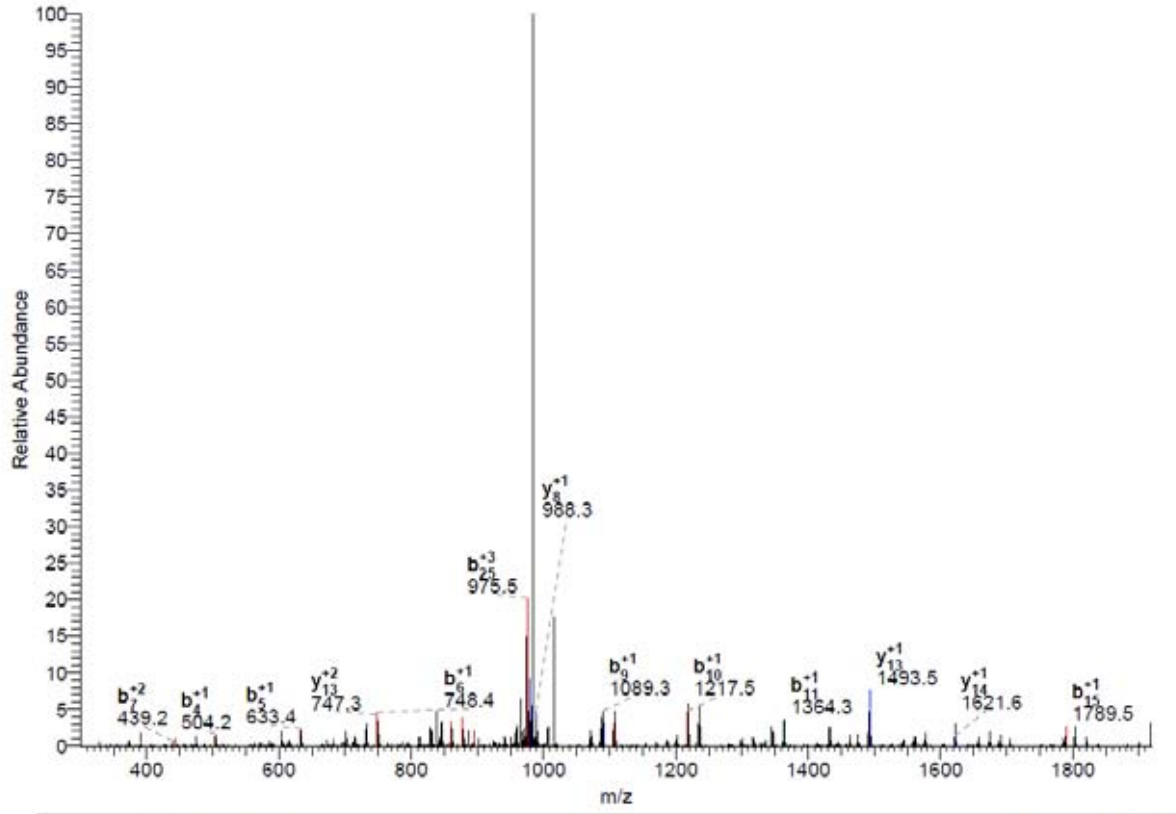
DTA for scan: 8246
Precursor ion: 1033.74
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
Y		164.07							
L		277.15				2935.44			
N		**391.20**				2822.36			
I		**504.28**				2708.32			
E		**633.32**				2595.23			
D		**748.35**				2466.19			
E		**877.39**				2351.16			
L		990.48				2222.12			
V		**1089.55**				2109.04			
Q		**1217.60**				2009.97			
M*		**1364.64**				1881.91			
L		1477.72				1734.87			
Q		**1605.78**				**1621.79**			
L		1718.87				**1493.73**			
A		**1789.90**				1380.65			
M*		1936.94				1309.61			
A		2007.98				1162.57			
C		2110.99				**1091.54**			
T		2212.03				**988.53**			
S		2299.07				887.48			
R		2455.17				800.45			
S		2542.20				644.35			
P		2639.25				557.32			
E		2768.29				460.26			
R		2924.40				331.22			
R						175.12			



#8246-8246 RT:46.43-46.43 NL: 2.26E3



Reference Scan(s)	Sequence	MH+	z	P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein retrotransposon protein, putative, Ty3-gypsy subclass				0.0004	6.2	0.0	0			
9690	R.VQKLLY#GVLIT#VR.K	1661.88	2	0.0004	2.357	0.078	343.0	2	17/48	

1 of 5 peptide matches reported, 4 removed due to filtering

Reference: LOC_Os03g36290.1|13103.m03970|protein retrotransposon protein, pu
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 1120 Monoisotopic MW: 130070.7 pI: 8.98

Protein:

MPFGLEIAGA TYQRMIQRCF STQIGRNVEA YVDDVVVKTK QKDDLIADLE
ETFANIRAFR MKLNPEKCTF GVPSGKLLGF MVSRRGIQAN LKKINTILNM
KPPSSQKIFD VQKLTGCMAA LSWFISRLGE WGMPFFMLLK KTDNFQWGPE
AQQAFDDFKK LLTTPPVVST VLVVEREEDG HVQKVQRLIY FVSEVLVDSK
TRYPRVQKLL YGVLITVRKL SHYFQSHSVT VVTSFSLEDI LHNHEANGRI
AKWALELMSL DLSFKPRTSI KSQALADFVA ECQEDTPVEK LEYWTMHFDG
SKRLLVHEYL NKPTQLARYA ASDIQDEKEK IERFIEGLRD ELRGQMISQD
HESFQSLVNK VVKLENDHRM VEGLRKRRMN LQRQSPTRNS LQVRRPYPFK
PSLPNARPGI APILPRPRLK MRDSPCYVCV KVGHFARLCP NRQNTGGQTS
YQKPSISGFI KKNPPIMKSG PNLGRSQVHH IKAEEAQQDL NILMEVPVVQ
EYLDVFPEEL SELPPDWEIE FIIELLPGTA PISKRPYKMP VNELEELKRQ
IVELEAKGFI RQSSSLWGAF VLFVKKKDGS LRMCVDYRAL NEVTIKNKYP
LPRIDDLFDQ LKGASVFSKI DLRSGYHQLK IRQEDISKTA FVTRYELYEF
TVMAFELTNA PAYFMNLMNK IFMEYLDQFI VVFIDDILIY SQNEEEHKQH
LRLVLEKLRQ QQLYAKFMLI LPDIHKDPEI YCDASRQGLG CVLMQEDLNL
QQRRWLELIK DYNLEILYHP RKANVVADAL SQKAYCNLQQ EIISEQLQAE
MLKLNLGIIK YGELNTLELK PTLMDQIREY QKDDPEIQIP KERKVQGKAA
DITEDLEGIL WYGNIIIVPQ SGNLREPLQI HEWKWDEIGM DFITGLPKTS
TGYDSIGVII DRLTKTARFI HVKTTYTGAQ LAELYMTRIV CLHGVPKRII
SDRGTQFTSH FWQKLHYELG SYLDFNTAFH PQTGGQIERL NQTLEVVLRA
CALDFSASWD TCLPYAEFSY NNNCQASLKM SRPKALYGRK CLTPLIWTKT
RERSIFGTDF VKEAEENVQL IRDRLKTTQS RQKSYADSRR RELTFVIGDY
VYPKISSLRR IKRFKKKKYQ

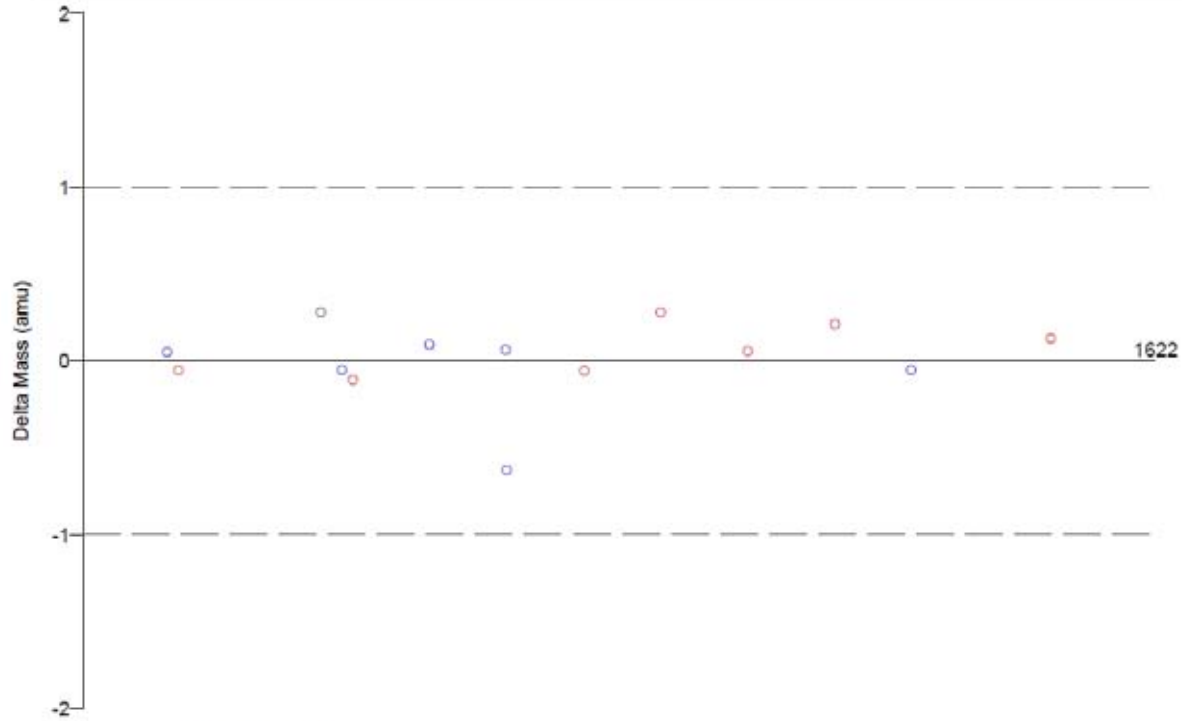
Protein Coverage:

Sequence		MH+	% Mass	AA	% AA
VQKLLYGVLITVR		1501.95	1.15	206 - 218	1.16
VQKLLYGVLITVRK		1630.04	1.25	206 - 219	1.25
	Totals:	1501.95	1.15	13	1.16

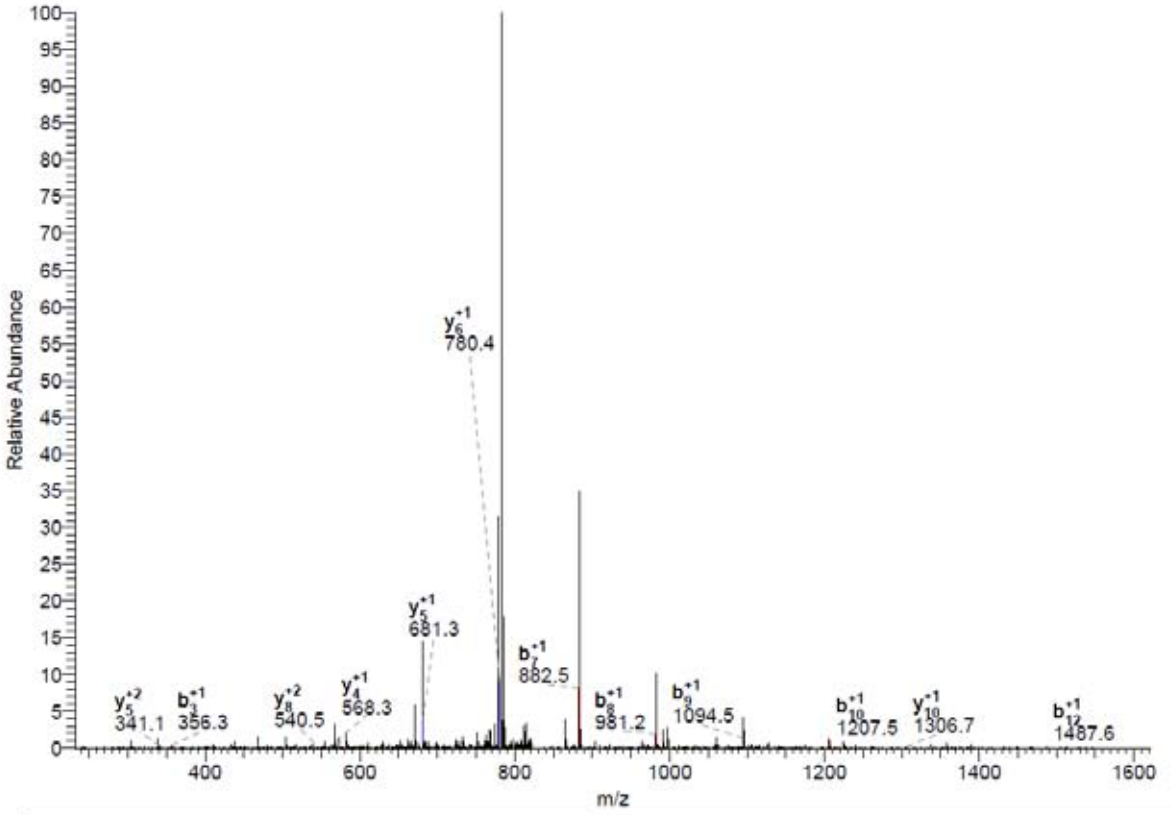
DTA for scan: 9690
Precursor ion: 831.70
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
V		100.08							
Q		228.13				1562.81			
K		**356.23**				1434.75			
L		469.31				**1306.66**			
L		**582.40**				1193.57			
Y#		825.43				1080.49			
G		**882.45**				837.46			
V		**981.52**				**780.44**			
L		**1094.60**				**681.37**			
I		**1207.69**				**568.29**			
T#		1388.70				455.20			
V		**1487.77**				274.19			
R						175.12			



#9690-9690 RT:54.58-54.58 NL: 3.20E3



Reference Scan(s)	Sequence	MH+	z	P P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein dehydration response related protein, putative, expres				0.0007	6.1	0.0	0			
8375	R.TRTAS#PCGRS#R.W	1351.53	2	0.0007	2.320	0.195	128.1	5	13/40	

1 of 4 peptide matches reported, 3 removed due to filtering

Reference: LOC_Os10g37770.1|13110.m03435|protein dehydration response relate
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 404 Monoisotopic MW: 42248.8 pI: 12.25

Protein:

MPNAGAARRC GRATHVDLLT VVLAAMLCWA SYTLSIWHNS RGAADSSVLG
LVVGATVCGD ADEELDFEAR HAADDAGLSV SSGPANSRVR RALSSSGPAP
AAAGTTVSRY RAPFPWPASR GVVWAGNSAR GAKAAADAAA AANKWARVDG
DMLRFTDAAA VRAYAYVVLR LVAAPVRAAV DVGAMHGGSW AAELMSRGVV
TVSVAAPWGA SDGAALVELA LERGVPAVLA AAGGAPSRRL PFPAGAFDMA
HCGRCLVPWH LHGGRFLMEI DRVLRPGGYW VHSARRRTAR MSAPPSRPPR
RACAGGASRT RTASPCGRSR WATSAATPAK TLPGSAPAKT RSSNGTATSS
RASRRSKKAR RRHGKPRRRR RSVATARRGR GGWRGTRRWP PSWGRRAGSA
TCST

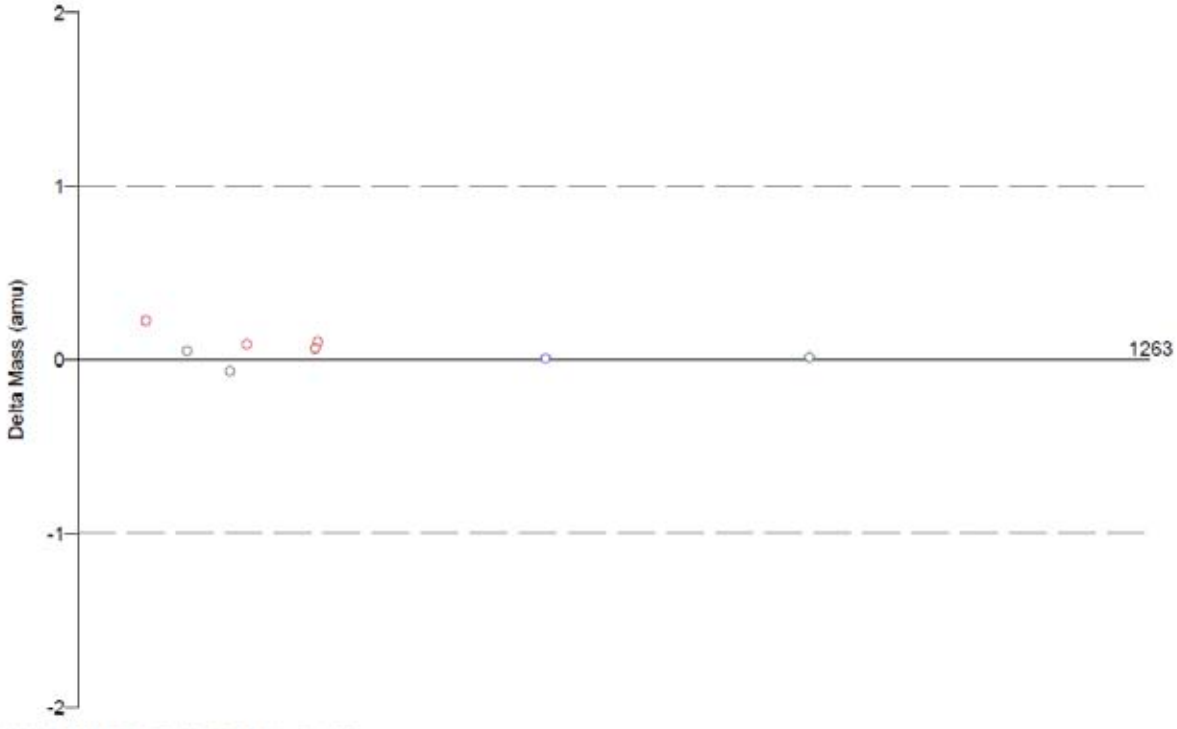
Protein Coverage:

Sequence		MH+	% Mass	AA	% AA
TRTASPCGR		948.47	2.24	310 - 318	2.23
TRTASPCGRSR		1191.60	2.82	310 - 320	2.72
TRSSNGTATSSR		1224.59	2.90	340 - 351	2.97
	Totals:	1191.60	2.82	11	2.72

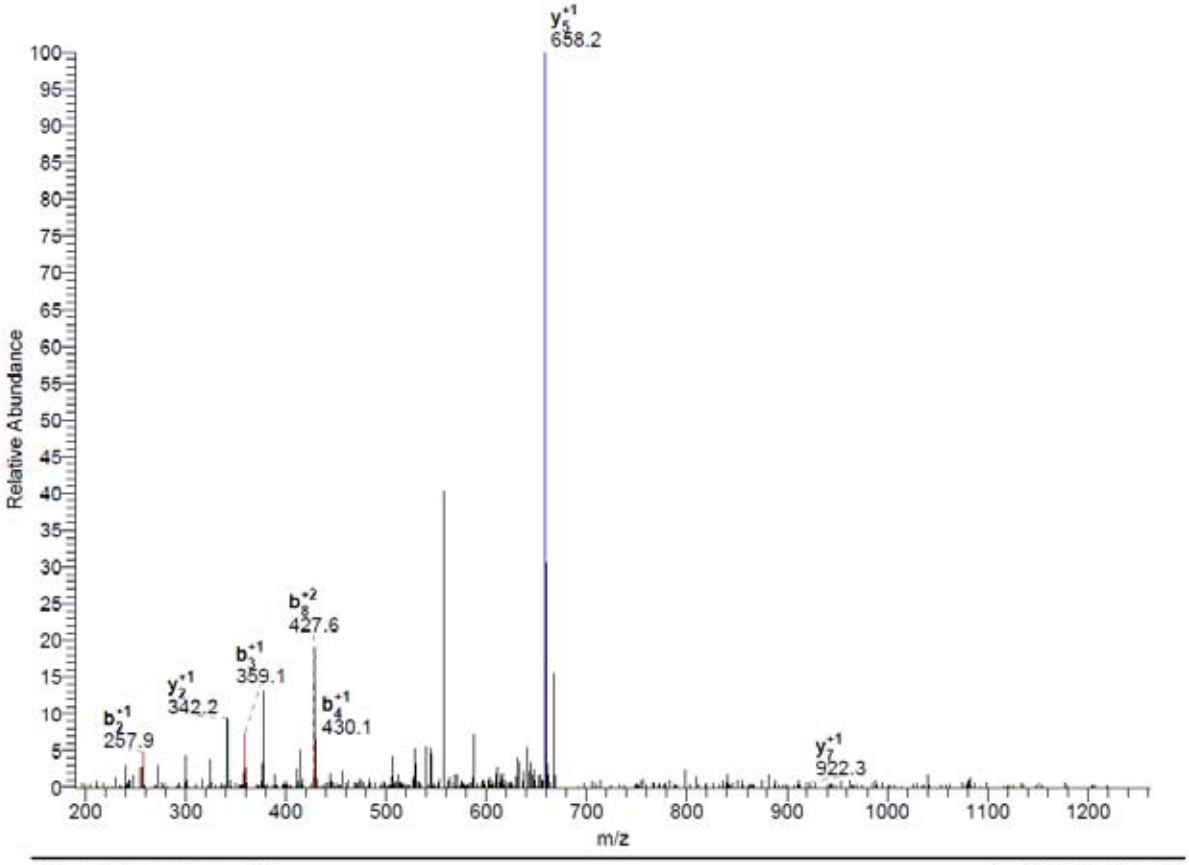
DTA for scan: 8375
Precursor ion: 676.49
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
T		102.05							
R		258.16				1250.49			
T		359.20				1094.38			
A		430.24				993.34			
S#		597.24				922.30			
P		694.29				755.30			
C		797.30				658.25			
G		854.32				555.24			
R		1010.42				498.22			
S#		1177.42				342.12			
R						175.12			



#8375-8375 RT:47.17-47.17 NL: 7.12E2



Reference Scan(s)	Sequence	MH+	z	P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein DnaK family protein, putative, expressed				0.0003	4.1	0.0	0			
8236	R.FDDPQTQKEMKMVPY#K.I	2064.92	2	0.0003	2.640	0.138	121.6	4	15/45	

1 of 2 peptide matches reported, 1 removed due to filtering

Reference: LOC_Os03g02260.1|13103.m00178|protein DnaK family protein, putati
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 676 Monoisotopic MW: 72521.6 pI: 5.60

Protein:

MAASLLLRAV RRRDLASPLG TLTANVQSKC AANVCSRWAG FARTFSAKAT
GNEVIGIDLG TTNSCVSVME GKNPKVIENS EGTRTTPSVV AFNQKGELLV
GTPAKRQAVT NPQNTFFGTK RLIGRRFDDP QTQKEMKMVP YKIVKALNGD
AWLETTDGKQ YSPSQIGAFV LTKMKETAES YLGKSVSKAV ITVPAYFNDA
QRQATKDAGR IAGLDVQRII NEPTAAALSY GTNNKEGLIA VFDLGGGTFD
VSILEISNGV FEVKATNGDT FLGGEDFDNT LLEFLVSEFK RTEGIDLSKD
RLALQRLREA AEKAKIELSS TAQTEINLPF ITADSSGAKH LNITLTRSKF
ESLVNSLIER TRDPCKSCLK DAGITTKDVD EVLLVGGMTR VPKVQEVVSE
IFGKAPSKGV NPDEAVAMGA AIQGGILRGD VKDLLLLDVT PLSLGIETLG
GIFTRLINRN TTVPTKKSQV FSTAADNQTQ VGIKVLQGER EMAADNKLLG
EFDLVGIPPA PRGMPQIEVT FDIDANGIVT VSAKDKATGK EQQITIRSSG
GLSEAEIQKM VHEAELHSQK DQERKALIDI RNTADTTIYS IEKSLGEYRD
KIPAEVASEI ETAIADLRNE MASDDIEKIK SKIEAANKAV SKIGQHMSGG
GSGGSQAGSQ GGGDQAPEAE YEEVKK

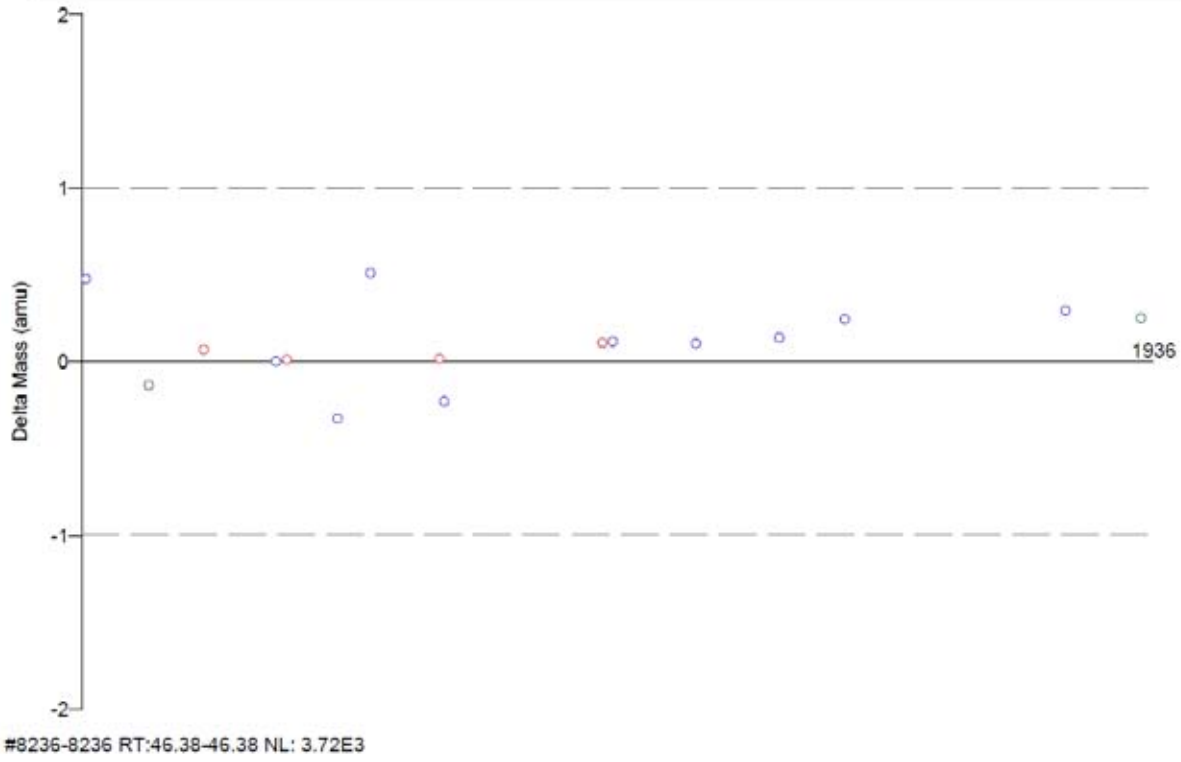
Protein Coverage:

Sequence		MH+	% Mass	AA	% AA
FDDPQTQKEMKMVPYK		1984.95	2.74	127 - 142	2.37
DPCKSCLKDAGITTK		1579.78	2.18	363 - 377	2.22
	Totals:	1984.95	2.74	16	2.37

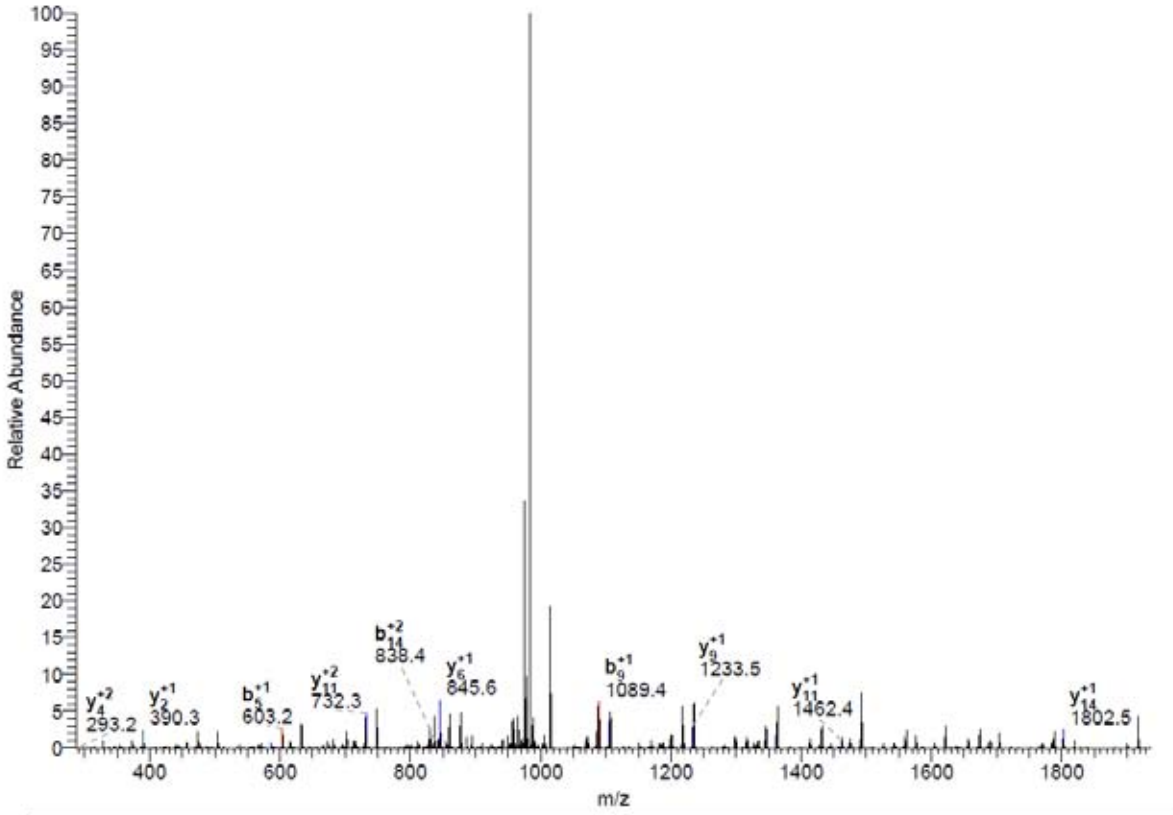
DTA for scan: 8236
Precursor ion: 1033.26
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
F		148.08							
D		263.10				1917.85			
D		378.13				**1802.82**			
P		**475.18**				1687.79			
Q		**603.24**				1590.74			
T		704.29				**1462.68**			
Q		832.35				**1361.64**			
K		960.44				**1233.58**			
E		**1089.48**				**1105.48**			
M		1220.53				976.44			
K		1348.62				**845.40**			
M		1479.66				717.30			
V		1578.73				**586.26**			
P		1675.78				487.20			
Y#		**1918.81**				**390.14**			
K						147.11			



#8236-8236 RT:46.38-46.38 NL: 3.72E3



Reference Scan(s)	Sequence	MH+	z	P P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions	Count
protein expressed protein				0.001	4.1	0.0	0			
11377	R.NPAY#EKEIPAS#S#TCMDK.K	2173.77	2	0.001	2.002	0.154	136.8	25	12/80	

1 of 3 peptide matches reported, 2 removed due to filtering

Reference: LOC_Os02g03840.2|13102.m12354|protein expressed protein
Database: C:\Xcalibur\database\Rice_database\all.pep.fasta
Number of Amino Acids: 3052 Monoisotopic MW: 337130.1 pI: 3.85

Protein:

MATEEEAREA TDGQEVQVEA ATDLQAQLDS KLSSIKEIKG EAQTMGSTEG
ILKIPEDEVL VEAPSETKIP SEHSLNGTAS SLNGHVNKEE NISNEQPHEI
NQESQEQVET SSDGTSTYPS NKSNEEEMTD SLSHGESASE DTTILKCENA
EAKDNGQQDV EGHVTNDTMV QEEICKTDDA IEQTVDAQDL DSSKESKEAQ
PAIMTDIPAD EVLAEAPAGI QSPLEPNVGD SDTVPGNKET DDPAKEDDTA
YLDHKESFPE DKVIAEHADE EVKAEDQQSK QADDMDAEVL QEEIPESKES
DLPETEQAEE ASVDDQEALN QQPALETNDS LSVKAEETCH QSNVATCGEK
TPEDDATTRE PTVDTKEEQN QGSVEEMKDA EAVDTEETVQ QSSVAFDEAI
QDHAATTNPS SDIQSIEPEE TKGPSDVKAE EVSSQSNVAF AEDAVQDKVI
PSEPPEDIRP VKKLEQEETK EADEASNETN SAIFSDLNQE DSIVASKQLE
TELAEEAAAF SNLDQEGGIA DSESQVADLE EAKEIEATET EEITHQPPAA
VSTELPKEDN STRSEPHNDD IQHSLEQDSI EVKDTEAAEI QEISQERTIA
TSKEDAVEDD GTAEGPTCVS QEVQDVESEE VKDTEPDNVV EASDVVTVDD
EGQENNVLTS ENIAELQLQG LESEEIKSPE PIETEAGFHT SHAAPSNDPV
EVNTTACETQ DTESAEEIKE TEGTKTESIP QESNISVSEE SNPEDSITAS
ETISNTQELT IAGSIEASED NIDTKTGEIT DQSNEVPAGE AAQGDNIPES
VSTADTQSMQ ELESSEMKKP ELVDLSGTPH QKDDAISQKQ NQEDNPTTCE
TNEIGSTEVS SVEASEDQAI AHQSNITQCE EQATEESITE SEPQILEIES
VQDMEDTEAT EPELVSEQNI VSTSEESVPE ENATTEEPAF HDREIQNDGA
ELTEQHDSVK AEELPNQSSG AIVEETAQEA DLVAGEPIDD VQEKDLEPEE
ISNTVDGETG EASHQTHAAV EDNWTGEVES SVEASEDQAI AHQSNITQCE
EQATEESLTE SEPQILEMES VQDIKDTEAT ESETIFQKNI VPTSEESVPE
ENATAKEPAF DDREIQNDGA ELTKEHDGVK DEEIPDQSSG AIVEETAQEA
NLLASEPTDD VQVKELEPEE ISNTVDGETE EASCQTHEVE MTESSEQMKD
TEPPVPEPES SEEMRDTAPT VPEPTLQDSR VASVEEIETH NNATIEQNVG
YQQLQEQESV EFKETEVLEP QGVIPSHNVS SSEEFNPQET VTKEEPGSDT
QAEESPVVIE DTEDVNNSAA LSEKIAPEEH VLATETTVDT PPVQETELEE
RQNIESVEAE DNIAASGLPG EEIDTEAIEI ESVPHESTTT SVKELNEDVK
SNVALAEEAA SEEHILETEV TVDKSSAQEP ELEEIKNTEP AEQEHNITET
GLPEEEMKDN EAMETETVPH DSNIESIKEL REDDIITASA LDVYTQQIPE
QESVEDMKCT DTTEHPGETP ESIVSTSDEL TKTGEITTVK EATFDTQQAQ
SFTDQETDLS SSSNPEEAVQ ESDLVKSEQE TDVQQEQELD STEETKGTED
YQQNGVSTCE ESVTEVEPNV DDQHVQENKS AVEVKENEDT ETEEISKQIN
PTTSENGAQE SSEQEADQPF YVQPVQQLEL TTDSKDNQLV EAEETSSQSN
IVTPEDPTAE DRVAYEIDPS VDIDQGHELE LVEEVKDTDA IEAEETSHAG
QAVSSAEKFS ESNLSAVELT HGIQQVHYLE TMEEMKGTEG TCDEEICYEQ
TATSEDPSPT DNGKSLQDYH VESNEENLGN GIGDVISVHE KIEDNIHESA
ELKDLQDNSL VYFIQDETSE LGEKTQISAN RIEENDDHIS TEDTEETSNN
TNLVKEGPNE HGSSQTSNVQ DNKQLHDVGL QTQVRERSVD IGQQDEDMKN
VNLDQQQKED EEIEKQKEDE EIEKQKEELQ TDEQKHDDKR VDFIIDTQVE
SIDAFQAEQT DSVVTEMLND EVTQHEPEDS IPRTTDAMVE NITEIKEETE
EENGPNSGGT LEVSAKNYNE DVHENTEKDA VVEKTSSSEH DEIAGEIRNE
EVEPCLASSL ERDLQVDSDL SNDQMLENNP IAVPQNDEYM YRAEEGYTNK
VNVDMHAIQE SDKVIEDAEE KQGMQNEDNV VHHDESLVTT QKEEASQVHT
DEQYSADTKM DDTAISYAEM THENTSTEPR EVEDTKEKKG FNDFPEFVVE
TSKQDDVDQD FSIHHQVEDE KSAETENNSA ESEAVQPKLD ITIAETNNDN
NLSTINPLPE HETENASDIN QNRQYQEATN EDAIDNIETG RVEKMETSYT
ATTEVVTLNE DICDKASGAD GVPPDGSLKT SEDNLDVSSV VTESEGENIN
KETEDHKLAL PVHPTQDENT TEQGFGLEDT EKESMSPEKA LPAEPEEQEE
NQVTKEQDEE DRHDAELGDA HEEDHKEAEQ DYLPVSSFLM NLILGKDNDD
PKKDSETEVE KEQEETTKDG SCLIASQQEE NLVAFPIEKS VDEKLTFEQE
KEKVEGSEET KEPVKEQSYD VEMDVQKSLE TDEELKRNTC DLEAPVYQDN
AQDEISSKLI SAKAADPIKK MEARDFELDE ESFDTVCQEN VEAATEIEDG
SLNRDQDDIT SPKASQEDAL EEVGTELPHE SLHENRHGAK DEQTLSLIEP
DTGNAEKLPN EADSVQSPPC TEQEESIESS YVEVRSTTEG QVESEVIETN
EEDQNTTAGG HTEEQIENLH DDKSKETCSE EISDEQAPEI TEPVSNTDRN
FAYEKEIPAS STCMDKKESM ISNNEVSNFE KALETHSDSP NLHVNQDKKD
ETADNQTVVD HNTVLDKLED SNRQEEQETA AQKLPKETEG NQEFMAITEP
VIKEENVHET VEGNTQAVKI KSNEEKELFD SQVQERGLNV VSPKATSEAD
EKFVEITKPE FSTDEEHSPK ADESNKPDEN TCDEKTKAEE ETNNITDEAT
VKIEERGAEQ KVSHKKHNIL SGVGSKVKHQ LAKVKKAIIG KPGHTKSESP
KS

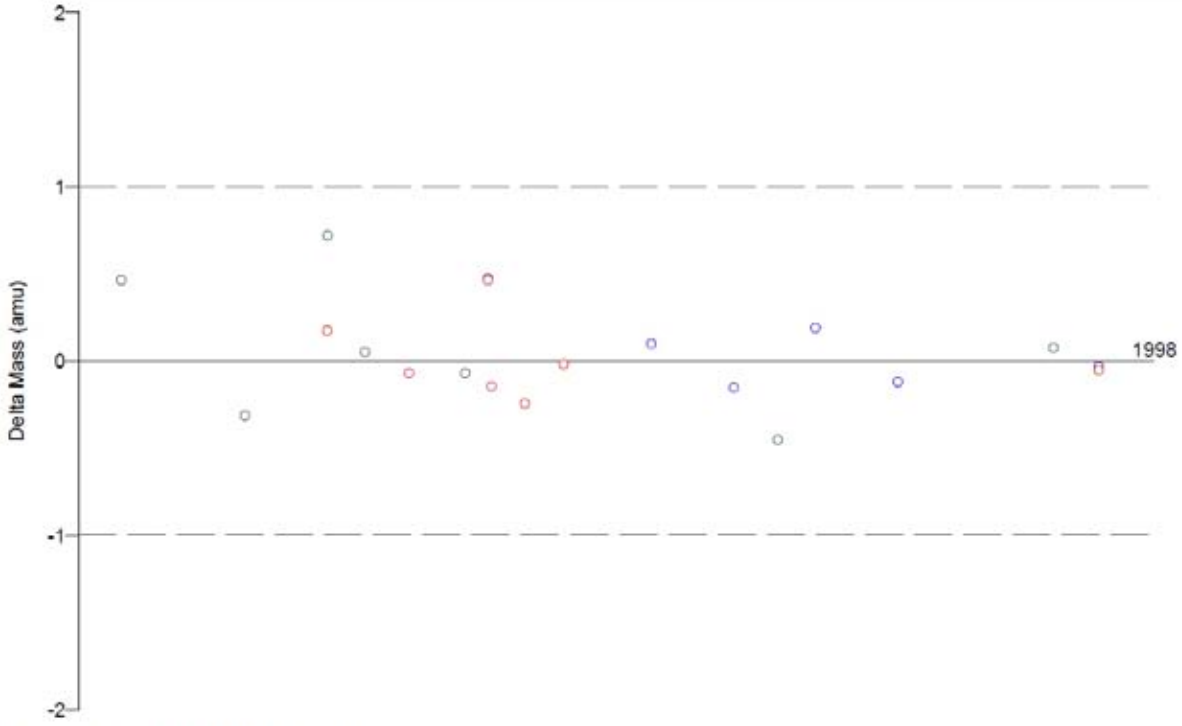
Protein Coverage:

Sequence		MH+	% Mass	AA	% AA
IPSEHSLNGTASSLNGHVNK		2062.03	0.61	69 - 88	0.66
AADPIK		614.35	0.10	2614 - 2619	0.20
NPAYEKEIPASSTCMDK		1933.87	0.57	2800 - 2816	0.56
	Totals:	1933.87	0.57	17	0.56

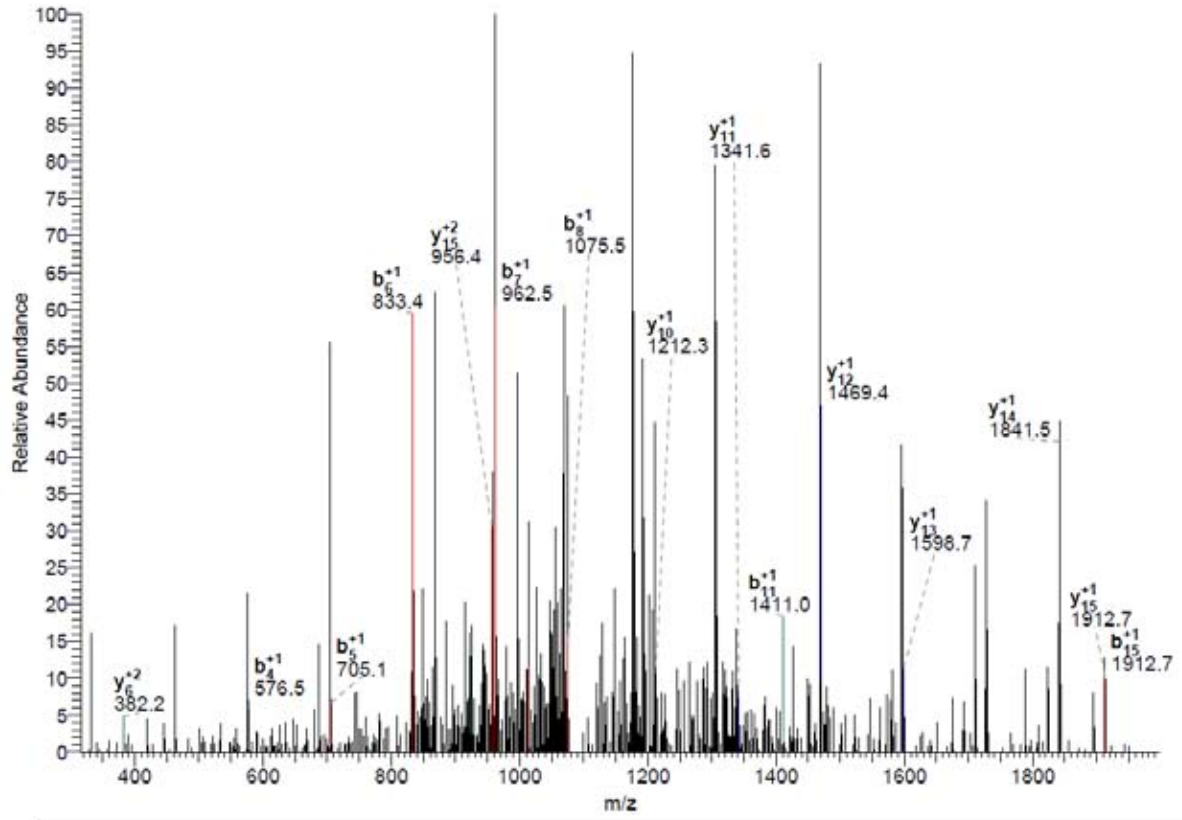
DTA for scan: 11377
Precursor ion: 1086.99
Mass type: Monoisotopic
Mod's: (M* +15.99492) (STY# +79.96633)

Ion series for charge: +1

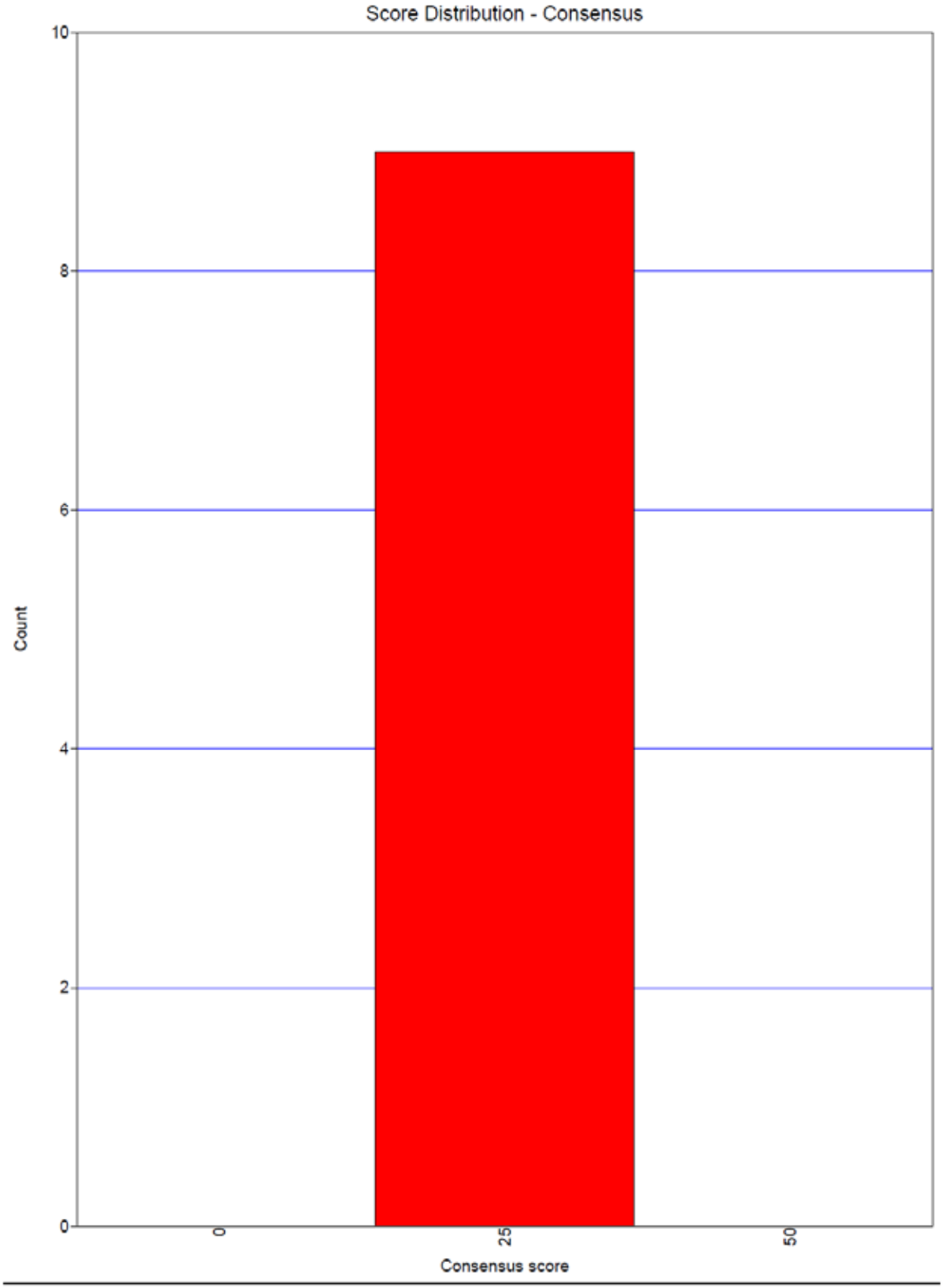
AA	A ions	B ions	B* ions	Bo ions	C ions	Y ions	Y* ions	Yo ions	Z ions
N		115.05							
P		262.12				2059.72			
A		333.16				**1912.65**			
Y#		**576.19**				**1841.62**			
E		**705.23**				**1598.59**			
K		**833.32**				**1469.55**			
E		**962.37**				**1341.45**			
I		**1075.45**				**1212.41**			
P		1172.50				1099.32			
A		1243.54				1002.27			
S#		**1410.54**				931.23			
S#		1577.54				**764.24**			
T		1678.58				597.24			
C		1781.59				496.19			
M		**1912.63**				393.18			
D		2027.66				262.14			
K						147.11			



#11377-11377 RT:64.11-64.11 NL: 1.33E2



BioworksBrowser rev. 3.3 25/08/2009 13:07:36


Score Distribution - Consensus
10
8
6
4
2
0
Count
0
25
50
Consensus score

```
[SEQUEST]
first_database_name = C:\Xcalibur\database\rice_all.pep.fasta.hdr
second_database_name =
peptide_mass_tolerance = 1.0000
peptide_mass_units = 0                  ; 0=amu, 1=mmu, 2=ppm
ion_series = 0 1 1 0.0 1.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0
fragment_ion_tolerance = 0.5000           ; for trap data leave at 1.0, for accurate mass data use values < 1.0
fragment_ion_units = 0                  ; 0=amu, 1=mmu
num_output_lines = 10                    ; # peptide results to show
num_results = 250                         ; # results to store
num_description_lines = 5               ; # full protein descriptions to show for top N peptides
show_fragment_ions = 0                  ; 0=no, 1=yes
print_duplicate_references = 0          ; number of duplicate references reported
enzyme_info = Trypsin(KR) 1 1 KR -
max_num_differential_per_peptide = 3            ; max # of diff. mod in a peptide
diff_search_options = 15.994915 M 79.966331 STY 0.000000 M 0.000000 X 0.000000 T 0.000000 Y
term_diff_search_options = 0.000000 0.000000
nucleotide_reading_frame = 0            ; 0=protein db, 1-6, 7 = forward three, 8-reverse three, 9=all six
mass_type_parent = 1                    ; 0=average masses, 1=monoisotopic masses
mass_type_fragment = 1                  ; 0=average masses, 1=monoisotopic masses
normalize_xcorr = 0                     ; use normalized xcorr values in the out file
remove_precursor_peak = 0               ; 0=no, 1=yes
ion_cutoff_percentage = 0.0000             ; prelim. score cutoff % as a decimal number i.e. 0.30 for 30%
max_num_internal_cleavage_sites = 2              ; maximum value is 12
protein_mass_filter = 0 0               ; enter protein mass min & max value ( 0 for both = unused)
match_peak_count = 0                    ; number of auto-detected peaks to try matching (max 5)
match_peak_allowed_error = 1            ;  number of allowed errors in matching auto-detected peaks
match_peak_tolerance = 1.0000              ; mass tolerance for matching auto-detected peaks
partial_sequence =
sequence_header_filter =
digest_mass_range = 600.0000 3500.0000

add_Cterm_peptide = 0.0000                ; added to each peptide C-terminus
add_Cterm_protein = 0.0000                ; added to each protein C-terminus
add_Nterm_peptide = 0.0000                ; added to each peptide N-terminus
add_Nterm_protein = 0.0000                ; added to each protein N-terminus
add_G_Glycine = 0.0000                    ; added to G
add_A_Alanine = 0.0000                    ; added to A
add_S_Serine = 0.0000                     ; added to S
add_P_Proline = 0.0000                    ; added to P
add_V_Valine = 0.0000                     ; added to V
add_T_Threonine = 0.0000                  ; added to T
add_C_Cysteine = 0.0000                   ; added to C
add_L_Leucine = 0.0000                    ; added to L
add_I_Isoleucine = 0.0000                 ; added to I
add_X_LorI = 0.0000                       ; added to X
add_N_Asparagine = 0.0000                 ; added to N
add_O_Ornithine = 0.0000                  ; added to O
add_B_avg_NandD = 0.0000                  ; added to B
add_D_Aspartic_Acid = 0.0000              ; added to D
add_Q_Glutamine = 0.0000                  ; added to Q
add_K_Lysine = 0.0000                     ; added to K
add_Z_avg_QandE = 0.0000                  ; added to Z
add_E_Glutamic_Acid = 0.0000              ; added to E
add_M_Methionine = 0.0000                 ; added to M
add_H_Histidine = 0.0000                  ; added to H
add_F_Phenylalanine = 0.0000              ; added to F
add_R_Arginine = 0.0000                   ; added to R
add_Y_Tyrosine = 0.0000                   ; added to Y
add_W_Tryptophan = 0.0000                 ; added to W
add_J_user_amino_acid = 0.0000            ; added to J
add_U_user_amino_acid = 0.0000            ; added to U
```

Appendix B2

Figure 1 – *KK01* candidate full-length ORF sequence

ATGGCTGCGTCCCCTACCAGCTCGCGCTCCGTGACGGAGACGGTGAACGGCTCCCACCGCTTCGTG
ATCCAGGGCTACTCCCTCGCCAAGGGGATGGGCGTCGGGAAGCACATCGCGAGCGAGACCTTCACC
GTCGGGGGATACCAGTGGGCTATCTACTTCTACCCAGACGGGAAGAACCCCGAGGACAACTCGGCC
TACGTTTCCGTCTTCATCGCGCTCGCTTCCGAGGGCACTGACGTCCGCGCTCTCTTTGAGCTCACA
CTGCAGGACCAGAGCGGCAAGGGCAAGCACAAGGTCCACTCCCACTTCGATCGCTCGCTCGAGTCT
GGCCCCTACACCCTCAAGTACCGAGGCTCTATGTGGGGTTACAAACGGTTCTTTCGGCGAACTGCC
CTTGAGACATCGGACTTTCTTAAAGATGATTGTTTGAGGATAAACTGCACTGTGGGTGTTGTGGTT
TCAACTATTGATTACTCCAGACCACACTCCATTCAGGTTCCAGACTCTGATATTGGCTACCACTTC
GGTTCTCTTTTGGACAGTAATGAAGGGGTTGATGTTATTCTTAATGTGAGTGGAGAGAGGTTTCAT
GCCCATAAGTTGGTGTTGGCTGCACGGTCTACTGTATTTAGATCCAAACTTTTTGACGATGAATCA
GAAGGAGACAAGAACGAGGTTAATGAGAGTGAGGATCTGAAGGAGATTGTTCTTGATGATCTAGAG
CCCAAGGTTTTCAAGGCAATGCTTCATTTCATCTACAGAGATACCCTCGTTGATGATTATGAGTTG
GATGCATCAAGCTCCATGGGCTCTATTTTTGATACTTTGGCGGCAAAGTTGTTGGCCGCAGCAGAC
AAGTATGACTTAGGAAGGTTAAGATTGCTATGTGAATCTTACTTGTGCAAAGGCATTTCTGTGGCC
TCGGTTGCAAATACTCTTGCAACTGGGGAGAGTCACCACGCCATGGAGCTTAAAGCCGTTTGCCTA
AAATTTGCTGCAGAAAATCTTTCAGCTGTGATCCGGACTGACGGGTTCGATTACCTCAAGGATAAT
TGCCCATCACTTCAGTCGGAGATACTGAGAACTGTTGCTGGGTGTGAGGAGCCGTGCAGTAGTGGC
GGGAAAAGCCAGAGTGTATGGGGACAGCTCTCGGACGGTGGCGACACTAGTGGCCGTAGGGTGAGG
CCAAGAATCTGA

Figure 2 – *KK03* candidate full-length ORF sequence

ATGGGCGACTCTGGCGGCTCCGTCGTGTCTATTGACGTCGAGCGCATCTCCTTCGGCGGGAAGGAA
CATCATATACAAACAAAACATGGACCTGTATCTGTTGCTGTGTATGGTGACCATGATAAGCATGCC
CTTATTACTTATCCAGATGTTGCTTTGAACCATATGTCTTGCTTCCAAGGACTACTATTCTGCCCG
GAGGCTGCCTCACTGCTGCTCCATAATTTTTGCATTTACCATATAAGCCCCCCTGGACATGAGTTA
GGAGCCACTCCGATTTTACCAAACAGTCCTGTGGCATCAGTTGATGAGTTAGCGGATCAGGTTGCA
GAAGTACTTGATTTCTTTGGATTGAGTTCTGTTATGTGCTTAGGTGTCAGTGCTGGTGCATACATT
CTTACTCTCTTCGCAACAAAGTATAGGGAGCGTGTGCTAGGTCTTATCCTCGTTTCACCTCTATGT
AGAACTCCCTCATGGACTGAGTGGTTTTATAACAAGGTGATGTCGAATTTGTTGTATTATTACGGG
ATGTGTGACGTTGTGAAGGACTGTTTGTTGCAGCGTTACTTTGGCAAGAGAGTGCGAGGAGGCTCT
GCTGTACCAGAATCTGATATTATGCAGGCGTGTAGAAGTTTCCTAGATCAACGGCAAAGCATGAAT
ATATGGCGGTTTATCCACACAATCAATGAGAGACATGACTTGACAGAAAGCCTGAAGCAGCTGCAG
TGCAGAACTCTGATATTTGTCGGTGAGAACTCGCAGTTCCACACTGAAGCTGTTCACATGACTGCA
AAGCTTGATAAGAGATACAGTGCTCTTGTTGAGGTGCAAGCGTGTGGGTCCGTCGTGACGGAGGAG
CAGCCGCACGCGATGCTGATACCAATGGAGTACTTCCTCATGGGATACGGCCTGTTCAGGCCGAGC
CATGTAAGCTCCAGCCCTCGCAGCCCTTTGAACCCGTTCTGCATATCGCCGGAGCTCCTCTCGCCC
GAGAGCATGGGTGTGAAGCTGAAGCCAATCAAGACGCGAGCCAATCTCAGGGTTTAG

Figure 3 – *KK04* candidate partial coding sequence (including 3′ untranslated region)

```
TATGGGGAGCATGCGGAGCACCTTGTTTCCTAACTTTGGCAGCATGTTTAGCGTGGCGGAACAGCA
GCAGGCTAAAGCTGACTGGGATGCTGAGAGTCATAGGGATGATGAAGATTATGCATCGGATCATGG
TGCTGATGACATTGAGGATAGCCTCAATAGCCCGCTTATTTCTCGTCAAGCGACAAGCGTGGAGGG
TAAGGAGATTGCTGCACCTCATGGAAGTATAATGGGTGGTGTGGGAAGAAGTAGCAGCATGCAGGG
AGGGGAGGCAGTAAGCAGCATGGGCATTGGTGGGGGGTGGCAGTTAGCTTGGAAGTGGACTGAGAG
AGAAGGTGCAGATGGGCAAAAGGAAGGCGGCTTCCAGCGTATTTACTTGCATGAGGAGGGCGTGTC
AGGTGATCGGAGGGGCTCTATATTGTCTATGCCAGGAGGTGATATTCCTCCTGGTGGTGAGTATAT
CCAGGCAGCCGCTCTAGTGAGCCAACCTGCTCTTTATTCGAAGGACCTGATAGAGCAGCAGCTTGC
TGGTCCTGCCATGGTACATCCATCTGAGGCAGTTGCCAAGGGTACAAAATGGGCAGAACTATTTGA
ACCTGGAGTGAAGCATGCACTGTTCGTTGGCATAGGATTACAGATCCTGCAGCAGTTTGCGGGTAT
CAATGGAGTCCTCTACTACACACCTCAGATACTTGAGCAAGCAGGTGTCGGGGTTCTTCTATCAAA
CATTGGACTAAGCTCTTCCTCTGCATCTATTCTTATTAGTGCCTTGACAACCTTGCTGATGCTTCC
CAGCATTGGCATCGCCATGAGACTCATGGATAGGTCAGGAAGAAGGTTTCTTCTCCTTTCGACAAT
CCCTGTCTTGATAGTAGCGCTAGCTGTCTTGGTTTTGGTGAATGTTCTGGATGTCGGAACCATGGT
GCACGCTGCGCTCTCAACGATCAGCGTCATCGTCTATTTCTGCTTCTTCGTCATGGGGTTTGGGCC
TATCCCAAATATTCTCTGCGCGGAGATTTTCCCCACCTCTGTCCGTGGCATCTGCATAGCCATCTG
CGCGCTAACCTTCTGGATCGGCGACATCATCGTGACATACACTCTCCCCGTGATGCTCAACGCCAT
CGGTCTCGCTGGAGTCTTCGGCATATATGCCATCGTTTGCATACTAGCCTTTGTATTCGTCTACAT
GAAGGTCCCTGAGACAAAGGGCATGCCCCTGGAGGTCATCACCGAGTTCTTCTCTGTCGGGGCAAA
GCAGGGCAAGGAAGCCACGGACTAGTTGCTCTGATCCGGTGATCCGCGTCGCTGGTGGTAATTTTG
TGGTGTCATAACTACTACTACACTGGTTAACCTGCGATGCTTTGGTGAAGAAACTTCAAAGAGAGC
AGATACGGAAGACTTTACATCGTGAGGCTGAATTGTGTTGTCGTAGGCCGGCTTTTGGAAGTAGGA
TATGTACTTAGATCATCTGTTCTTTTCGCTTTGGAACTTTCTATTTGTGTTATTCAGAATTTCTTG
CCCATGTAACTAGTGCTGTTATCACAATTTATGTCGATTATGTGTTTGCCTTAAAAAAAAAAAAAA
AAAAAAAAAAAAA
```

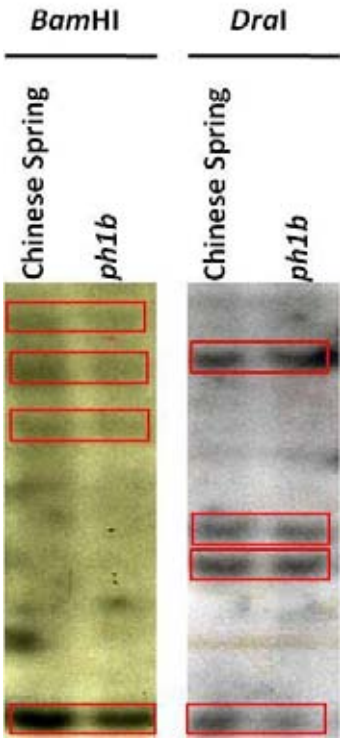
Figure 4 – *KK06* candidate full-length ORF sequence

```
ATGGCCATCGGATCTCTCCTCGCCTCGCGGCTGGCCAGGTCCGGCCACGCCCTCGCCACGCGCGCC
ATCGCTCAGGCTCCCCGGACCCACAATCATCATCACGCGACGTCTCCCCTGCTCTCCAGGCTCGGA
GCCGTGGCCCGCGCTTTCAGCTCAAGGCCCGCCGCCGCCGATGTCATCGGGATCGATTTGGGCACG
ACGAACTCATGTGTCTCCGTCATGGAAGGAAAGACACCACGGGTGATTGAGAATGCTGAAGGTGCG
AGGACAACGCCCTCCATTGTTGCCACAAACAGCAAAGGCGAGATCTTGATTGGCATCACTGCCAGT
CGACAGGCAGTGACAAATGCCGAGAACACAGTTCGTGGGTCCAAGCGCCTGATTGGTAGAGCTTTT
GATGACCCACAGACTCAGAAGGAAATGAAGATGGTGCCTTACAAGATTGTCAGGGGAACAAATGGT
GATGCCTGGGTGGAGATGGCTGGGAAGTCATACTCCCCGAGTCAGATTGGTGCGTTTGTTCTTACC
AAGATGAAGGAAACTGCAGAGGCTTACCTTGGCAAGTCGGTCTCCAAGGCTGTTATTACAGTCCCA
GCTTATTTCAATGATGCTCAGCGTCAGGCCACCAAGGATGCTGGTAGGATTGCTGGGTTGGACGTG
ATGAGGATTATCAATGAGCCCACTGCTGCAGCTCTGTCGTATGGAATGAACAACAAGGAGGGCCTG
ATTGCTGTATTTGACTTGGGTGGTGGCACATTTGATGTATCAATCCTTGAGATTTCTAATGGCGTT
TTTGAGGTCAAGGCGACCAATGGAGATACTTTCCTTGGTGGTGAGGACTTTGATGCTACATTGCTC
AACTACCTGGTTAGTGAATATAAGAACTCTGACAACATTGATCTGAGCAAGGACAAGTTGGCCCTG
CAAAGGCTCAGGGAAGCTGCTGAGAAGGCGAAGGTTGAGCTTTCCTCGACTCCACAGACTGAAATT
AATCTCCCATTCATCACAGCAGATGCCTCTGGTGCGAAGCATTTCAACATTACCTTGACCAGATCA
AAGTTTGAGTCTCTTGTGGGCAATCTCATTGAGAGAACTCGCATCCCATGCACAAACTGCCCCAAG
GATGCTGGTGTTTCTGCAAAGGAGATTGATGAGGTTCTACTGGTTGGTGGTATGACAAGGGTGCCA
AAGGTCCAGGATATTGTTTCTCAGATATTCGGCAAGTCTCCAAGCAAGGGTGTCAATCCTGATGAG
GCTGTTGCCATGGGGGCTGCCATTCAGGGTGGCATCTTGAGGGGTGATGTGAAGGAGCTCTTGTTG
CTGGATGTAACACCCCTTTCCCTTGGTATTGAGACTCTTGGAGGCATCTTCACAAGGCTAATCAAC
AGGAACACCACAATTCCAACCAAGAAGAGCCAGACCTTCTCCACTGCTGCAGATAACCAGACGCAG
GTTGGAATCAAGGTGCTTCAGGGTGAGAGGGAGATGGCCACAGATAACAAGCTTCTTGGTGAATTC
CAGCTTGAAGGTATTCCACCAGCCCCGAGGGGTATGCCACAGATCGAGGTCACTTTTGACATTGAT
GCCAATGGTATTGTCAAGGTGTCAGCAAAGGACAAGTCCACTGGCAAAGAGCAGGATATCACCATC
AAATCTTCAGGTGGTTTGTCTGACAGTGACATTGAGAAGATGGTGAAGGAGGCGGAGCTAAATTCT
CAGAGGGATCAGGAAAGGAAGTCGTTGATTGACCTCAGGAACTCTGCAGATACCACCATCTACAGC
ATCGAGAAGAGCGTCAGCGAGTACAAAGACAAGGTCCCTGCTGAAGTCGTCACGGAAATCCAGTCT
```

GCTGTCTCTGATCTGAGAGCAGCAATGGCTGGTGATGATTCGGACGCCATCAAGCAGAAGCTGGAG
GCGGCGAACAAGGCAGTCTCAAAGATTGGACAGCACATGCAGGGTGGTGGCGGCGCTGCCGGCGGC
GACAGCGGCAGCAGTGGTGGTGGCGACCAGACTCCGGAAGCTGAATACCAGGACCCCAAGGAGGCC
AAGATGTAG

Appendix B3

Figure 1 – Genotype southern blot for *KK06* candidate. Genotype Southern blots using *BamH*I and *Dra*I digested genomic DNA from Chinese Spring and the *ph1b* mutant revealed that the *KK06* gene candidate is located outside of the *ph1b* deletion. This result was based on equivalent banding patterns (denoted by red boxes) seen in both the Chinese Spring and *ph1b* lanes.



Appendix C1

Author declaration permitting the use of the publication, Khoo *et al.* (2011) Poor Homologous Synapsis 1 (PHS1) interacts with chromatin but does not co-localise with ASYnapsis 1 (ASY1) during early meiosis in bread wheat. *BMC Plant Biology* (submitted), in this thesis.

I the undersigned give permission for the paper;

Khoo KHP, Able AJ, Able JA (2011) Poor Homologous Synapsis 1 (PHS1) interacts with chromatin but does not co-localise with ASYnapsis 1 (ASY1) during early meiosis in bread wheat. *BMC Plant Biology* (submitted)

to be included into the thesis of Kelvin Khoo Han Ping for the academic program, Ph.D in Sciences at the University of Adelaide, Australia.

The statement of contributions below is sourced directly from the relevant publication.

Authors' contributions

KK performed all experimental procedures and drafted the manuscript.

KK, JA and AA participated in the design of the study.

JA and AA drafted the manuscript.

All authors read and approved the final manuscript.

Jason Able [redacted]

Amanda Able .. [redacted] .....

Appendix C2

Table 1 – List of primers used for isolation and characterisation of *TaPHS1* in Khoo *et al.* (2011) (Additional File 1).

Primer name	Primer sequence (5´ → 3´)	T_m (°C)
***TaPHS1* gene isolation primers**		
*TaPHS1*_F1	CATTTTCGGCGTCATCGTCGTCG	65
*TaPHS1*_R1	CTACAGTGACAAGTCGCCACCCAGTTCA	
***TaPHS1* gene expression primers (to amplify ORF for protein production)**		
*TaPHS1*_F2	ATGGCGGGCGCCGGC	62
*TaPHS1*_R1	CTACAGTGACAAGTCGCCACCCAGTTCA	
***TaPHS1* peptide primers (to amplify conserved regions for protein production)**		
*TaPHS1*_Reg1_F1	CGGCGGAGGCAGAGGTGG	65
*TaPHS1*_Reg1_R1	GAAGGAGTGGGTGGGGCGAGA	
*TaPHS1*_Reg2_F1	GTCTACGAGGAGCACTATGTATCTATCCTCAACTT	66
*TaPHS1*_Reg2_R1	AGGGAAACGTACAGCAAACTTCTGGAT	
*TaPHS1*_Reg3_F1	AAGGAACTCTCAAGCAACACCAT	59
*TaPHS1*_Reg3_R1	CGCTTCATCTGGCCTGTATTG	
*TaPHS1*_Reg4_F1	GGCGGGGACGACTCTTTTCAT	62
*TaPHS1*_Reg4_R1	Primer used was *TaPHS1*_R1	
Plasmid vector sequencing primers		
T7 (pGEM®-T Easy)	TAATACGACTCACTATAGGG	50
SP6 (pGEM®-T Easy)	ATTTAGGTGACACTATAG	
GW1 (pCR®8/GW/TOPO®)	GTTGCAACAAATTGATGAGCAATGC	50
GW2 (pCR®8/GW/TOPO®)	GTTGCAACAAATTGATGAGCAATTA	
Quantitative real-time PCR (Q-PCR) primers		
*TaPHS1*_QF1	CACTCGGATTGATGCTGCTG	55
*TaPHS1*_QR1	TGACAAGTCGCCACCCAGTT	
Southern blot probe primers		
*TaPHS1*_SF1	TGAACTGAATTGTGTCGGATGAA	54
*TaPHS1*_SR1	AAAGCTCACGAACACCACCCT	

Figure 1 – *TaPHS1* full-length ORF sequence (GenBank accession: GQ851928)

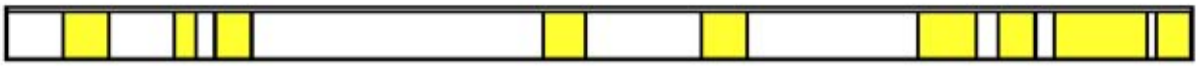
```
ATGGCGGGCGCCGGCGGCAGGAGCAGGGAGCGGCTCACGTCGCGCGCGGAGGAGGCCGCGGGGGGCAAGC
GGCGGAGGCAGAGGTGGGAGGTGGAGTTCGCGCGCTACTTCGCGAAGCCGCGGCGCGCCCCCTCGACGCC
GCCGCCCCCCGGCCTCCGCTACATATCCCGCGGCAAGCAACTCCACCAGGGCACCTGGCTCCTGGCAGCC
TCCCCCGCCGCCCTCTGCATCTCTCGCCCCACCCACTCCTTCGCCGCCCGCGTCCTCACCGTCTCCATCG
GCGACGTCGTCTACGAGGAGCACTATGTATCTATCCTCAACTTCTCATGGCCACAGGTTGCATGCGTGAC
AGAGTGCCCAGTTAGAGGGAGCAGGGTGGTGTTCGTGAGCTTTTGTGATAGGTCCAAACAGATCCAGAAG
TTTGCTGTACGTTTCCCTCGCCTCAGTGATGCAGAATCATTCTTGAATAGTGTGATTGTGAAGGAACTCT
CAAGCAACACCATGGATATCATGCCATCTGGAAGCGACTATATGTGTGAGCTTGAAGATTCATCATCATC
TGAATATATTCCTTCTAATGGACTTCAATACAGGCCAGATGAAGCGGTAAGCTTTGAGGAGCCAACTTCT
GATCACAGGACAGACGCACCTGCTGTTGGCTACCACATGGAACCAGATCAGCCTGTTCTTCAATCTCCCA
TTGCTACAAATATCAATAGCATCTACTCTGGTTTTCCCGAGGGTTACTCTGGCTTCCCTGAGGGCTACTC
TGGTTTCCCTGAGGGTTACTCTGGTTCAGTTAAAATTGAGAGAGATGGAGGACCCTTTCCAGCAACTATA
ACTGATCATGCTCCTGAAAAGGCCTACATACTGGACACTCGGATTGATGCTGCTGGTGGAAATTCGGTTG
CTGATAAAGGAAAGGGTGCTGGTAAAGAGATTGATGTTAGTGACGTAACACGTGATATTTTGGCAGGGAT
AGAGACATATGGCGGGGACGACTCTTTTCATGATATGCTGTCCAAGCTCGACAAAGCCATTGATGAACTG
GGTGGCGACTTGTCACTGTAG
```

Figure 2 – *TaRAD50* full-length ORF sequence

ATGAGCACGGTGGACAAGATGCTGATCAAGGGGATCCGGAGCTTCGACCCGGAGAACAAGAACGTCATCACCT
TCTTCAAGCCCCTCACCCTCATCGTCGGCTCCAACGGAGCCGGCAAAACCACCATCATCGAGTGCCTGAAGCT
GTCCTGCACCGGCGAGTTGCCCCCCCAACTCCCGCTCCGGCCACACCTTCGTCCATGACCCCAAGGTAGCAGGT
GAGACTGAAACGAAAGGGCAGATTAAGCTGCGGTTTAAGACAGCAGCAAGAAAGGATGTGGTGTGCATCCGCT
CTTTCCTGCTTACCCAGAAGGCATCAAAGATGGAGTTTAAGGCAATCGAGAGCGTCCTCCAGACTATAAATCC
ACACACGGGCGAGAAGGTTCGTCTCAGCTACAGATGCGCTGATATGGATAGGGAGATTCCAGCCCTAATGGGT
GTTTCGAAGGCTATATTGGAGAATGTTATATTTGTGCACCAAGACGAATCCAATTGGCCATTACAGGACCCTT
CAACACTTAAGAAAAAGTTTGACGACATCTTTTCTGCTACCCGCTATACCAAAGCTCTCGAAGTCATAAAGAA
ACTTCACAAGGACCAAGCACAAGAAATCAAGACTTTTAGGTTAAAGTTGGAGAACCTTCAGACTCTTAAAGAT
CAAGCATACAGGCTTCGTGACAGTATTGCACAAGATCAAGAGAAGTCAGACGCCTTAAAAACTCAAATGGAGG
ACCTGAAAACAAACATCCAAGCTGTGGAAAACAAAATCCTTCGTACCGAAACAAGTATGGTGGACTTGAGGAA
ACTTCAGGGGCAAATTAGCACCAAAGCAACTGCTAGAAGTACATATTTTACACTTCAGCAGCAACAGTATGCC
GCTCTTTCTGAGGAAAATGAAGATACTGATGAGGAGCTGAAGGAGTGGCAAACAAAATTTGAAGAAAAAATTG
CATTACTGGAAACGAAAATCGCTAAACTTGAAAGAGAGATGAATGATGAATATGCAAAAAGCTCTCTGCTATC
TGAGACCATCAATGATTCAACACGCGAAATAGGAAAGCTCCAGGCAAAAGCTGACGCTCACATGTCCGTGAAG
CATGAAAGAGATTCACCCATCAAAACGATATTCAATAAACATAATCTTGGGCCAGTTCCTGATGCTCCTTTTA
CCAATGATATTGCCATGAACCTTACAAACAGAACTAAAGCAAGACTATCGAATCTCGAGGACGATTTGCAGGA
AAAGAAAAAAACTAACGAGACACAGTTAGAATTTCTTTGGGGACGTTATCTTAAAGTAAATGCTCGCTACTCT
GAAGTTGATGGCCAGATACAGTCTAAGAAGGAATCTAAGATAGGTGTTTTAAGGCGCATAAAGGATAAAGAGA
ATGAGCGAGATGCTGCAAAGACGGAGCTTTCAAGGCATAATCTGGCCCGTATTGATGAAAGAGAGCGGCATCT
GCAAATTGAAGTTGAGAGGAAGACAATTGCGCTGGGAGAAAGAGACTATGATTTGATTATAAGTCAGAAGCGC
TCAGAGATATATACCTTGGATCACAAGATAAAAGCACTTCACCGAGAGAAAGATAACATAGCAACTGATGCTG
ATGACAGAGTAAAATTAGAACTCAAGAAGGATGAGTTGGAGAAGTGCAAGAAGAAACTTAAAAAGATATATGA
TGAACATAAGGATAAATTTAGAAGTGTCCTTAAGGGAAGGCTTCCTCATGAGAAGGATGTCAAGAAGGAGATT
ACTCAAGCTTTCGGGTCTGTAGACTCAGAATACAATGATTTGAACTCAAAATCTCAGGAAGCGGAACAACAGT
TGAAATTGGCACAAATGAAAATTGATGCTGCTAAAAGCCACTTATCAAAGCTCCAAAAAGTTTTGGATGCAAA
AAGAAAGCATCTGAACTCGAAACTTCAATCCATTTTTAAGGTATCTGTTGACATCAACGCTTACCCCAAGATT
TTGAAAGATGCCATGGACGAGAGAGATAAACAGACAAATAATTTCAGTTATGCTAAGGGAATGCGCCAAATGT
ATGAACCTTTTGAAACAGTGGCCCGCCAGCATCACAAGTGTCCTTGCTGTGATCGTGCTTTCACACCTGATGA
AGAGGACCTCTTTGTAAAGAAGCAAAGGACAACAGGTACAAGTACCGCAAAGCGCTTGAAAGTGCTGGCAGAG
AACTTATCAGTTGCTGAAGACTTATTCAATCAATTGGATAATCTCCGTGTGATTTATGATGAATATGTGAAAC
TGGAGAAAGAGACTATACCTTTAGCAGAGAAGGATTTGGAACAACTTTCGGCAGATAAAAGTGAGAAGGAACA
GATATCTGATGATCTTGTGAGTGTTCTTGCTCAAGTTAAAATGGACAGGGATGGAGTGGAAATCTTGTTACGT
CCAGTTGATACTATTGACAGGCATGTGCAAGAAATACAGGAGTTAGAACCACAAGTCAAAGATCTTGAATATA
AGCTTGATTCTCGTGGCCAAGGTGTTAAATCTGTCGACAAAATTCAGTTGGAGCTGATCTCTGTGCAGAGGGC
AAGGGACACATTGACTGGTGAAGTAGACGATCTCAGGGATCGGCAAAAAATGCTTAGCGAGGATCTATCAAAT
GCTCAGATGCGGTGGCATGCTCTTAGGGAAGAAAAACTAAGAGCTTCAAGTGTATTGCTGAAGTTCAAAAAGG
CTGGAGAAGATTTGGTACATTTTGCCGAGGAAAAGGAGCAACTGATCCTAGATCAGAAGCATTTAGAAGAAGC
TCTTGTTCCATTGTCAAAAGAGAGAGAAAGCTTGTTGCAAGAATATAAAGCTTTGAAGGAAAGGTTTGATCAG
GAGTATGATCAGCTGGCAGAAAGAAAAAGGGGATTCCAGCAAGAAATCGATGTACTTGGAACCCTTAACACAC
GCATCAAAGGGTACCTGGATTCAAACAAAGTAGAGAAGCTTAATGAGTTGCAGGAAAGGCATACCCTATCCCT
GTCTCAGTTACAGAAATGTGAGGCAAGAAAGCAAGACATCTCGGTTGAGCTTGACAAGAGCAAACGGCTATTA
CGGAGCCAAGATCAATTGAAAAGAAACATTGATGACAATCTGAACTACAGGAAAACAAAGGCTGAAGTGGATC
GACTTACTCATGATATCGAGTTACTTGAAGATAATGTGCTTTCCATCGGCAGCATGTCTACTATAGAAGCTGA
TCTTAAGCGGCATGCTCAAGAAAAGGAGAGGCTTCTTTCAGAGTATAATAGGTGTCAAGGAACAATTTCTGTT
TACCAAAGTAATATTTCAAAGCACAAACTTGAACTTAAACAGACACAATACAAGGATATTGAGAAGAGATATT
TTAACCAACTTCTTCAACTAAAGACAACTGAGATGGCAAACAAGGACTTGGACCGATACTATGCTGCTTTGGA
CAAGGCTCTTATGCGGTTCCATACCATGAAAATGGAGGAGATAAATAAAATAATTAAGGAATTGTGGCAGCAG
ACATACAGAGGCCAAGATATTGATTGCATAAGCATCAATTCTGATTCTGAGGGTGCAGGCACTCGATCATATA
GCTACCGTGTTGTCATGCAAAATGGTGGTGCTGAGCTGGAAATGCGAGGGCGATGCAGTGCTGGTCAGAAGGT
GCTTGCTTCTCTTATAATCAGACTCGCACTGGCAGAGACATTCTGTCTGAATTGTGGCATATTAGCTCTGGAT
GAGCCAACCACCAATCTTGATGGACCGAATGCTGAGAGCCTAGCTGCTGCCCTATTGAGAATAATGGAGAGTA
GGAAAGGCCAGGAGAACTTCCAGCTGATTATAATCACCCACGATGAGCGGTTTGCTCAGCTTATTGGTCAGAG
GCAACTTGCCGAGAAGTATTACCGAATCTCCAAGGATGAGCAGCAACACAGCAAAATAGAAGCCCAAGAGATA
TTTGACTAG

Figure 3 – Summary report of *Ta*PHS1 mass-peptide identification

Reference: Translation\of\pDEST17-TaPHS1\clone\sequence
Database: C:\Xcalibur\database\Whole Rice Proteome FASTA & TaPHS1.fasta
Number of Amino Acids: 382 Monoisotopic MW: 41918.6 pI: 5.60



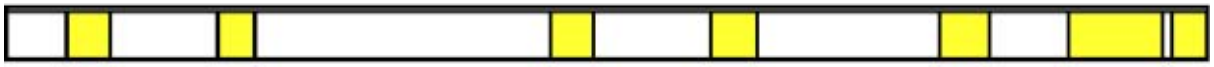
Protein:

MSYYHHHHHH LESTSLYKKA GSEFALMAGA GGRSRERLTS RAEEAAGGKR
RRQRWEVEFA RYFAKPRRAP STPPPPGLRY ISRGKQLHQG TWLLAASPAA
LCISRPTHSF AARVLTVSIG DVVYEEHYVS ILNFSWPQVA CVTECPVRGS
RVVFVSFCDR SKQIQKFAVR FPRLSDAESF LNSVIVKELS SNTMDIMPSG
SDYMCELEDS SSSEYIPSNG LQYRPDEAVS FEEPTSDHRT DAPAVGYHME
PDQPVLQSPI ATNINSIYSG FPEGYSGFPE GYSGFPEGYS GSVKIERDGG
PFPATITDHA PEKAYILDTR IDAAGGNSVA DKGKGAGKEI DVSDVTRDIL
AGIETYGGDD SFHDMLSKLD KAIDELGGDL SL

Protein Coverage:

Sequence	MH+	% Mass	AA	% AA
KAGSEFALMAGAGGR	1422.72	3.39	19 - 33	3.93
AGSEFALMAGAGGR	1294.62	3.09	20 - 33	3.66
QRWEVEFAR	1220.62	2.91	53 - 61	2.36
WEVEFAR	936.46	2.23	55 - 61	1.83
RAPSTPPPPGLR	1245.71	2.97	68 - 79	3.14
APSTPPPPGLR	1089.61	2.60	69 - 79	2.88
VVFVSFCDR	1071.53	2.56	152 - 160	2.36
LSDAESFLNSVIVK	1521.82	3.63	174 - 187	3.66
PDEAVSFEEPTSDHR	1715.75	4.09	225 - 239	3.93
IERDGGPFPATITDHAPEK	2051.02	4.89	295 - 313	4.97
DGGPFPATITDHAPEK	1652.79	3.94	298 - 313	4.19
AYILDTR	851.46	2.03	314 - 320	1.83
IDAAGGNSVADK	1117.55	2.67	321 - 332	3.14
GAGKEIDVSDVTR	1346.69	3.21	335 - 347	3.40
EIDVSDVTR	1033.52	2.47	339 - 347	2.36
DILAGIETYGGDDSFHDMLSK	2284.04	5.45	348 - 368	5.50
AIDELGGDLSL	1102.56	2.63	372 - 382	2.88
Totals:	14259.98	34.02	135	35.34

Reference: Translation\of\pDEST17-TaPHS1\clone\sequence
Database: C:\Xcalibur\database\Whole Rice Proteome FASTA & TaPHS1.fasta
Number of Amino Acids: 382 Monoisotopic MW: 41918.6 pI: 5.60



Protein:

MSYYHHHHHH LESTSLYKKA GSEFALMAGA GGRSRERLTS RAEEAAGGKR
RRQRWEVEPA RYFAKPRRAP STPPPPGLRY ISRGKQLHQG TWLLAASPAA
LCISRPTHSF AARVLTVSIG DVVYEEHYVS ILNFSWPQVA CVTECPVRGS
RVVFVSFCDR SKQIQKFAVR FPRLSDAESF LNSVIVKELS SNTMDIMPSG
SDYMCELEDS SSSEYIPSNG LQYRPDEAVS FEEPTSDHRT DAPAVGYHME
PDQPVLQSPI ATNINSIYSG FPEGYSGFPE GYSGFPEGYS GSVKIERDGG
PFPATITDHA PEKAYILDTR IDAAGGNSVA DKGKGAGKEI DVSDVTRDIL
AGIETYGGDD SFHDMLSKLD KAIDELGGDL SL

Protein Coverage:

Sequence	MH+	% Mass	AA	% AA
AGSEFALMAGAGGR	1294.62	3.09	20 - 33	3.66
WEVEPAR	936.46	2.23	55 - 61	1.83
RAPSTPPPPGLR	1245.71	2.97	68 - 79	3.14
APSTPPPPGLR	1089.61	2.60	69 - 79	2.88
VVFVSFCDR	1071.53	2.56	152 - 160	2.36
LSDAESFLNSVIVK	1521.82	3.63	174 - 187	3.66
PDEAVSFEEPTSDHR	1715.75	4.09	225 - 239	3.93
DGGPFPATITDHAPEK	1652.79	3.94	298 - 313	4.19
IDAAGGNSVADK	1117.55	2.67	321 - 332	3.14
EIDVSDVTR	1033.52	2.47	339 - 347	2.36
DILAGIETYGGDDSFHDMLSK	2284.04	5.45	348 - 368	5.50
AIDELGGDLSL	1102.56	2.63	372 - 382	2.88
Totals:	11717.68	27.95	112	29.32

Appendix D1

Figure 1 – Summary report of *Ta*ZYP1 mass-peptide identification

Reference: LOC_Os04g37960.1|12004.m08785|protein expressed protein
Database: C:\Xcalibur\database\Rice_database\Rice_database.fasta
Number of Amino Acids: 887 Monoisotopic MW: 101920.4 pI: 5.70

Protein:

MQKLGLSGLR GLEGFRSLAG STSTAAKAPN PKPSSDIGGS TYGSFANLKI
TAEKLVKEQA SVKTDLEMTV YFSAFTIIYG NFTGWVYGNI IPTVVIFHGA
NVLVISWIKE AKLQQAVNEN AKLKVKQTED SKLWQGLDSK VSSTKTLCNQ
LTETLQQLAS QTERAEEDKK FFEEILGKNS KAFEEFNCLL HDSSIKLECA
EQMIISGKQE MLRIKQEKEE MDQSYKEQLY ASDTTIREKN SLIKQLEDSI
EQNKARLLYV DSRLECMEQE LKLKEDVCIC LKENLASTES EKNDLKLRNE
GYTLEVQKLS KDNKELNELL SGFTVKVTEL DKEHTSISSH VTQLISSFER
YDGKVHEEKM LMIKSAKDKF EMLQMQYVNL ISENNALQTE IEELKSRIIE
LQKTQEIVMV QHVEECQVAE DKIRRLESEA EISASNISQL EKVASDLEGR
VQKLLEDSRS AENHKQELLQ KILKLESDNQ ELLGRVQSVL NEKSNDTESL
QGEIAKRDQQ VETLENQVNQ LCSIIDEKEQ LHTCAVEREK NLEEQKLQVQ
ASLAATESQL TEAKEQYDIM LEGKKIELSK HLKELSLKND QAINEIRRKY
ELEKVEIINI EKEKAEKLIK EMENKCNEKI SENRQDSESS FEMVARIQQD
NEKKESTLRA YHKEELQRIQ SQAENELRER LSSLRKDHEI QMKSLTKKHE
ENCQKLQDEL ELQKSKEEKQ RALLQLQWKV MGETQQVDQE VNSKKEYSVS
SIKRRDPYIR KEHQLQLVSP ETKRKDVNLS GIIQSPITNM LRKVEKGTQD
IPKHRKVTNH EYEVETANGR ITKRRKTKST VMPGEPNTQK SLMDTADKDP
TKMKKVVAGS HPHPANIGEL FSEGSLNPYA EDPYAFG

Protein Coverage:

Sequence	MH+	% Mass	AA	% AA
LGLSGLR	715.45	0.70	4 - 10	0.79
GLEGFRSLAGSTSTAAK	1652.86	1.62	11 - 27	1.92
LWQGLDSK	946.50	0.93	133 - 140	0.90
EQLYASDTTIR	1296.64	1.27	227 - 237	1.24
RLESEAEISASNISQLEK	2004.02	1.97	425 - 442	2.03
LESDNQELLGR	1273.64	1.25	475 - 485	1.24
VQSVLNEK	916.51	0.90	486 - 493	0.90
KQYDIMLEGK	1224.63	1.20	565 - 574	1.13
QYDIMLEGK	1096.53	1.08	566 - 574	1.01
AYHKEELQR	1173.60	1.15	660 - 668	1.01
LQDELELQK	1115.59	1.09	706 - 714	1.01
ALLQLQWK	999.60	0.98	722 - 729	0.90
STVMPGEPNTQK	1338.64	1.31	829 - 840	1.35
Totals:	6803.51	6.68	58	6.54

Reference: LOC_Os04g37960.1|12004.m08785|protein expressed protein
Database: C:\Xcalibur\database\Rice_database\Rice_database.fasta
Number of Amino Acids: 887 Monoisotopic MW: 101920.4 pI: 5.70

Protein:

MQKLGLSGLR GLEGFRSLAG STSTAAKAPN PKPSSDIGGS TYGSFANLKI
TAEKLVKEQA SVKTDLEMTV YFSAFTIIYG NFTGWVYGNI IPTVVIFHGA
NVLVISWIKE AKLQQAVNEN AKLKVKQTED SKLWQGLDSK VSSTKTLCNQ
LTETLQQLAS QTERAEEDKK FFEEILGKNS KAFEEFNCLL HDSSIKLECA
EQMIISGKQE MLRIKQEKEE MDQSYKEQLY ASDTTIREKN SLIKQLEDSI
EQNKARLLYV DSRLECMEQE LKLKEDVCIC LKENLASTES EKNDLKLRNE
GYTLEVQKLS KDNKELNELL SGFTVKVTEL DKEHTSISSH VTQLISSFER
YDGKVHEEKM LMIKSAKDKF EMLQMQYVNL ISENNALQTE IEELKSRIIE
LQKTQEIVMV QHVEECQVAE DKIRRLESEA EISASNISQL EKVASDLEGR
VQKLLEDSRS AENHKQELLQ KILKLESDNQ ELLGRVQSVL NEKSNDTESL
QGEIAKRDQQ VETLENQVNQ LCSIIDEKEQ LHTCAVEREK NLEEQKLQVQ
ASLAATESQL TEAKEQYDIM LEGKKIELSK HLKELSLKND QAINEIRRKY
ELEKVEIINI EKEKAEKLIK EMENKCNEKI SENRQDSESS FEMVARIQQD
NEKKESTLRA YHKEELQRIQ SQAENELRER LSSLRKDHEI QMKSLTKKHE
ENCQKLQDEL ELQKSKEEKQ RALLQLQWKV MGETQQVDQE VNSKKEYSVS
SIKRRDPYIR KEHQLQLVSP ETKRKDVNLS GIIQSPITNM LRKVEKGTQD
IPKHRKVTNH EYEVETANGR ITKRRKTKST VMPGEPNTQK SLMDTADKDP
TKMKKVVAGS HPHPANIGEL FSEGSLNPYA EDPYAFG

Protein Coverage:

Sequence	MH+	% Mass	AA	% AA
LGLSGLR	715.45	0.70	4 - 10	0.79
SLAGSTSTAAK	993.52	0.97	17 - 27	1.24
LWQGLDSK	946.50	0.93	133 - 140	0.90
EQLYASDTTIR	1296.64	1.27	227 - 237	1.24
LECMEQELK	1122.52	1.10	264 - 272	1.01
NDLKLR	758.45	0.74	293 - 298	0.68
LLEDSRSAENHKQELLQK	2138.12	2.10	454 - 471	2.03
QELLQK	758.44	0.74	466 - 471	0.68
LESDNQELLGR	1273.64	1.25	475 - 485	1.24
VQSVLNEK	916.51	0.90	486 - 493	0.90
EQLHTCAVER	1185.57	1.16	529 - 538	1.13
KQYDIMLEGK	1224.63	1.20	565 - 574	1.13
QYDIMLEGK	1096.53	1.08	566 - 574	1.01
KIELSK	717.45	0.70	575 - 580	0.68
LQDELELQK	1115.59	1.09	706 - 714	1.01
Totals:	4503.31	4.42	38	4.28

Figure 2 – Dual 3-dimensional immuno-fluorescence localisation of *Ta*ASY1 and *Ta*ZYP1 in *Taasy1-1.9.2* pachytene cells. Three representative replicates of *Taasy1-1.9.2* mutant pachytene meiocytes showing reduced levels of *Ta*ZYP1 signal compared to wild-type. Left most panel shows diffuse *Ta*ZYP1 signal with foci formation on the chromatin while centre and right panels show undetectable levels of *Ta*ZYP1.

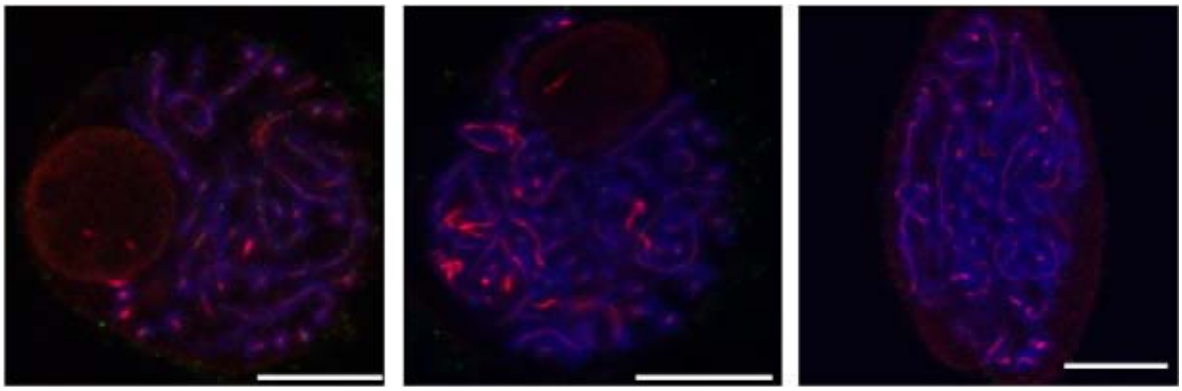
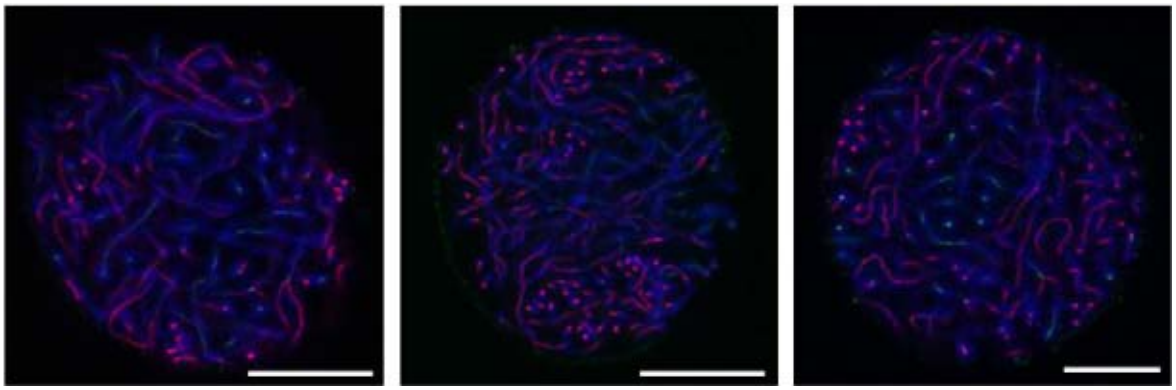


Figure 3 – Dual 3-dimensional immuno-fluorescence localisation of *Ta*ASY1 and *Ta*ZYP1 in *Taasy1-2.2.3* pachytene cells. Three representative replicates of *Taasy1-2.2.3* mutant zygotene-pachytene transition meiocytes showing no significant increase in *Ta*ASY1 and *Ta*ZYP1 protein signals compared to wild-type.



Appendix E

Copyright permission for Figure 1.1

Dear Kelvin,

RE: Permission to reuse one figure from McLeish and Snoad, Looking at chromosomes, 1958 in a University of Adelaide thesis

It is our policy to grant permission free of charge to students who are looking to quote excerpts from our titles for strictly academic purposes. This is subject to the following conditions:

- Your work must be of a entirely academic nature; no commercial publication of the work at hand is allowed;
- You must acknowledge our edition directly below the reproduced material as follows : «author/editor, title, year of publication, publisher (as it appears on our copyright page) reproduced with permission of Palgrave Macmillan. This material may not be copied or reproduced without permission from Palgrave Macmillan»

Your thesis may be reproduced through academic aggregators but should it become eligible for commercial publication at a later stage, you will need to re-apply for our permission.

Best wishes

Lauren Russell

Rights Co-ordinator
Palgrave Macmillan
Houndmills
Basingstoke
Hampshire
RG21 6XS
+44 (0)1256 329242 Ext. 3518
fax +44(0)1256 353774