MOLECULAR INVESTIGATIONS OF GENE FAMILIES ENCODING FOLDASE ENZYMES IN WHEAT

A thesis submitted for the degree of

DOCTOR OF PHILOSOPHY

by

JOSHUA CARL JOHNSON

BSc. (Hons.) (Applied Biology)

School of Molecular Sciences

Faculty of Health, Engineering and Science

Victoria University

Melbourne

AUSTRALIA

September 2005

DECLARATION

I, Joshua Johnson, declare that the PhD thesis entitled *Molecular Investigations of Gene Families Encoding Foldase Enzymes in Wheat* is no more than 100,000 words in length, exclusive of tables, figures, appendices, references and footnotes. This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work.

Joshua Johnson

September, 2005

ACKNOWLEDGEMENTS

My sincere thanks and gratitude goes to my principal supervisor, Dr. Mrinal Bhave, for her constant willingness to give of her time and mental energy for lengthy discussions during the planning and undertaking of this project and preparation of this thesis. Her great knowledge of molecular biology and ability to critically analyse the data generated during this study was an important factor in the development and publication of this thesis.

I am very grateful to Dr. Bryan Clarke, at Plant Industry, Commonwealth Scientific and Research Organization (CSIRO), Australia, for facilitating my research at these laboratories for a few weeks during my project and discussions on my project. Special thanks go to Tanya Phongkham and Lynette Rampling for their advice and technical assistance during these trips to Canberra.

I would like to extend many thanks Professor Rudi Appels at Murdoch University, Australia, for his invaluable assistance and advice on the genetic mapping components of my project. In addition, I would like to express my gratitude for his willingness to discuss and provide advice on other aspects of this project.

My thanks to Dr. Sarah Fraser for her warm help and willingness to give of her time for morale-boosting chats during my studies. In addition many thanks go to all of the PhD and Honours students in the lab, in particular, Belinda, Denijal, Kate, Meredith, Idris, Chutima, Sandra, Matthew and Goran, for their support and willingness to help when I needed it. Thanks to the technical staff at Victoria University, especially Dale, Charmaine, Michael, Marianne, Ira, Stacey, Joe and Min for their assistance and friendship throughout this project.

My thanks to the Grains Research and Development Corporation (GRDC), for providing me with financial support throughout this project through a Grain Industry Research Scholarship.

Thanks to my parents, brothers and sisters for their constant love and support. Special thanks must go to Norma for her generosity and support throughout my undergraduate and postgraduate studies.

Finally, thanks and love to my wife, Rebekah and son, Judah, who have been by my side and supported me through the ups and downs of my studies, which have been such a time consuming part of my life over recent years.

SUMMARY

The storage proteins of wheat (*Triticum aestivum* L.) form large polymers commonly referred to as gluten. These polymers are critical in determining important traits, especially in the processing qualities of dough. Vast amounts of research on the storage proteins and the genes encoding them have revealed the impact of different alleles on the properties of wheat dough. However, little is known about factors that may play a role in the folding, association and deposition of the storage proteins. One such factor could conceivably be the enzymes involved in catalysing folding of nascent proteins appropriately, the 'foldase' enzymes, including members of the protein disulfide isomerase (PDI) and cyclophilin families. Little is known about the genes encoding these enzymes in wheat, this information being limited to expression data showing the up-regulation of these genes in the developing endosperm, cDNA clones previously isolated in our lab and a report of partial *PDI* genes. One strategy to determine an association of the genes with important traits is to develop molecular markers and identify any genetic linkage between these gene-specific markers and quantitative trait loci related to dough quality. Thus, this project involved the characterisation of the gene families encoding PDI and cyclophilin in wheat and assessment of these genes for inter-cultivar polymorphism that could be used to develop 'perfect' molecular markers for these genes. Further, orthology between the *PDI* loci in wheat and the *esp2* locus in rice, which has been associated with irregular storage protein deposition, is described.

Cloning and sequencing of the *PDI* genes of *T. turgidum* subsp. *durum* (*TtPDI4A* and *TtPDI4B*) and that of *Ae. tauschii* (*AetPDI4D*) revealed a conserved 10 exon / 9 intron structure and high sequence conservation (>95%). These sequences facilitated the isolation and sequencing of sections of the corresponding genes from *T. aestivum* cv. Katepwa by allele-specific PCR, revealing almost complete conservation (97-100%) between the genes in the hexaploid and its tetraploid and diploid progenitors. This data suggested the presence of three, two and one *PDI* genes in these species, respectively, one per genome. Investigations into an additional *PDI* gene reported to be located on chromosome 1B suggested that, if present, it is a partial gene and is unlikely to be expressed.

The characterisation of the *PDI* gene family permitted the analysis of these genes in eight cultivars used in the development of genetic maps. Partial sequencing of the three *PDI* genes identified a single nucleotide polymorphism (SNP) in each of the genes on chromosomes 4AL and 4BS, however, no polymorphism was identified in the *PDI* gene on chromosome 4D of the eight cultivars after sequencing over 80% of this gene. The SNP identified in the *PDI* gene on chromosome 4A was found to be polymorphic between three crosses, while the SNP in the *PDI* gene on 4BS was found to be polymorphic in one. These SNPs were used to develop molecular markers for both the 4AL and 4BS genes, exploiting restriction fragment length polymorphisms (RFLPs) for *Sml* and *Bsu*36I, respectively, to identify each allele through the generation of characteristic-sized restriction fragments due to the presence of the SNPs. The molecular markers were subsequently used to screen the doubled haploid progeny of each of the crosses containing the polymorphic SNPs revealing co-segregation of the PDI4A marker (*XvutPDI*) with a number of amplified fragment length polymorphism (AFLP) and simple sequence repeat (SSR) markers localised to chromosome 4A as well as the *GERMIN* marker, linked to an oxalate oxidase gene (*germin*) involved in a number of important traits including, germination, stress response and pathogen defence. Similarly, the PDI4B marker (*XvutPDIb*) was found to be linked to a number of anonymous, random SSR, RFLP and AFLP markers localised to chromosome 4B and the *XGERMINB* marker, linked to a homeologue of the *germin* gene described above. A literature search revealed several quantitative trait loci (QTLs) mapped to chromosomes 4A and 4B.

Reports on the rice mutant, esp2, revealed the presence of aberrant storage protein deposition in conjunction with the absence of PDI expression, suggesting a direct role for PDI in storage protein deposition and thus presence of a hypothetical QTL for such a trait at the *esp2* locus of rice. As orthology has been established between the genomes of many of the cereal crops, it was possible that this hypothetical QTL may be present in wheat also. The lack of PDI expression in esp2 suggested that the mutation could be directly in the *PDI* gene or its flanking regulatory sequences. To address these questions, a comparison of the *esp2* locus and the *PDI* gene of rice revealed they were present at similar positions in the rice genome on the short arm of chromosome 11 distal to the *S20163S* RFLP marker. In order to determine whether flanking genes present at the *PDI* locus of rice were also present at similar loci in wheat, an analysis of a 1Mb

region of the TIGR rice genome assembly 2004 on chromosome 11 containing the *PDI* gene to identify putative expressed wheat orthologues was undertaken. This work revealed 34 tentative consensus (TC) sequences with high sequence similarity to sequences at the *PDI* locus of rice. Subsequent analysis revealed that 5 of these TCs contained sequence data from clones that had been used as probes in physical mapping experiments as part of the US Wheat EST project. Two of these probes were specific for the *PDI* gene of wheat and the three others for genes putatively orthologous to those flanking the rice *PDI* gene. Analysis of this data revealed that all probes identified loci on the group 4 chromosomes of wheat; however, the reported physical loci for the two *PDI* probes were contradictory. To clarify this contradiction the *PDI* genes of wheat were localised to chromosomal 'bins' using AS-PCR, confirming their presence on, and refining their locations to, bins on the centromeric section of chromosome 4AL and distal sections of chromosomes 4BS and 4DS. Thus, the present data provides strong evidence of an orthology between the *PDI* loci of rice and wheat and therefore the presence of a QTL involved in storage protein deposition at the *PDI* loci of wheat.

Characterisation of three isoforms of genes encoding cyclophilin A from eight wheat cultivars revealed that they comprise a small, intronless multi-gene family, lack any inter-cultivar polymorphisms and localise to chromosomal arms 6AS, 6BS and 6DS, in a region where genes for other quality traits are localised, the locus at 6AS possibly having duplicated genes. Further, cDNAs encoding two novel, endosperm-expressed classes of cyclophilins were isolated in the course of this work, one putatively encoding a cyclophilin with a signal peptide targeting it to the plastid and the other, a nuclear protein with cyclophilin-like domains that appears to be related to SR-cyclophilins which may be involved in RNA processing. *In-silico* analyses have further led to identification of another form of cyclophilin A and a potentially ER-localised, endosperm-expressed cyclophilin B. The plastid and ER-localised forms are of particular relevance to events occurring during endosperm maturation. The results thus provide valuable data and molecular tools for isolation and analysis of these genes, to address their roles and any association with wheat quality traits.

PUBLICATIONS

Refereed Journal Articles

Johnson, J.C., Appels, R. and Bhave, M. (2005). The *PDI* genes of wheat and their syntenic relationship to the *esp2* locus of rice. Funct Integr Genomics. Available online 8/2005.

Johnson, J.C. and Bhave, M. (2004). Characterisation and physical mapping of cyclophilin A genes and identification of new classes of cyclophilins in wheat. J Cereal Sci. 40: 137-150.

Johnson, J.C. and Bhave, M. (2004). Molecular characterisation of the protein disulphide isomerase genes of wheat. Plant Sci. 167: 397-410.

Johnson, J.C., Clarke, B.C. and Bhave, M. (2001). Isolation and characterisation of cDNAs encoding protein disulphide isomerases and cyclophilins in wheat, J Cereal Sci. 34: 159-171.

Book Chapters

Bhave, M. and Johnson, J. (2005). The 'foldase' enzymes PDI and cyclophilins: roles in storage protein folding/deposition and wheat grain quality. In: *Biotechnology for Food, Agriculture and the Environment*, Thangadurai, D. (Ed.), Regency Publications, New Delhi. (In Press).

Conference Proceedings

Johnson, J.C., Appels, R. and Bhave, M. (2005). Genetic mapping of the *PDI* genes of wheat and synteny between these and the *esp2* locus of rice. Presented at: The 12th RACI Convention, 3-7 July, 2005, Sydney, Australia. (In Press).

Johnson, J.C. and Bhave, M. (2002). Characterisation of the PDI genes of wheat, In: *Cereals 2002, Proceedings of the 52nd Annual Cereal Chemistry Conference*, Black, C.

K., Panozzo, J. F., Wrigley, C. W., Batey, I.L. and Larsen, N. (eds.), 9-12 Sept, 2002, Christchurch, New Zealand. pp. 206-209.

Johnson, J.C., Clarke, B. and Bhave, M. (2000). Isolation and characterisation of cDNAs encoding protein disulfide isomerases and cyclophilins in wheat, In: *Cereals 2000, Proceedings of the 11th ICC Cereal and Bread Congress and of the 50th Australian Cereal Chemistry Conference*, Wooton, M., Batey, I. L. and Wrigley, C. W. (eds.), 11-14 Sept, 2000, Surfers Paradise, Queensland, Australia. pp. 382-385.

Conference Poster Presentations

Johnson, J.C. and Bhave, M. (2003). The mapping of genes encoding foldase enzymes in wheat. Presented at the XIX International Genetics Congress, Melbourne, Victoria, Australia, 2003.

Johnson, J.C. and Bhave, M. (2002). Characterisation of the *PDI* genes of wheat. Presented at The 52nd Annual Cereal Chemistry Conference, Christchurch, New Zealand, 2002.

Johnson, J.C., Clarke, B and Bhave, M. (2001). The diversity of protein disulfide isomerase genes expressed in the developing wheat endosperm. Presented at the 22nd Lorne Genome Conference, Lorne, Victoria, Australia, 2001.

ABBREVIATIONS

AFLP	amplified fragment length polymorphisms
ANGIS	Australian National Genomic Information Service
AS-PCR	allele-specific PCR
BAC	bacterial artificial chromosome
BiP	binding protein
bp	base pairs
C-terminal	carboxyl terminal
CD	*Triticum aestivum* cv. CD87
cDNA	complementary DNA
Cr	*Triticum aestivum* cv. Cranbrook
CS	*Triticum aestivum* cv. Chinese Spring
CsA	cyclosporin A
CSIRO	Commonwealth Scientific and Industrial Research Organisation
Cyp	cyclophilin
dATP	adenosine deoxyribonucleoside triphosphate
dCTP	cytosine deoxyribonucleoside triphosphate
dGTP	guanosine deoxyribonucleoside triphosphate
DH	doubled haploid
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	equimolar mixture of dATP, dCTP, dGTP and dTTP
DPA	days post-anthesis
DTT	dithiothreitol
dTTP	thymine deoxyribonucleoside triphosphate
EDTA	ethylenediaminotetra-acetic acid
Eg	*Triticum aestivum* cv. Egret
ER	endoplasmic reticulum
EST	expressed sequence tag
FAO	Food and Agricultural Organization of the United Nations
FKBP	FK506 binding protein
g	times gravitational force

gDNA	genomic DNA
GRDC	Grains Research and Development Corporation
H-bonding	hydrogen bonding
Hb	*Triticum aestivum* cv. Halberd
HMW	high molecular weight
Indel	insertion/deletion
IPTG	isopropylthio-β-D-galactoside
Ka	*Triticum aestivum* cv. Katepwa
kb	kilobase pairs
kDa	kilodaltons
LB	Luria-Bertani broth
LMW	low molecular weight
MAS	marker-assisted selection
mRNA	messenger RNA
N-terminal	amino terminal
nm	nanometre
NWMMP	National Wheat Molecular Marker Program
ORF	open reading frame
PCR	Polymerase Chain Reaction
PDI	protein disulfide isomerase
pI	isoelectric point
PPIase	peptidyl-prolyl *cis-trans* isomerase
QTL	quantitative trait loci
RAPD	randomly amplified polymorphic DNA
RFLP	restriction fragment length polymorphism
RIL	recombinant inbred lines
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
Sc	*Triticum aestivum* cv. Sunco
SDS	sodium dodecyl sulfate
SDS-PAGE	SDS polyacrylamide gel electrophoresis
SNP	single nucleotide triphosphate

Ss	*Triticum aestivum* cv. Sunstar
SSR	simple sequence repeats
Ta	*Triticum aestivum* cv. Tasman
TaGI	*Triticum aestivum* gene index
TE	Tris-EDTA buffer
TIGR	The Institute for Genomic Research (Rockville, MD, USA)
U	units
UTR	untranslated region
UV	ultraviolet
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

GENERAL INTRODUCTION AND LITERATURE REVIEW

List of Figures and Tables in Chapter 1

Chapter 1 General Introduction and Literature Review

ABSTRACT

The storage proteins of wheat form large polymers commonly referred to as gluten. These are critical in determining important traits, especially in the processing qualities of dough. Specifically, the polymer-forming ability of the storage proteins, and thus the size of storage protein bodies in the wheat endosperm, are positively associated with the elasticity and extensibility of the wheat dough, with highly visco-elastic doughs generally preferred for the production of leavened products. Characterisation of the storage proteins and the genes encoding them has revealed the importance of different alleles in conferring beneficial (or undesirable) qualities to wheat dough. However, little is known about factors that may play a role in the folding, association and deposition of the storage proteins. One such factor could conceivably be enzymes involved in the ensuring the appropriate folding and association of proteins, the 'foldase' enzymes, including members of the protein disulfide isomerase (PDI) and cyclophilin families. PDI is responsible for the reduction, oxidation and isomerization of disulfide bonds in nascent proteins in the endoplasmic reticulum and cyclophilins are members of the peptidyl-prolyl *cis-trans* isomerase (PPIase) family, which catalyze the *cis-trans* isomerization of peptide bonds preceding prolyl residues. Support for the involvement of these enzymes in protein body formation is limited to the apparently non-random, rapid association and deposition of the storage proteins and the up-regulation of these enzymes prior to and during protein body formation in the wheat endosperm. However, little is known about the genes encoding these enzymes, limited to cDNA clones previously isolated in our lab and a report of partial *PDI* genes. Thus, this project aimed to characterise the gene families encoding PDI and cyclophilin in wheat and assess these genes for inter-cultivar polymorphism that could be used to develop 'perfect' molecular markers for these genes. Further, orthology between the *PDI* loci in wheat and a locus in rice associated with irregular storage protein deposition was investigated. This chapter will describe the necessary background information on the storage proteins, their deposition into protein bodies and the biochemical activities of PDI and cyclophilins, the genes encoding these enzymes and their role in storage protein deposition. Genetic analyses in wheat and our current knowledge of orthology between crop species are also reviewed.

1.1 WHEAT PRODUCTION AND WHEAT QUALITY

1.1.1 The importance of wheat research

Common wheat (*Triticum aestivum* L.) is arguably the most important food crop grown in the world in being cultivated in more areas throughout the world than any other crop and being the primary source of dietary protein for the world's population and its livestock. As such, the global demand for wheat is high, with current demand outstripping production. Global utilization of wheat was calculated to be 614.8 million tonnes in 2002/2003 compared to production of 569.6 million tonnes of wheat during the same period (Food and Agricultural Organization of the United Nations (FAO; www.fao.org)). Although demand is predicted to fall by 1.3% during 2003/2004, production is predicted to fall further, by 1.9% over the same period. In Australia, wheat production is quite low by world standards, representing about 3% of the world's total. However, the small population of Australia means 80% of all Australian wheat is exported, representing 8-15% of the wheat traded on the global market, worth $3,412 million in 2000-2001, or 4% of Australia's total exports (Australian Wheat Board; www.awb.com.au). The demand for wheat poses a significant challenge to increase the yield of wheat crops to maintain adequate supplies. In addition, the quality of this wheat must be of an appropriate standard to meet the needs of the end-user. To meet this challenge, research must be undertaken to increase the yield of wheat, whilst maintaining the quality characteristics required for end-uses.

1.1.2 Genetic and environmental determinants of overall grain quality

The wheat grain consists primarily of the starchy endosperm, making up >80% of the dry weight of the grain, with the remainder of the grain consists of the embryo and seed coat or bran layers that make up the pericarp/testa (Figure 1.1). The endosperm is separated from the other grain components during the milling process to generate white flour and it is the properties of this flour that determine the potential end-uses of wheat. There are a number of factors that can influence these potential end-uses, including (i) the composition and quality of starch; (ii) the quantity and quality of protein; (iii) grain hardness; (iv) lipids and lipoproteins; (v) non-starch carbohydrates; and (vi) the colour of the endosperm and pericarp (Morris, 1998). Most of these quality traits are

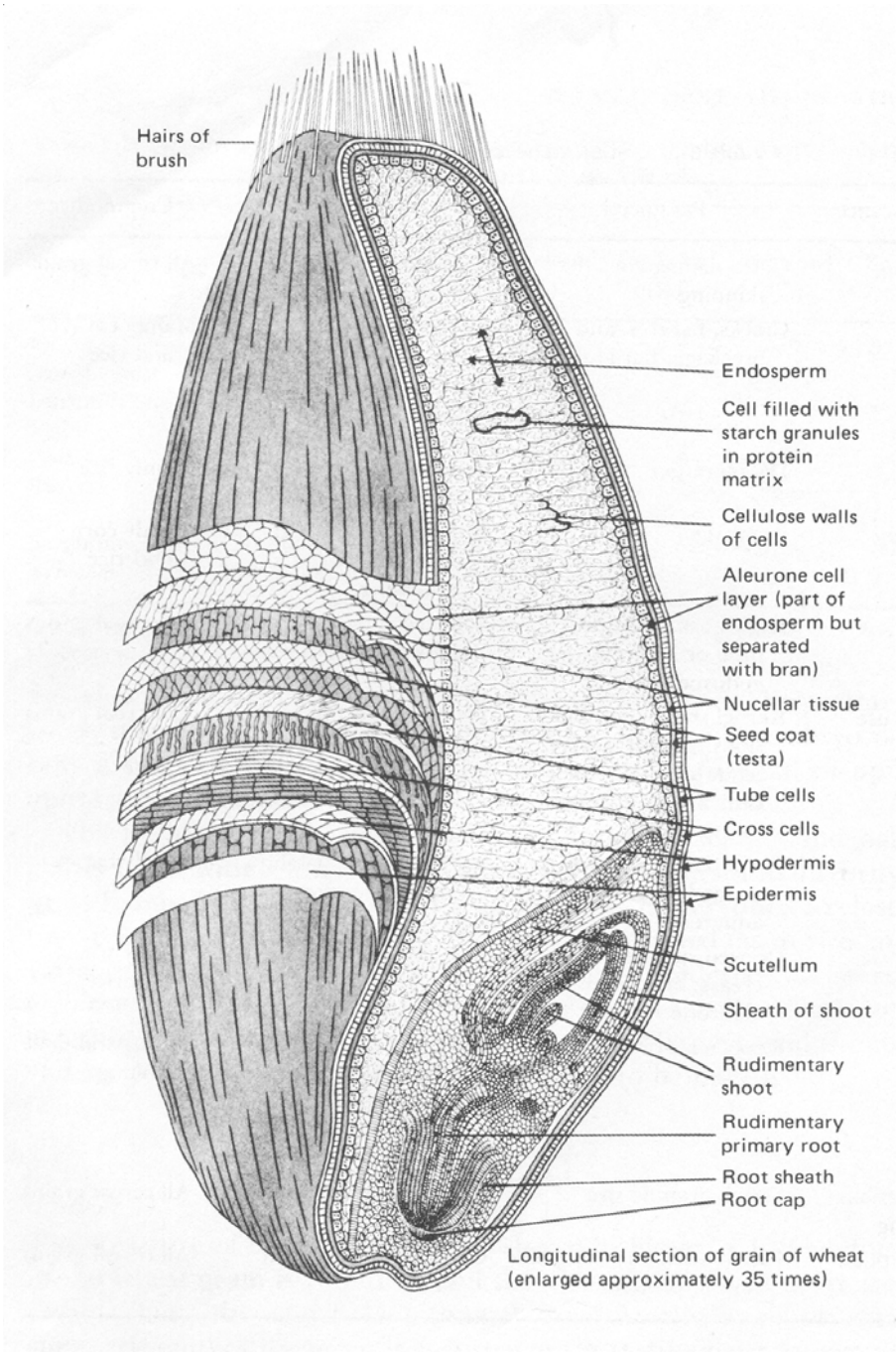


Figure 1.1 Basic structure of the wheat grain.

A longitudinal section of a wheat grain highlighting the main tissue types (Pomeranz, 1987).

determined genetically, by the expression of specific genes during the maturation of the endosperm, making this particular tissue the focus of much research. In addition, these traits can be affected by environmental conditions, primarily temperature, drought, abiotic and biotic stressors and availability of nutrients.

About 50% of the protein content of the wheat endosperm is in the form of storage proteins. The physiological role of the storage proteins is to act as a reservoir of amino acids for the germinating plant; however, researchers have focussed on the storage proteins due to their nutritional value and even more so, their role in determining the visco-elastic properties of wheat dough. The latter trait is important in determining end use of wheat as those cultivars that produce highly elastic (strong) doughs are generally used for pan breads, while those that producer weaker doughs are used for the production of noodles, biscuits, cakes and flat breads (Shewry *et al.*, 2003). The role of the storage proteins in determining the visco-elastic properties of dough is discussed below (Section 1.4.2).

1.2 WHEAT TAXONOMY AND THE GENOMES OF WHEAT

Taxonomy in wheat is an area of some uncertainties at both the generic and species levels, with 13 different classification systems currently in use (Wheat Genetics Resource Center at Kansas State University website, available at http://www.k-state.edu/wgrc/Taxonomy/taxintro.html). Wheat and its wild relatives (*Triticum* spp. & *Aegilops* spp.) are members of the tribe *Triticeae*, of the Pooideae subfamily of the grass family Poaceae or Gramineae (Germplasm Resources Information Network (GRIN); http://www.ars-grin.gov/). The *Triticeae* tribe is further subdivided into *Triticinae* (consisting of many genera including *Triticum*, *Aegilops* and *Secale*), and *Hordeinae* (consisting of *Hordeum* and related genera). In this project, the bread wheat lineage followed is as described by Van Slageren (1994) i.e., the hexaploid wheat as *T. aestivum* L., the tetraploid wheat as *T. turgidum* and the diploids as *T. urartu*, *T. monococcum* subsp. *monococcum*, *Ae. speltoides* and *Ae. tauschii*.

The genera *Triticum* and *Aegilops* contain 13 diploid and 18 polyploid species, most of which are allopolyploids with genomes consisting of genetic material proposed to have

originated from two (or more) evolutionary distinct diploid species (as opposed to autopolyploids, wherein polyploidy is achieved through duplication of a single diploid genome). The diploid species of *Triticum* and *Aegilops* species contain one of eight distinct genomes, designated as A (A and A^b/A^m), D, S (S, S^b, S^s, S^{sh}, S^l), M, C, U, N and T and the polyploids contain an additional two, B and G, as the diploid sources of these genomes remain unclear. The genomes of the polyploids are designated by their proposed diploid sources, for example, AABB for the tetraploid *T. turgidum* spp. durum (durum wheat) and AABBDD for the hexaploid *T. aestivum* (common wheat).

1.3 EVOLUTION OF THE *T. aestivum* GENOME

1.3.1 The putative origins of the A, B and D genomes of *T. aestivum*

Common wheat, *T. aestivum* (AABBDD, 2n=6x=42) is a hexaploid, or more accurately, an amphidiploid, a polyploid containing three evolutionary distinct diploid genomes (A, B and D). Research into the evolution of the wheat genome supports the origin of *T. aestivum* approximately 8000 years ago through spontaneous hybridisation of one of the (cultivated or wild) subspecies of the tetraploid *T. turgidum* with the wild diploid goat grass *Ae. tauschii* (Sax, 1922; McFadden and Sears, 1946; Lubbers *et al.*, 1991; Dvorak *et al.*, 1998) (Figure 1.2).

1.3.2 Current theories on the donors of the A and B genomes to *T. turgidum*

Although there is strong evidence supporting *T. turgidum* as the tetraploid donor of the A and B genome to *T. aestivum*, the diploid donors of the A and B genomes have been, and remain, the subject of much controversy. Much of the early cytogenetic evidence and later, C-banding analysis, isoenzyme and amino-acid sequence analysis supported *T. monococcum* as the diploid donor of the A genome to polyploid wheat (Sax, 1922; Lilienfeld and Kihara, 1934; Morris and Sears, 1967; Gill and Kimber, 1974; Jaaska, 1980; Jones and Mak, 1983). The identification of *T. urartu* as also carrying an A genome, raised the possibility that this species could be the donor of the A genome, which has since been supported by investigations into repetitive DNA (Dvorak *et al.*, 1988; Dvorak *et al.*, 1993) and comparison of gene sequences (Huang *et al.*, 2002) and isoenzymes (Jaaska, 1997) between diploid and tetraploid wheat.

Investigations into the diploid donor of the B genome to polyploid wheat has been more controversial. It is generally accepted that members of the *Sitopsis* section of *Aegilops* (*Ae. speltoides*, *Ae. sharonensis*, *Ae. bicornis*, *Ae. longissima* and *Ae. searsii*) are the most likely donors of the B genome, however, many alternative theories exist. The 'polyphyletic origin' school of thought proposes that the B genome has evolved from >1 of these diploid species as suggested by work with isoenzymes by Vittozzi and Silano (1976), studies of the *Pur-B1* locus of *T. turgidum* and *T. aestivum* which encode purothionins with amino acid sequences very similar to those analysed in three of the five putative donors of the B genome from the *Sitopsis* section (Kerby *et al.*, 1990) and RFLP data, demonstrating that >1 member of the *Sitopsis* section has the same patterns as *T. aestivum* (Takumi *et al.*, 1993). There is also extensive evidence supporting the alternative school of thought, of a 'monophyletic origin' i.e., only one species of the *Sitopsis* section donated the B genome to polyploid wheat. Many studies identified *Ae. speltoides* as the likely donor, e.g., comparisons of the rRNA spacer at the major *NOR* locus of many diploid wheats to that in *T. aestivum* found only *Ae. speltoides* shares close homology (Gill and Appels, 1988), analysis of two repetitive DNA sequences and a low-copy number DNA sequence revealed that these sequences from the B genome of *T. aestivum* were most closely homologous to those of *Ae. speltoides* than the other members of the *Sitopsis* section (Talbert *et al.*, 1991; 1995), an *Ae. speltoides*-derived repetitive sequence that was not present in other members of the *Sitopsis* section but found in the B genome of *T. turgidum* and *T. aestivum* (Daud and Gustafson, 1996), C-banding analysis revealing that *Ae. speltoides* is more similar to the B genome of polyploid wheat than to other members of the *Sitopsis* section (Badaeva *et al.*, 1996) and single-copy DNA sequences at fourteen loci in various wheat species (Blake *et al.*, 1998). However, other researchers believe that the single donor is not *Ae. speltoides*, as argued by Miyashita *et al*. (1994), who base it on RFLP differences between *Ae. speltoides* and *T. turgidum* / *T. aestivum* and Mori *et al.* (1997), who base it on variation in the *coxII* intron. An alternative, unidentified, member of the *Sitopsis* section has been suggested to be the donor, based on glutenin subunits in wheat (Randhawa *et al.*, 1997) and it has also been suggested that no currently existing species of the *Sitopsis* section is the donor of the B genome, although they probably share a common ancestor (Blake *et al.*, 1999; Huang *et al.*, 2002).

Taking into consideration the above evidence, the origin of the B genome remains equivocal. While significant amounts of evidence support *Ae. speltoides* as being the donor of the B genome to polyploid wheat, the various alternative lines of argument (discussed above) suggest that it is possible that the true donor has yet to be identified, or is now extinct.

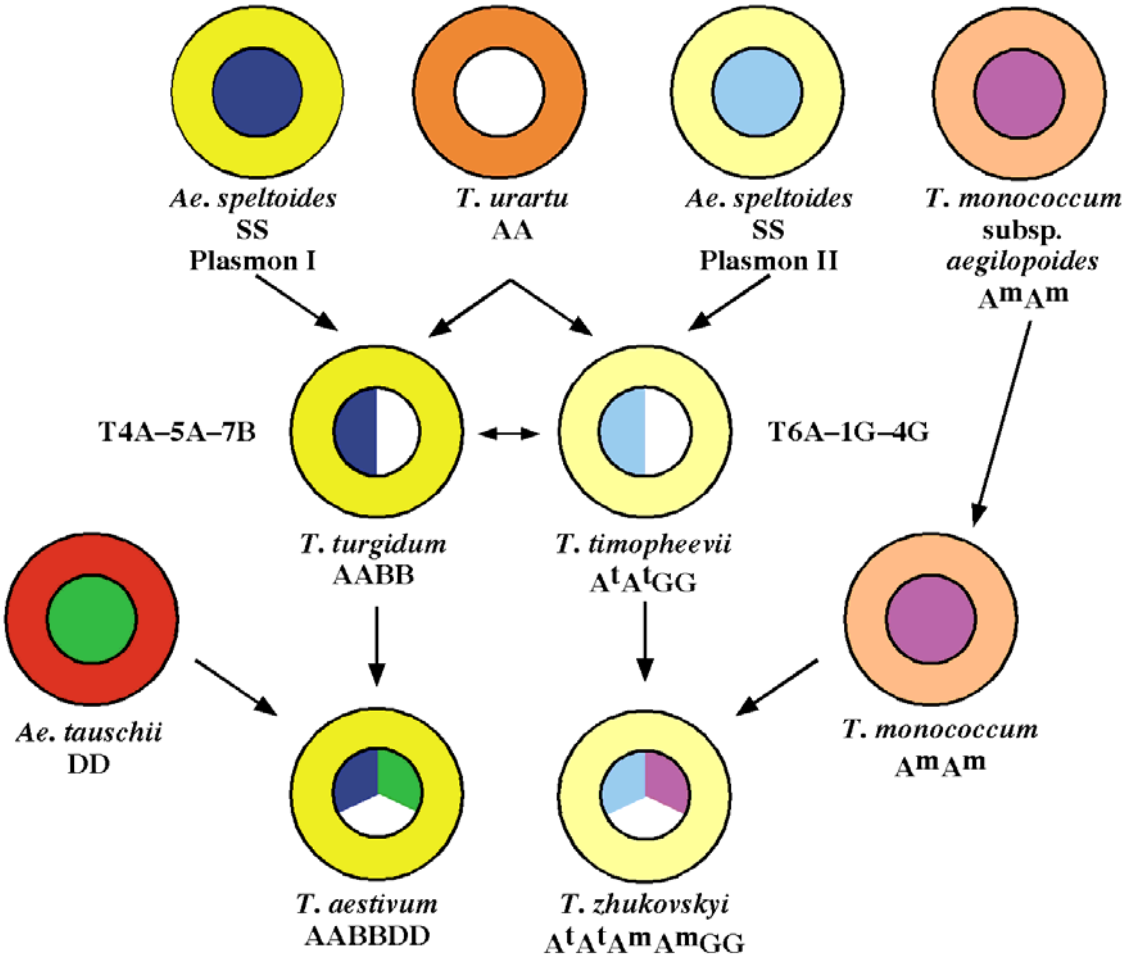


Figure 1.2 Current theory of the evolution of wheat.

The proposed diploid and tetraploid progenitors of the A (*T. urartu* (AA) and *T. turgidum* (AABB)), B (*Ae. speltoides* (SS) and *T. turgidum* (AABB)) and D (*Ae. tauschii* (DD)) genomes of the hexaploid *T. aestivum*. The depiction of *Ae. speltoides* (S genome) as the donor of the B genome remains controversial (Section 1.3.2) (Figure from the Wheat Genetics Resource Center - http://www.ksu.edu/wgrc/Extras/evolve.html).

1.4 THE STORAGE PROTEINS OF COMMON WHEAT

1.4.1 Classification of the storage proteins

As mentioned above, the protein content of (6.5%-18%) of the wheat grain consists of approximately 50% storage proteins. When white flour is mixed with water to form dough, these storage proteins form a continuous, elastic network in the mature grain that collectively is referred to as 'gluten'. Early investigations into the protein composition of gluten revealed that the major storage protein fraction consisted primarily of the alcohol/water mixture-soluble prolamins (Osborne, 1924), so-named due to their high content of the amino acids proline and glutamine. Subsequently the definition was widened to include proteins insoluble in alcohol/water because they were bound together in disulfide-bond stabilised polymers. The early classification system divided the prolamins into two main fractions that are still commonly used, i.e., the monomeric gliadins which are soluble in alcohol/water solutions and the polymeric glutenins which are insoluble in these. Subsequent analysis of the amino acid sequences of the glutenins and gliadins led to their reclassification into three groups: the sulfur-poor prolamins (S-poor), the sulfur-rich prolamins (S-rich), and the high molecular weight (HMW, occasionally referred to as HMM) prolamins (Miflin *et al.*, 1983) (Figure 1.3). To understand how each of these fractions affects the functionality of wheat dough, their structure and their ability (or inability) to form polymers must be described.

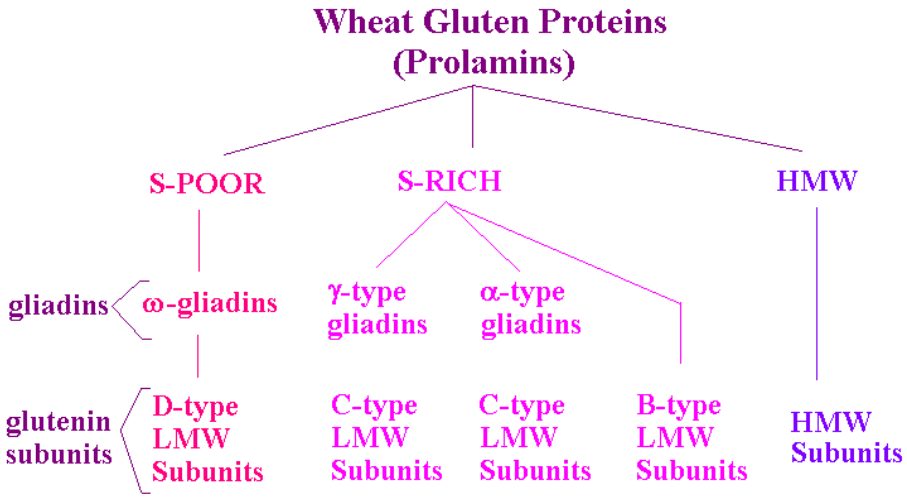


Figure 1.3 Classifications of wheat prolamins

The classes and nomenclature of wheat prolamins (Shewry *et al.*, 2003).

1.4.1.1 S-poor prolamins

The S-poor prolamins of wheat are called so due their general lack of cysteine residues and hence the inability to form disulfide bonds and thus to form polymeric structures in gluten. The basic structure of the S-poor prolamins, typified by the C-hordein in barley and the ω-secalins in rye, is almost entirely of a repetitive octapeptide motif (Pro-Gln-Gln-Pro-Phe-Pro-Gln-Gln) flanked by short, unique sequences of 12 residues at the C-terminal and 6 at the N terminal (4 in ω-secalins) (Shewry *et al.*, 1995) (Figure 1.4). In wheat, the largest group of S-poor prolamins are the ω-gliadins, found as monomers in the wheat endosperm due their absence of polymer-forming cysteine residues. However, Masci *et al.* (1991a; 1991b; 1993; 1999) have reported a polymer-forming class of S-poor prolamins, the D-type glutenin subunits, which appears to be a mutant form of ω-gliadin with a Ser-Cys mutation in the central repetitive domain. This suggestion is based on the identification of at least three genes encoding D-type glutenin subunits on chromosomes 1B and 1D at or near the *Gli-1* loci (loci of ω-gliadins), with the proteins encoded by the genes on chromosome 1D having similar electrophoretic mobility and N-terminal sequences to the ω-gliadins (Masci *et al.*, 1991b; Masci *et al.*, 1993).

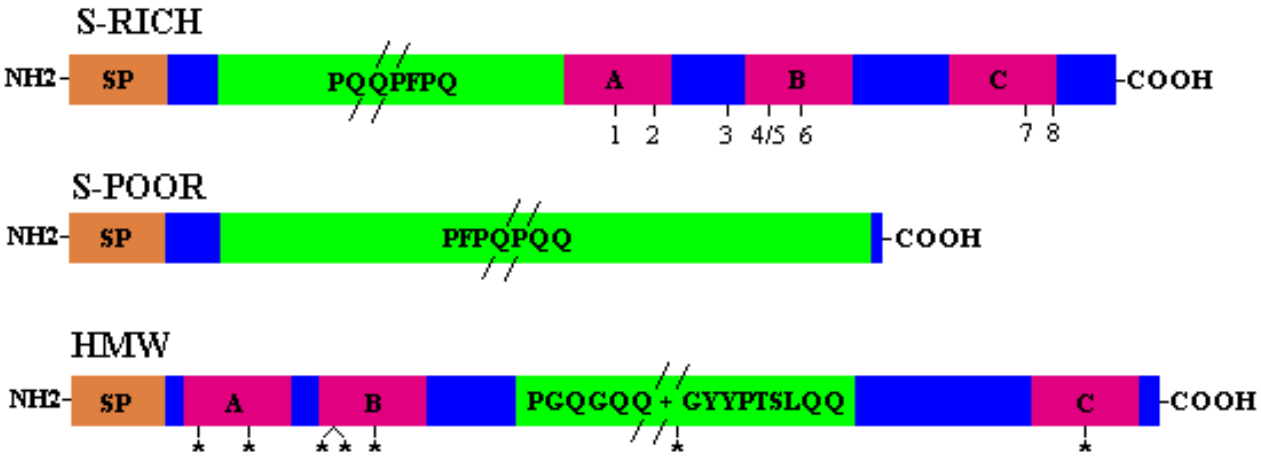


Figure 1.4 Generalised structures of the prolamin families

The proteins depicted are γ-gliadin (wheat) for the S-rich prolamin class, C hordein (barley) for the S-poor class and HMW subunit 1By9 (wheat) for the HMW prolamin class (Shewry *et al.*, 1995). Amino (NH2) and Carboxyl (COOH) termini are shown. 1-8: SC1-SC8 cysteine residues involved in intra-subunit disulfide bonds in the S-rich prolamins. *****: Cysteine residues in the HMW subunits. A/B/C: Related regions in the S-rich and HMW subunits. SP: Signal peptide.

1.4.1.2 S-rich prolamins

As their name suggests, the S-rich prolamins contain cysteine residues, allowing the formation of inter- and intra-subunit disulfide bonds. This group of prolamins represents the major storage protein group in wheat, accounting for 80-90% of the total prolamin fractions and comprise a diverse family, containing a monomeric fraction, wherein all of the cysteine residues are involved in intra-protein disulfide bonds and a polymeric fraction wherein at least some cysteines are involved in inter-protein disulfide bond formation. Traditionally, the monomeric S-rich prolamins are classified into three groups, the α, β and γ gliadins, based on their mobility on SDS-PAGE, but analysis of amino acid sequences has revealed a close relationship between the α- and β- gliadins (Tatham *et al.* 1990), which are now commonly grouped together as the α-gliadins (Figure 1.3). The γ-gliadins form a second, distinct family, which also includes the polymeric LMW-glutenins, containing cysteine residues involved in both intra- and inter-protein disulfide bonds. The S-rich LMW glutenins are also divided, based on mobility on SDS-PAGE, into the C-type LMW-glutenins, related to the α and γ gliadins and a third group, the B-type LMW-glutenins (Figure 1.3) (Payne and Corfield, 1979; Jackson *et al.*, 1983).

All S-rich prolamins consist of 250 to 300 residues, with 1/3 to 1/2 of their structure consisting of proline/glutamine-rich repeat motifs which vary between the three subfamilies: α-gliadins: PF/YPQ_{3-6}; γ-gliadins: $PFPQ(PQQ)_{1-2}$; and LMW-glutenin: $P_{1-2}FP/SQ_{2-6}$ (Anderson and Blechl, 1998), while their N-termini consist of short, unique amino acid sequences (Fig 1.4). The remainder of these proteins consist of non-repetitive sequences that contain three short conserved domains, known as A, B and C, that are related to one another and to the HMW-prolamins, providing evidence of a close evolutionary relationship between the different members of the prolamin superfamily (Kreis *et al.*, 1985; Kreis and Shewry, 1989) (Figure 1.4). Sequences outside of these domains are more variable within the S-rich prolamin family.

The pattern of disulfide bond formation in the S-rich prolamins has been an area of considerable research due to its implications to the functional properties of dough. The non-repetitive C-terminal sequences of the S-rich prolamins contain cysteine residues at

8 conserved positions, SC1 to SC8 (Shewry and Tatham, 1997). The α-gliadins contain 6 of these residues (missing SC2 and SC3), the γ-gliadins contain all 8 and the LMW-glutenins contain 6 (missing SC6 and SC8) (Shewry and Tatham, 1997). The pattern of intra-protein disulfide bonds in these proteins, as proposed by Thompson *et al* (1994) and further clarified by Müller *et al.* (1998), involves bonds between residues SC1/SC4, SC2/SC3, SC5/SC7 and SC6/SC8 (Figure 1.4). The LMW-glutenins contain two additional residues that are not able to form an intra-protein disulfide bond and are therefore believed to be free for polymer formation through inter-protein bonds. One of these cysteine residues is located in the non-repetitive sequence at the N-terminal, and the other between SC5 and SC7 (Köhler *et al.*, 1993; Keck *et al.*, 1995; D'Ovidio *et al.*, 1997). Although gliadins are considered to be primarily monomeric, the isolation of bound γ-gliadin subunits from the polymeric glutenin fraction has revealed a γ-gliadin with a Phe→Cys substitution in the repetitive domain. This γ-gliadin does not have a second free cysteine to form an intra-protein disulfide bond, thus allowing incorporation of this subunit into polymers with LMW and HMW glutenins (Shewry and Tatham, 1997).

1.4.1.3 HMW prolamins

The HMW prolamins represent the most extensively studied group of gluten proteins due to their central role in determining the functional properties of wheat dough. An understanding of the genes encoding these proteins is essential to understand how allelic variations and the deposition of these proteins with other glutenin (and some gliadin) subunits in the endosperm affects dough rheology (Section 1.4.2).

The chromosomal locations of the genes encoding the HMW glutenin subunits were first determined during the 1970's by Bietz *et al*. (1975) using the nullisomic/tetrasomic lines of wheat (Section 1.8.3). It is now firmly established that the HMW glutenins are encoded by two tightly linked genes on the short arms of the group 1 chromosomes, at the *Glu-A1, Glu-B1 and Glu-D1* loci. These two genes encode units of high and low molecular weight, referred to as x- and y-types, respectively. Despite the presence of six HMW glutenin gene loci in bread wheat, only three to five genes are expressed in various cultivars due to the silencing of different alleles. The y-type gene at *Glu-A1* is

always silent in *T. aestivum* and the x-type gene at *Glu-A1* and the y-type gene at the *Glu-B1* locus are only expressed in some cultivars. Early analysis of allelic variation of the HMW glutenin genes, through electrophoretic analysis of 300 wheat cultivars, revealed three alleles encoded by the *Glu-A1* locus, eleven by *Glu-B1* and five by *Glu-D1* and a numbering system was devised for describing these alleles (Payne and Lawrence, 1983).

The proteins encoded by the x- and y-type genes reveal a similar structure, consisting of unique N- and C-termini and a central domain consisting of a repetitive motif of variable length from 481 to 696 residues (Figure 1.4) that explains much of the variation in the size of the subunits, varying from 627 residues (1Dy10) to 827 (1Dx5) and molecular weights of 67,476 Da to 88,128 Da (Shewry *et al.*, 2003). One of the major differences between the proteins encoded by these two gene types is the presence of tripeptide, hexapeptide and nonapeptide motifs in the x-type subunits and only the latter two in the y-type subunits. These repeat motifs are rich in glutamine (~35 mol%), glycine (~20 mol%) and proline (~10 mol%) with high conservation of glutamine positions, essentially forming a glutamine backbone, which may have some functional significance in the association of these subunits in the gluten polymer (Section 1.4.2). Further, the relatively high amount of proline may be significant in determining the conformation of these proteins due to *cis/trans* proline isomerism. Van Dijk *et al.* (1996), using model cyclic peptides of the repetitive domain of the HMW glutenins, found 50% of proline residues occur in the *cis* conformation in the repetitive motif YPTS, while 90% of the remaining prolines occur in the *trans* conformation. Another difference with potential functional significance is the number of cysteines in the x- and y-type subunits. The unique N- and C- terminal domains contain most (or all, in some alleles) of the cysteine residues in the protein, with three (x-type) and five (y-type) in the N-terminal domain and one (in all subunits) in the C-terminal domain (Shewry and Tatham, 1997) (Figure 1.4). An additional cysteine is present in the repetitive domain of some subunits (e.g. 1Dx5) (Shewry and Tatham, 1997).

HMW glutenins are found as polymers in the storage protein bodies, thus requiring the formation of inter-protein disulfide bonds as well as non-covalent interactions with other subunits. Little is known about the actual organization of the individual subunits

within these polymers, but some data is available from the reduction of the gluten polymer and subsequent stabilisation of the resulting oligomers. These studies reveal that many of these disulfide-bound oligomers consist of HMW x-y dimers (Lawrence and Payne, 1983; Werner *et al.*, 1992), in a head-to-tail arrangement, i.e., the C-terminal of one subunit is connected to the N-terminal of the adjacent subunit (Tao *et al.*, 1992). This arrangement of alternating x-y subunits forming the backbone of the gluten polymer is consistent with the model proposed by Graveland *et al.* (1985). The analysis of peptides produced by enzymic digestion of gluten polymers and analysis of disulfide bonds between the products (Köhler *et al.*, 1993; Keck *et al.*, 1995) revealed further disulfide bonding between HMW subunits and LMW subunits, suggesting that the LMW subunits are present as branches off the HMW subunit backbone (Figure 1.5). In addition, these studies identified y-type homodimers (Köhler *et al.*, 1993; Keck *et al.*, 1995), not detected in other studies (Lawrence and Payne, 1983; Werner *et al.*, 1992), possibly due to their low frequency. These different arrangements of disulfide bonds have led to a theoretical model of the glutenin polymer based around an HMW-glutenin backbone with LMW-glutenin branches (Shewry *et al.*, 2003) (Figure 1.5).

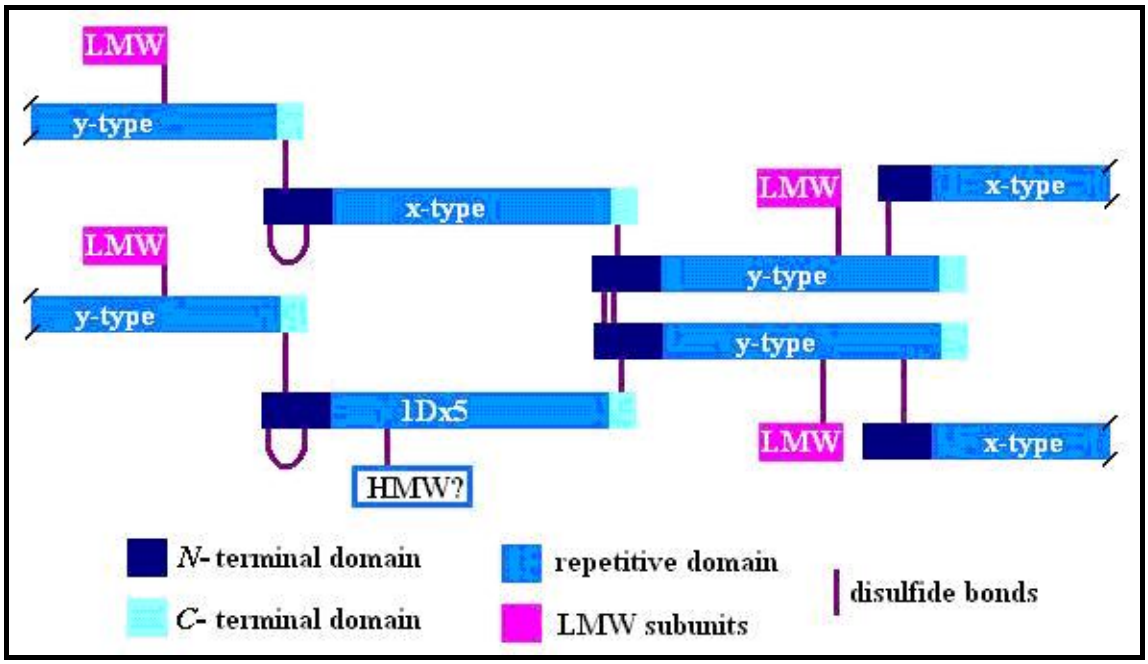


Figure 1.5 Hypothetical structure of wheat glutenin polymers.

The basic structure shows a head-to-tail arrangement of x- and y-type HMW glutenin subunits with branches of LMW-glutenins (Shewry *et al.*, 2003). The presence of an additional cysteine in the repetitive domain of the 1Dx5 subunit may allow the interaction with another HMW glutenin subunit, leading to the stronger doughs observed in cultivars containing this allele.

1.4.2 Role of the storage proteins in determining dough rheological properties

The molecular basis of the balance between elasticity and extensibility in wheat gluten that allows the use wheat dough in a wide range of end-use products has been the subject of investigation for many years. A direct association between particular storage proteins and bread-making quality was established by Payne *et al.* (1979) followed by the association of other HMW subunits with good bread-making qualities, including 1Dx5 + 1Dy10 and 1Ax2* (Moonen *et al.*, 1982, 1983). Field *et al.* (1983) demonstrated that cultivars that contain higher amounts of polymeric protein have stronger (highly visco-elastic) doughs, while those with a lower polymeric:monomeric protein ratio have weaker doughs (Fido *et al.*, 1997). This provides a general rule that the glutenin component is largely responsible for the elastic properties and the monomeric gliadins are responsible for the extensible properties of gluten (Fido *et al.*, 1997), the balance between these two properties being responsible for the unique visco-elastic properties observed in the dough of a particular wheat cultivar. Theories on molecular basis of the visco-elasticity of the gluten network revolve around the structure of the HMW subunit backbone and its interactions with other subunits in the polymer as well as with the monomeric gliadin fraction. Thus, the molecular characterisation of the individual glutenins and gliadins (especially the HMW glutenins) has been, and continues to be, the focus of research aimed at determining how differences in the individual proteins affect the organization of the gluten polymer and consequently the bread-making qualities of different cultivars of wheat.

Evidence for an increase in dough strength through the presence of more disulfide bonded polymeric protein has been suggested by different experimental approaches. For example, the incorporation of purified glutenin polymers was found to increase the resistance of dough to extension, but only when free thiol groups were available to allow incorporation of the added proteins to the gluten polymer (Schropp and Wieser, 1996; Antes and Wieser, 2001). Likewise, incorporation of exogenous HMW subunits into the gluten polymer through a reduction/reoxidation cycle, allowing the formation of disulfide bonds between the incorporated and endogenous subunits, always led to an increase in dough strength, while incorporation of LMW subunits led to varying results, depending on the type of subunit incorporated (Sissons *et al.*, 1998). Further, an

increase in HMW:LMW subunit ratio led to an increase in dough strength and reduced extensibility (Uthayakumaran *et al.*, 2000). Incorporation of analogue glutenin subunits comprising most of the repetitive domain of the HMW subunit 1Dx5 with 0, 1 or 2 cysteine residues at each of the N- and C-termini revealed that incorporation of the subunit containing two free cysteine residues at each end resulted in stronger doughs (Buonocore *et al.*, 1998). In a similar experiment involving incorporation of analogue glutenins based C-hordein from barley, it was found that subunits with a free cysteine at a single terminus act as chain terminators, reducing the size of the gluten polymers and thus dough strength, while subunits with a free cysteine at each terminus increased dough strength, presumably through chain extension and increase in polymer size (Tamás *et al.*, 2002).

Although the disulfide bonds between glutenin subunits clearly play an important role in determining the elastic properties of gluten through the stabilization of this massive polymer, other studies highlight the importance of non-covalent interactions, especially hydrogen bonding. This was suggested by experiments wherein esterification of glutamine residues in gluten resulted in reduced resistance to extension and the use of deuterium oxide (D_2O) instead of water results in increased resistance (Beckwith *et al.*, 1963; Mita and Matsumoto, 1981). More recent spectrophotometric analyses have further highlighted the role of hydrogen bonding in determining the physical properties of gluten (Belton *et al.*, 1994; Belton *et al.*, 1995; Wellner *et al.*, 1996; Belton *et al.*, 1998; Gilbert *et al.*, 2000). These studies led to a hypothetical model being proposed by Belton (1999) describing the basis of elasticity in wheat gluten, wherein it is proposed that at low hydration levels there are many protein-protein interactions between the storage proteins, primarily through H-bonding between glutamine residues in a β-spiral conformation. An increase in hydration leads to a β-sheet-like conformation between neighbouring protein chains in a ‘train’ conformation, with further hydration leading to an increase in H-bonding between the glutamines and water at the expense of glutamine-glutamine interactions, resulting in ‘loops’ of peptide chains H-bonding with water, the equilibrium between ‘loops’ and ‘trains’ being determined by the hydration level (Figure 1.6). Extension of the dough will thus lead to stretching of ‘loops’ and separation of ‘trains’. This mechanical alteration from the equilibrium of ‘loops’ and ‘trains’ would lead to the storing of potential energy and a gradual increase in

resistance, as is observed during dough mixing (Figure 1.6). Continued application of extension would, however, lead to breaking of disulfide bonds, breakdown of the gluten polymer and subsequent weakening of the dough, also observed through over-mixing of dough. It must be noted that both an increase in length of the repetitive domains of the HMW glutenin subunits as well as that in the number of subunits making up the glutenin backbone will increase the number of protein-protein interactions and therefore resistance to extension, explaining the contribution that HMW glutenin alleles with longer repetitive domains and/or an additional cysteine residue available for inter-protein bonding (e.g. 1Dx5) make to bread-making qualities. Further, Belton (1999) describes the contribution of the significant number of monomeric gliadins present in the dough as providing a viscous medium that also contributes to the resistance to extension through non-covalent interactions such as Van de Waals forces.

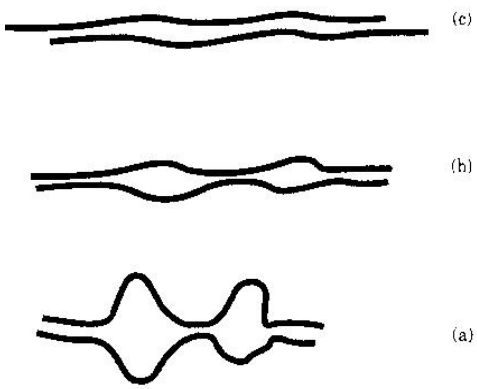


Figure 1.6 Effect of dough extension on gluten polymer interactions

(a) The equilibrium configuration of loops and trains between neighbouring proteins; (b) The deformation of 'loops' caused by dough extension; (c) The eventual separation of 'trains' caused by further dough extension breaking the inter-protein hydrogen bonds and allowing the chains to slip over each other. This deformation from the equilibrium provides the potential energy that explains the elastic properties of gluten.

1.4.3 Storage protein deposition and protein body formation

The models described above, outlining the structure and role of the wheat storage proteins in the gluten polymer, provide the basis for the central role of the storage proteins in determining the unique visco-elastic properties of dough. However, it must be noted that these storage proteins must be first synthesised in the developing

endosperm cells then be folded into their correct conformations and interact with other storage proteins to form the building blocks of the gluten polymer described above. These processes and the role molecular chaperones and, more precisely, 'foldase' enzymes may play in directing and regulating these are described below.

A comparison of the putative protein products encoded by the storage protein cDNA sequences and the N-termini of mature proteins reveals an N-terminal signal peptide responsible for targeting the storage proteins to the secretory pathway within the developing endosperm cells. This pathway involves synthesis of the storage proteins on the surface of the rough endoplasmic reticulum (RER), translocation into the lumen of the ER, cleavage of the signal peptide, folding and disulfide bond formation. As the physiological role of the storage proteins is to provide an adequate supply of amino acids which can be mobilised rapidly during seed germination and growth of the developing plant, it is expected that the storage proteins would be packaged in an orderly manner, as opposed to a random interaction and aggregation, as supported by the non-random organization of the gluten polymer (Sections 1.4.1.2 and 1.4.1.3). Further, the presence of soluble intermediates in the folding of glutenins and gliadins, preventing the random aggregation of these proteins in the ER, has been demonstrated (Shimoni and Galili, 1996; Orsi *et al.*, 2001).

The storage proteins are observed as large protein bodies in the endosperm cells, wherein they coalesce during late grain development to form a continuous protein network surrounding the starch granules and two pathways have been identified in this process. Some proteins, primarily the monomeric gliadins, appear to follow the standard secretory pathway, i.e., via the Golgi apparatus to the vacuole, where they coalesce to form protein bodies. The other pathway involves direct accumulation of protein bodies, presumed to consist primarily of glutenins, within the lumen of the ER (Rubin *et al.*, 1992). The mechanisms of this selection process or its regulation are unknown. Analysis of amino acid sequences of the storage proteins reveals no classical ER-retention signal, –KDEL or –HDEL (Denecke *et al.*, 1992). The expression of HMW glutenins reveals accumulation of these proteins directly in the ER (Shani *et al.*, 1994) and a similar experiment involving expression of γ-gliadin in transgenic tobacco reveals that it is transported via the secretory pathway, presumably to the vacuole,

where it is degraded (Napier *et al.*, 1997). Although no direct evidence is available, it is presumed that the monomeric gliadins in their soluble intermediate form are competent for transport via the Golgi apparatus, whereas the glutenins form high M_r polymers through the formation of disulfide bonds in the oxidising environment of the ER, which subsequently form insoluble aggregates that cannot be transported via the classical secretory pathway. Levanony *et al*. (1992) have proposed that the ER-derived protein bodies are later enclosed in vesicles and transported to the vacuole, where they fuse to form a single population.

The organization of subunits within the gluten polymer, as well as the selective transport of gliadins in a soluble form implies roles for molecular chaperones and/or 'foldase' enzymes, i.e., in preventing the random aggregation of the nascent storage proteins, directing the correct folding of these proteins and association of the glutenin subunits to form polymers. These enzymes include the molecular chaperone Binding Protein (BiP) as well as the foldase enzymes, protein disulfide isomerases (PDI) and peptidyl-prolyl *cis-trans* isomerases (PPIase). The focus of this study is on PDI and a subfamily of PPIase, the cyclophilins. As such, a review of our current knowledge of these enzyme families is necessary.

1.5 PROTEIN DISULFIDE ISOMERASE (PDI)

1.5.1 The structure of PDI

PDI is a member of the thioredoxin superfamily of proteins, characterised by the presence of at least one thioredoxin fold, consisting of a conserved secondary structure encompassing the redox-active site CXXC. PDI was first isolated from rat liver by Goldberger *et al*. (1963) and subsequently became the focus of many studies to elucidate its *in vivo* roles: PDI catalyses the oxidation, reduction and isomerization of incorrectly paired disulfide bonds in nascent proteins, suggesting an important role in ensuring the correct folding of nascent proteins into their native conformations. In support of this central role, it appears to be ubiquitous amongst eukaryotes and has been localised to the oxidising environment of the ER at near-millimolar concentrations (Lyles and Gilbert, 1991) and also occurs at other cellular locations at much lower concentrations (Turano *et al.*, 2002). Analysis of the amino acid sequence of PDI has

revealed that it has a modular structure, primarily made up of four domains, all of which have a thioredoxin-like conformation. Two of these domains, termed a and a', are repeated domains that contain the redox-active site, CXXC, which is involved in the catalytic functions of the enzyme (Figure 1.7), while the other two domains, b and b', which do not share obvious amino acid sequence homology with the a and a' domains and do not contain the redox-active site, but still assume a thioredoxin fold-like conformation (Kemmink *et al.*, 1999) and play an important role in the activity of PDI (described below). In addition, PDI has an acidic domain at its C-terminal (the c domain) that is believed to be a high volume, low-affinity, Ca^{2+}-binding domain (Van Nguyen *et al.*, 1989). The amino acid sequence also reveals a putative N-terminal signal peptide for translocation into the ER lumen, and a C-terminal tetrapeptide, KDEL (HDEL in yeast), which acts as an ER retention/retrieval signal (Denecke *et al.*, 1992) (Figure 1.7).

1.5.2 The catalytic activity of PDI

Experiments in the yeast *Saccharomyces cerevisiae* have shown that PDI is essential for viability (LaMantia *et al.*, 1991; Scherens *et al.*, 1991) and suggested that it is the disulfide isomerase activity, not the oxidation or reduction of disulfide bonds, that is essential (Laboissiere *et al.*, 1995); presumably other redox-active proteins within the ER can complement the oxidation/reduction activities in PDI-null mutants. However, recent reports dispute these findings; Xiao, *et al.* (2004) revealed that an isomerase-deficient strain of *S. cerevisiae* is viable, even when all PDI homologues in the ER are deleted, although the efficiency of the oxidative folding of carboxypeptidase Y was greatly compromised, while Solovyov *et al.* (2004) report that the oxidative domains of PDI are essential for viability. The PDI-mediated isomerization of disulfide bonds in nascent proteins has been analysed through an investigation of folding intermediates in a number of model proteins including RNase (Creighton, 1979; Konishi *et al.*, 1982), bovine pancreatic trypsin inhibitor (Creighton and Goldenberg, 1984), and hirudin (Chatrenet and Chang, 1992). These studies suggest that the process of disulfide bond formation is 'hit and miss', as proteins do form incorrect disulfide bonds, which are subsequently reduced through the formation of a disulfide bond with one of the redox-active sites of PDI, providing a free thiol group in the nascent protein which can form

an alternative disulfide bond. The second active site is believed to be important for this process, acting as a molecular clock which can release PDI from the substrate if the rearrangement of disulfide bonds is too slow (Gilbert, 1997) and mutations at this site lead to accumulation of PDI-substrate complexes and a subsequent rapid decrease in isomerase activity as the enzyme becomes unable to release itself (Walker *et al.*, 1996). This 'hit and miss' process continues, with formation of different arrangements of disulfide bonds until the 'correct' conformation is achieved and the nascent protein can continue through the secretory pathway to its ultimate destination.

Figure 1.7 Schematic representation of PDI

a and a': the thioredoxin-like redox-active domains; b and b': the thioredoxin-like redox-inactive domains; c: the acidic C-terminal domain; SP: signal peptide; CXXC: redox-active site; KDEL: ER retention/retrieval tetrapeptide.

1.5.3 Additional roles of PDI

In addition to its catalytic activities, PDI has been implicated as a molecular chaperone in a number of studies, although in a substrate-specific role. This activity appears to be independent of the redox activities and has been found *in vivo* for lysozyme (Hayano *et al.*, 1995) and *in vitro* for glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (Cai *et al.*, 1994), rhodanase (Song and Wang, 1995), phospholipase A_2 (Yao *et al.*, 1997) and proinsulin (Winter *et al.*, 2002), however, the folding of antibody Fab fragments does not appear to require this activity (Lilie *et al.*, 1994). An important property of PDI that permits it to perform chaperone functions and may also play an important role in its catalytic activities, is its ability to bind to a wide range of, primarily misfolded, proteins. The nature of these interactions is unclear, however the b' domain and the C-terminal 57 amino acids, encompassing the last 3 amino acids of the a' domain and most of the c domain, appeared to be central to the peptide-binding functions (Noiva *et al.*, 1993; Klappa *et al.*, 1998a). Further, elimination of the C-terminal peptide-binding site led to

a reduced rate of protein folding in lysozyme and acidic phopholipase A_2, both disulfide-containing proteins (Dai and Wang, 1997). Based on these findings, a model of synergy between the peptide-binding sites and the catalytic activities has been postulated by Freedman *et al.* (2002), wherein misfolded proteins are bound by the peptide-binding sites and catalysis of disulfide bond formation and isomerization is performed by the a and a' domains.

The protein-binding and chaperone properties of PDI appears to be central in another *in vivo* role of this enzyme, in acting as the subunit of other enzyme systems. For example, PDI has been found to act as a subunit of the microsomal triacylglycerol transfer protein (MTP), wherein dissociation of PDI from the other subunit of this enzyme results in its aggregation and loss of MTP activity (Wetterau *et al.*, 1991). Likewise, PDI has been identified as the β-subunits of prolyl-4-hydroxylase (P4H), an ER-localised enzyme tetramer ($\alpha_2\beta_2$) involved in the synthesis of collagens. On its own, PDI does not have P4H enzyme activity and mutations of the CXXC sites do not affect the ability of PDI to act as the subunit of this enzyme system (Vuori *et al.*, 1992). The role of PDI in this enzyme system is thus believed to be to ensure the α-subunit, which lacks an ER-retention/retrieval signal, remains in the ER in a soluble, active conformation (Vuori *et al.*, 1992).

1.5.4 PDI-related proteins

PDI represents the first and best characterised member of a larger family of PDI-related proteins, wherein most members have redox-active sites and are believed to catalyse reactions similar to PDI and perform similar chaperone functions (Ferrari and Söling, 1999). The main differences between the different members of this family may be in their modular structure (Figure 1.8), tissue distribution, substrate specificity, expression levels and regulation. The classification and nomenclature of the PDI family members have varied in different reports, however in a review by Ferrari and Söling (1999) they have been given standard names based on their amino acid sequences; the following description of the PDI family members follows that system.

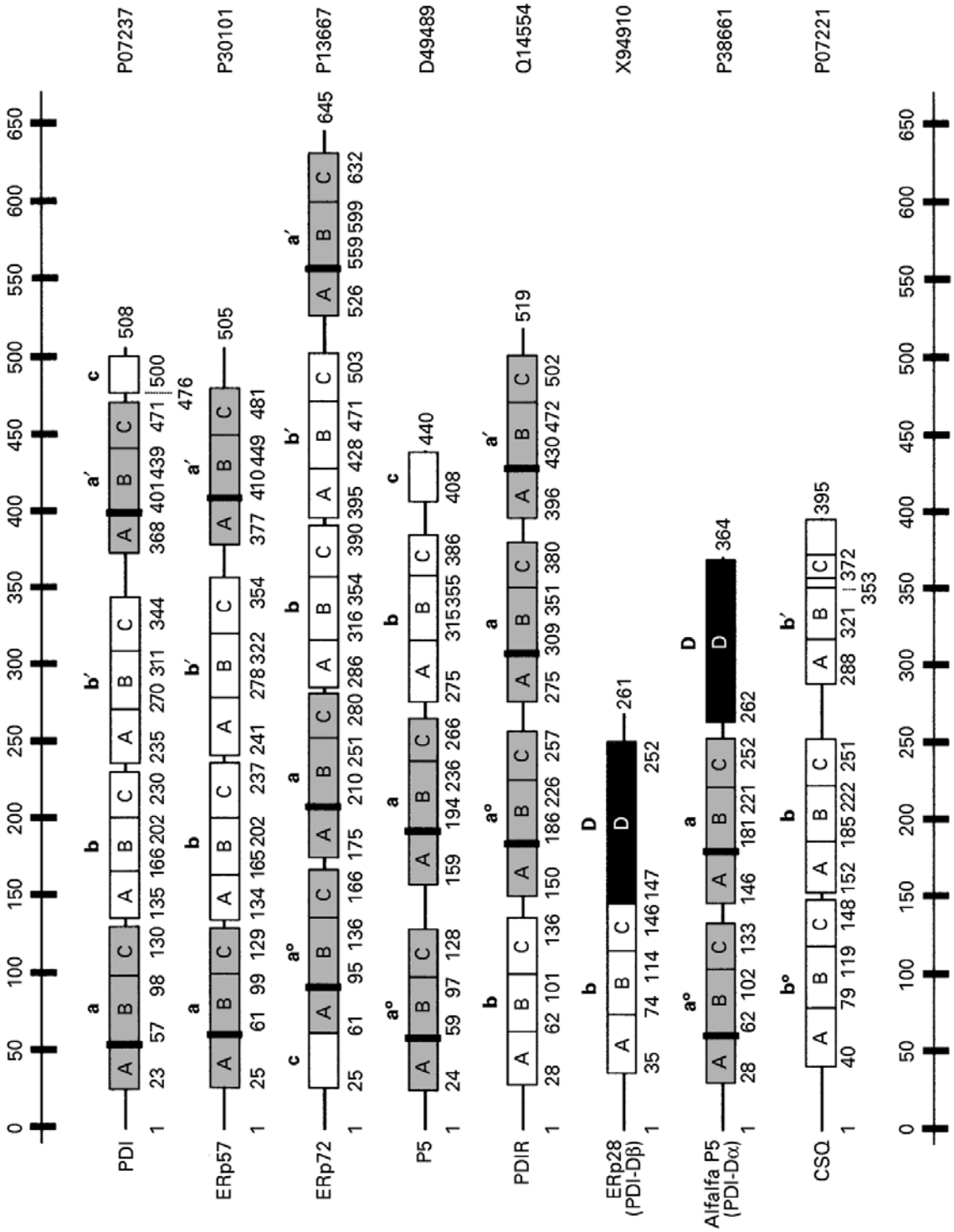


Figure 1.8 PDI and related proteins

a, a' and a^o: thioredoxin-like, redox-active domains; b, b' and b^o: thioredoxin-like, redox-inactive domains; c and D: domains found in some PDI proteins; black bars: the redox-active site position; A, B and C: conserved secondary structures of the thioredoxin fold. Accession numbers are provided next to each diagram. Figure from Ferrari and Söling (1999).

- ERp57 is an ER-localised member of the PDI family, with a structure similar to PDI and the ER-retention/retrieval signal of QEDL but lacking the c-domain. Unlike PDI, it has been shown to interact with calnexin or calsequestrin and through this interaction with monoglycosylated glycoproteins (Elliott *et al.*, 1997; Oliver *et al.*, 1997). This interaction has led to suggestions that ERp57, calnexin and calsequestrin are responsible for the folding of glycoproteins in the ER (High *et al.*, 2000), whereas glycosylation status does not affect the interaction of PDI with nascent proteins (Oliver *et al.*, 1997). ERp57 appears to also have a different specificity for peptide binding as it cannot substitute for PDI in the P4H enzyme system (Koivunen *et al.*, 1996).
- ERp72 has three redox-active thioredoxin domains (Figure 1.8) and like PDI and ERp57 it is localised to the ER, appears to be induced by stress and in antibody-producing cells and been shown to have peptide binding ability (Klappa *et al.*, 1998b), although it does appear to interact with substrate proteins directly through disulfide bond formation (Feng *et al.*, 1996).
- Less is known about other members of the PDI family. PDIp has a modular structure similar to ERp57, however it has only been isolate in vertebrates and is solely expressed in the pancreas (Desilva *et al.*, 1997; Klappa *et al.*, 1998b). P5 has also only been identified in vertebrates and represents the smallest redox-active member of the PDI family that contains an inactive b-type module (Ferrari and Söling, 1999). Little is known about PDIR, however, its three thioredoxin-like active sites (Figure 1.8) all have different amino acid sequences (Hayano and Kikuchi, 1995). PDI-D proteins are the only members that contain a domain unrelated to thioredoxin, termed the D-domain (Figure 1.8) and can be either redox-active (PDI-Dα) or -inactive (PDI-Dβ). A new subfamily of plant PDI has recently been postulated based around the PDIL1 enzyme in carrot (see below).

1.5.5 PDI in plants

Genes encoding PDI have been isolated from a number of plant species including wheat (Shimoni *et al.*, 1995a; Johnson *et al.*, 2001), barley (Chen and Hayes, 1994), maize (Li and Larkins, 1996), rice (Takemoto *et al.*, 2002), castor bean (Coughlan *et al.*, 1996) and alfalfa (Shorrosh and Dixon, 1991). Analysis of the putative protein products of

these genes revealed high conservation of PDI sequences in plants ranging from 80% identity between maize, wheat and barley to 62% identity with alfalfa (Li and Larkins, 1996). In addition to genes encoding the archetypal PDI, the PDI-Dα family has been isolated in alfalfa (Shorrosh and Dixon, 1992) and a novel PDI family encoded by *PDIL1* in carrot and closely related genes in tomato and *Arabidopsis* have been described (Xu *et al.*, 2002).

In wheat, Roden *et al.* (1982) established the presence of PDI in the developing endosperm, an observation subsequently supported by Shimoni *et al.* (1995b), who demonstrated that PDI is also co-localised with the protein bodies. Analysis of the expression patterns of PDI in the wheat endosperm, through Northern (Grimwade *et al.*, 1996; DuPont *et al.*, 1998) and Western blotting (Shimoni *et al.*, 1995b; DuPont *et al.*, 1998), revealed that PDI is upregulated in the developing wheat endosperm, with rapid increase in both transcripts and enzyme from 6-9 days post-anthesis (DPA), with subsequent decline in transcript levels, while the protein levels remain high until around 20 DPA and then decline. Investigations into the gene family encoding PDI in hexaploid wheat have revealed four genes, localised to the short arms of chromosomes 1B, 4B and 4D and the long arm of chromosome 4A and restriction fragment length polymorphisms (RFLPs), especially for the gene on chromosome 1B, and absence of this gene in some genotypes of common and durum wheat (Ciaffi *et al.*, 1999; 2000). The expression of the four genes in wheat has not been directly investigated, however, our previous work led to isolation of three distinct types of *PDI* cDNA clones from a wheat endosperm cDNA library suggesting that at least three of the *PDI* genes are actively expressed in the endosperm (Johnson *et al.*, 2001).

1.5.6 Evidence supporting a role for PDI in the formation of protein bodies

The potential role for PDI in facilitating the efficient formation and deposition of protein bodies in the wheat endosperm is clear (Section 1.4.3); however, direct experimental evidence for this is limited. An elegant, early experiment involved the *in vitro* folding of γ-gliadin in the presence of dog microsomes (Bulleid and Freedman, 1988). When the microsomes were depleted of luminal proteins, the ability of the γ-gliadins to form disulfide bonds was impaired, however, addition of purified PDI

improved the folding of γ-gliadins. Shimoni and Galili (1996) used pulse chase experiments to reveal the *in vivo* orderly assembly of gliadins into soluble intermediates before their deposition into protein bodies. This study also showed that treatment with the reducing agent DTT led to the premature aggregation of the gliadins *in vivo* while treatment of the soluble, oxidised gliadins with DTT *in vitro* also led to the rapid aggregation of the individual subunits, suggesting an important role for correct disulfide bond formation in the deposition of protein bodies. Comparison of the expression levels of PDI (Section 1.5.5) and the storage protein in the developing endosperm, at both the mRNA and protein levels, have been used to both argue for (DuPont *et al.*, 1998) and against (Grimwade *et al.*, 1996) a role for PDI in the deposition of the protein bodies. A recent study on the *N*-methyl-*N*-nitrosourea (MNU)-induced rice mutant esp2, which exhibits defective storage protein segregation and deposition, has now provided compelling evidence for a role for PDI in storage protein deposition (Takemoto *et al.* 2002). In esp2 the glutelin-containing protein bodies form as normal, however, the prolamin-containing protein bodies are smaller and contain aggregates of both prolamins and glutelin-precursors unlike those observed in wild type rice. It was only possible to extract the glutelins after removal of the cysteine rich prolamins, suggesting the formation of disulfide bonds in the aggregated polymer. The mutation in esp2 was found to be an absence of PDI expression, providing direct evidence for a role of PDI in the deposition of storage proteins in rice, prolamins in particular. The direct relevance of this observation to wheat is currently unclear; investigation into orthology between the *esp2* locus in rice and the corresponding *PDI* loci in wheat would be required to suggest the involvement of one (or more) of the wheat PDI genes in a similar process. With this in mind, our current knowledge of gene orthology between members of the grass family, especially rice and wheat, and the potential significance of such information, is reviewed briefly later (Section 1.9).

1.6 PEPTIDYL PROLYL *CIS-TRANS* ISOMERASES (PPIASE)

1.6.1 The PPIase superfamily

In addition to the formation of disulfide bonds, another rate-limiting step in the folding of nascent proteins into their final conformation is the isomerization of peptide bonds preceding proline into *trans* or *cis* forms (Herzberg and Moult, 1991). In general, most peptide bonds in mature proteins are in the *trans* conformation, but approximately 6.5% of the peptide bonds preceding proline residues in correctly folded proteins are found in the *cis* conformation (Galat and Rivière, 1998). As most proteins are translated with their peptide bonds in the *trans* conformation, the isomerization of these peptide bonds into their *cis* form must occur for them to attain their final folded state (Figure 1.9). This process was recognized as one of the major rate-limiting factors in the folding of nascent proteins and led to the discovery of a protein with PPIase activity (Fischer *et al.*, 1984). A simultaneous search for the intracellular target of the powerful immunosuppressant drug cyclosporin A (CsA) identified an 18 kDa protein, cyclophilin, which was found to be identical to the PPIase enzyme, with CsA inhibiting its activity (Handschumacher *et al.*, 1984).

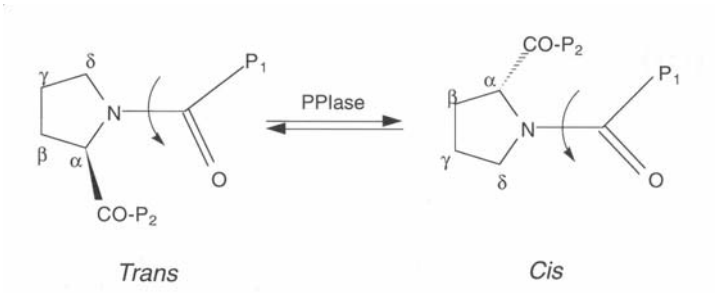


Figure 1.9 The reaction catalysed by PPIase

The *cis-trans* isomerization of peptide bonds preceding prolyl residues in nascent proteins catalysed by members of the PPIase superfamily (Galat and Rivière, 1998).

In addition to the cyclophilin family of PPIase enzymes, identification of a number of other proteins with PPIase activity has since led to the establishment of distinct PPIase subfamilies. The second major subfamily identified were the FKBP (FK506 binding protein) proteins, so named due to their affinity and inhibition by a different class of immunosuppressants, FK506 and rapamycin (Harding *et al.*, 1989). The ‘trigger factor’,

first identified as having PPIase activity in *E. coli* (Stoller *et al.*, 1995) does not bind to FK506 or rapamycin, however, has limited sequence homology to FKBP and is placed either in the same subfamily as FKBP (Callebaut and Mornon, 1995; Göthel and Marahiel, 1999) or in a separate subfamily (Galat and Rivière, 1998). Another subfamily of proteins possessing PPIase activity, first identified in *E. coli*, are known as the parvulins (Rahfeld *et al.*, 1994) which are irreversibly inhibited by juglone (Hennig *et al.*, 1998). As the current study focussed on the cyclophilin subclass of PPIase the various classes of these enzymes and their putative roles in cellular metabolism are described below.

1.6.2 Classes of cyclophilins

Cyclophilins are an ubiquitious class of proteins, found in bacteria, fungi, animals and plants and expressed in a wide range of tissue types (Galat and Rivière, 1998). The 18 kDa cyclophilin, first identified in mammals, appears to be the most abundant cyclophilin subtype, however a number of other classes of cyclophilins with discrete intracellular locations and putative functions have been identified. In yeast, at least eight forms (genes *CPR1* through *CPR8*) of cyclophilin have been identified, ranging in molecular mass from 17 kDa to 45 kDa (Göthel and Marahiel, 1999), while humans have at least sixteen proteins containing at least one domain related to cyclophilin (Galat, 2003; 2004). Of these four smaller cyclophilin forms have been identified ranging in size from 18 kDa to 23 kDa, consisting of a cyclophilin domain in the cytoplasmic hCyp-18a (CyP A) to a cyclophilin domain with a signal peptide to target the cyclophilin to its subcellular location (Table 1.1). A number of paralogues to these smaller cyclophilins have also been identified in the human genome (Galat, 2003; 2004). In addition to these small cyclophilins, a number of larger proteins containing cyclophilin-like domains (CLD) have been identified ranging in size from 26 kDa to 320 kDa, which contain additional non-cyclophilin-like domains (Table 1.1). The mammalian 18 kDa cyclophilin is referred to as cyclophilin A (hCyp-18a; CyPA) and represents the smallest cyclophilin, with the other classes of cyclophilins and CLD-containing proteins exhibiting homology to it, with the presence of N-terminal signal peptides and/or non-homologous N or C-terminal extensions. In addition to the cytoplasmic CyPA the hCyP-22b/p (CyPB) proteins are localised to the ER

(synonymous with S-cyclophilin as this form of cyclophilin has been observed to be secreted from the cell). The yeast ER-resident cyclophilin carries a typical –HDEL ER-retention signal at its carboxyl terminus (Frigerio and Pelham, 1993), however, several cyclophilin B's lack this or the alternative KDEL motif (Denecke *et al.*, 1992) and may have other non-typical motifs instead; e.g., 'VEKPFAIAKE' in vertebrate cyclophilin B's that localises them to specialised sub-compartments of the ER lumen (Arber *et al.*, 1992), or 'HEEL' in a fungal homologue (Derkx and Madrid, 2001). hCyP-22c/p (CyPC) appears to be a larger cytoplasmic or membrane-associated form of cyclophilin (Price *et al.*, 1991; Friedman *et al.*, 1993; Galat, 2003) and hCyP-22d/p (CyPD) is localised to the mitochondria (Bergsma *et al.*, 1991). In addition to the cyclophilins containing signal peptides for subcellular localisation, a number of proteins containing CLDs have been characterised in mammals, many of them apparently nuclear-localised, including:

- A unique 28 kDa protein from bovine retina, a structural homologue of the *ninaA* gene from *Drosophila melanogaster*, involved in folding of opsins (Ferreira *et al.*, 1995);
- A T-cell specific 33 kDa protein localised to the nucleus and containing a CLD and RNA binding domain (Mi *et al.*, 1996);
- Cyp-40, localised to the cytoplasm and, which contains a domain with homology to the P59 steroid receptor complex (Kieffer *et al.*, 1993);
- The 358 kDa protein, nucleoporin, which makes up part of the nuclear pore (Wu *et al.*, 1995);
- The nuclear proteins, CyP-88 which contain a number of nuclear localization signals and serine/arginine (SR) domains which may be involved in pre-mRNA splicing (Nestle *et al.*, 1996).
- The natural killer T-cell-specific receptor Cyp-158/169 involved in tumour recognition (Anderson *et al.*, 1993).

In summary, the diverse cyclophilin family is defined by the presence of a domain with sequence similarity to the archetypal cyclophilin A, with other classes roughly divided into (i) those that have this cyclophilin A-like domain and additional signal peptides that target them to their subcellular location or, (ii) larger multidomain proteins that contain a domain with homology to cyclophilin.

Table 1.1 Classes of mammalian cyclophilins and CLD-containing proteins

Name	Size (kDa)	Intracellular Location	Species	Reference(s)
Cyclophilin A	18	Cytoplasm	*Homo sapiens* *Bos taurus* *Rattus norvegicus* *Mus musculus* *Sus scrofa*	(Harding *et al.*, 1986; Haendler *et al.*, 1987; Danielson *et al.*, 1988; Takahashi *et al.*, 1989; Hasel and Sutcliffe, 1990)
Cyclophilin B	20-23	ER/secreted	*H. sapiens* *R. norvegicus*	(Iwai and Inagami, 1990; Price *et al.*, 1991)
Cyclophilin C	22	Cytoplasm	*M. musculus*	(Friedman and Weissman, 1991)
Cyclophilin D	22-24	Mitochondria	*H. sapiens*	(Bergsma *et al.*, 1991)
CLD-containing Proteins				
CyP-28	28	ER (retina)	*B. taurus*	(Ferreira *et al.*, 1995)
CyP-33	33	Nucleus (T-cells)	*H. sapiens*	(Mi *et al.*, 1996)
CyP-40	40	Cytoplasm	*H. sapiens* *B. taurus*	(Kieffer *et al.*, 1993)
CyP-58	58	Nucleus	*H. sapiens*	(Wang *et al.*, 1996)
CyP-88	88	Nucleus	*H. sapiens*	(Nestle *et al.*, 1996)
CyP-158/169	158/169	Cell Membrane	*H. sapiens* *M. musculus*	(Anderson *et al.*, 1993)
CyP-358 (nucleoporin)	358	Nucleus	*H. sapiens*	(Wu *et al.*, 1995)

1.6.3 Cyclophilins in plants

Cytoplasmic *cyclophilin A* cDNA sequences have been reported from a number of plant species, including canola (*Brassica napus*), tomato (*Lycopersicon esculentum*), maize (*Zea mays*) (Gasser *et al.*, 1990), rice (*Oryza sativa*) (Buchholz *et al.*, 1994), bean (*Phaseolus vulgaris*) (Marivet *et al.*, 1992), foxglove (*Digitalis lanata*) (Scholze *et al.*, 1999), mouse ear cress (*Arabidopsis thaliana*) (Hayman and Miernyk, 1994; Lippuner *et al.*, 1994) and wheat (*Triticum aestivum*) (Johnson *et al.*, 2001). Comparison of these with those from mammals reveals a seven amino acid insertion apparently characteristic of plant cyclophilin As. Other classes have been identified in different subcellular compartments including, the detection of cyclosporin A-sensitive PPIase activity in a chloroplast fraction (Breiman *et al.*, 1992), subsequently revealed to be multiple forms of cyclophilins in the thylakoid lumen, including pCyPB of fava beans (Luan *et al.*,

1994), TLP40 of spinach, associated with intraorganelle signalling and dephosphorylation of photosynthetic proteins (Fulgosi *et al.*, 1998; Vener *et al.*, 1999) TLP20, identified in spinach, rice and *Arabidopsis thaliana*, responsible for most of the PPIase activity within the thylakoid lumen and, thus hypothesised to be the general folding catalyst of chloroplast proteins (Edvardsson *et al.*, 2003) and AtCYP20-2, a light-regulated cyclophilin associated with the photosynthetic membranes of *A. thaliana* (Romano *et al*. 2004a). Cyclosporin A-sensitive PPIase activity has also been reported from a mitochondrial fraction in pea plants (Breiman *et al.*, 1992), but the proteins or genes encoding these enzymes have not been reported. The first report of an ER-localised plant cyclophilin B was the direct isolation of a cyclophilin from a microsomal fraction in maize that had high sequence similarity to the mammalian cyclophilin B (Sheldon and Venis, 1996). A recent, extensive *in silico* analysis of *Arabidopsis* cyclophilins (Romano *et al.*, 2004b) revealed the largest *cyclophilin* gene family in any organism studied to date, consisting of 29 members, including intronless cytoplasmic forms, an intron-containing chloroplast stromal form (Chou and Gasser, 1997), an ER-targeted form lacking the ER-retention motifs mentioned above (Saito *et al.*, 1999), as well as other multidomain cyclophilins.

The high level of gene expression over a relatively short time in the developing wheat endosperm that leads to production of the storage protein bodies implicates folding catalysts such as cyclophilins in these processes. As discussed earlier (Section 1.4), the storage proteins contain a relatively high number of proline residues, some of which may be have peptide bonds in the *cis* conformation, suggesting a potentially important role for PPIases in accelerating and ensuring the correct folding of the nascent storage proteins. It is unclear whether an ER-resident form of cyclophilin is present in the developing endosperm, or whether any other cyclophilin forms play any roles here. It has been demonstrated that cyclophilins are upregulated in the developing endosperm (Grimwade *et al.*, 1996); however, this study did not differentiate between the various forms of cyclophilin. We subsequently characterised three forms of *cyclophilin A* from a cDNA library generated from developing wheat endosperm tissue (Johnson *et al.*, 2001), however, their specific role(s) in this tissue is currently unknown.

1.6.4 Putative roles of cyclophilins

The enzymic activity of cyclophilins, the *cis-trans* isomerization of peptide bonds preceding proline residues, suggests a role in ensuring the correct and efficient folding of nascent proteins. Early efforts confirmed that cyclophilins could catalyse this isomerization reaction in model peptides and ribonuclease A, *in vitro* (Fischer and Bang, 1985; Lang *et al.*, 1987). Evidence for an *in vivo* protein folding role have since been reported including the treatment of hepatoma cells with cyclosporin A leading to the significant slowing of the folding of transferrin and folding of the collagen triple helix (Lodish and Kong, 1991; Steinmann *et al.*, 1991). The cyclophilin-like protein, ninaA, from *Drosophila melanogaster*, was shown to be essential for the folding and translocation of rhodopsin types I and II *in vivo* (Colley *et al.*, 1991), and knockout of mitochondrial cyclophilins in *Neurospora crassa* and *Saccharomyces cerevisiae* resulted in a marked slowing of folding of imported proteins (Matouschek *et al.*, 1995; Rassow *et al.*, 1995).

In addition to their role in protein folding, there are a number of reports describing other physiological roles for cyclophilins including:

- Acting as an intracellular chaperone. The affinity of cyclophilins for their proline substrate is believed to facilitate the binding and stabilisation of proteins to prevent their aggregation (Freskgard *et al.*, 1992). The observed upregulation of many cyclophilins under a variety of stress conditions supports their chaperone role in the stress response (Andreeva *et al.*, 1999), which may or may not be associated with their protein folding activities. In plants, upregulation of cyclophilins has been observed in response to heat (Luan *et al.*, 1994), abiotic stress (Marivet *et al.*, 1992), high salinity, cold (Marivet *et al.*, 1994) and fungal infection (Godoy *et al.*, 2000). Further, the mammalian cytoplasmic CyP-40 has been found to interact with other intracellular chaperones including HSP90 and HSP104 (Duina *et al.*, 1996; Abbas-Terki *et al.*, 2001).
- Involvement in intracellular signalling pathways, due to their interactions with calcineurin, a phosphatase involved in calmodulin and Ca^{2+}-signalling pathways. The binding of cyclosporin A to cyclophilins leads to the inhibition of calcineurin, blocking of the signalling pathway and ultimately the immunosuppressant effects of

cyclosporin A (Liu *et al.*, 1991; Schreiber, 1992). In plants, Luan *et al.* (1993) has demonstrated the involvement of cyclophilins in Ca^{2+}-dependent intracellular signalling in guard cells of *Vicia faba*.

- Involvement in developmental pathways, as they have been observed to be upregulated in developing tissues (Gasser *et al.*, 1990; Marty *et al.*, 1993; Marivet *et al.*, 1994; Grimwade *et al.*, 1996). The nature of the cyclophilins' role in development is unclear, however a role in acceleration of protein folding during tissue development is conceivable.
- Assembly and stabilisation of steroid receptors (Lebeau *et al.*, 1992);
- Regulation of transcription factors (Yang *et al.*, 1995);
- A role in the mitochondrial permeability pore and its implications to cell death (Crompton, 1999).

Clearly, their various intracellular locations, their ability to bind various proteins and their catalytic activity, all allow these enzymes to potentially serve numerous intracellular roles. Further, analysis of cyclophilin structure in bacteria, fungi, plants and animals reveals general conservation throughout evolution (Galat, 1999), indicating a strong selection pressure for maintenance of these proteins and their biological role(s).

1.7 POSSIBLE INTERACTIONS BETWEEN PDI AND PPIASE

1.7.1 A synergistic role for PDI and PPIase?

The roles of PPIase and PDI in accelerating the folding of nascent proteins has prompted investigations into a tantalising possibility, i.e., whether these two functions are associated with each other, especially that of cyclophilin B, which is co-localised with PDI in the ER. Schönbrunner *et al.* (1992) found that *in vitro* refolding of ribonuclease T1, especially the formation of correct disulfide bonds, was facilitated by the presence of a PPIase in conjunction with PDI, and drew the conclusion that formation of disulfide bonds could be more efficient if the prolyl isomers are in the correct conformation. Further, it was found that cyclophilin B improved the chaperone activity of PDI *in vitro* and the complex of cyclosporin A and cyclophilin B completely inhibited this activity (Horibe *et al.*, 2002). In the nematode *Caenorhabditis elegans*, *PDI* and *cyclophilin* genes are clustered together and co-expressed in a functionally

related manner (Page, 1997), suggesting that the efficiency of protein folding may be improved with the presence of both enzymes together. Interestingly, Meunier *et al.* (2002) found that a number of ER-localised foldase enzymes and molecular chaperones, including PDI and cyclophilin B, formed large multiprotein complexes, which were found to be associated with unfolded immunoglobulin heavy chains *in vivo* suggesting a coordination between various proteins involved in the folding and integrity of nascent proteins in the ER.

1.8 GENETIC ANALYSIS IN WHEAT

1.8.1 Benefits and problems with genetic analysis in wheat

Marker technology allows DNA sequence variations in different wheat cultivars to be followed in crosses between the alternative genotypes into their subsequent generations. These markers do not directly affect biological processes; they are generally random polymorphisms between the two parents, however they can be linked to genes or sequences that are involved in determining phenotype. This has obvious benefits to wheat breeders as a sequence variation could represent a particular allele that confers a beneficial (or undesirable) phenotype on the plant for which a marker will permit its rapid detection in the subsequent progeny of the crosses and the selection of desirable alleles. Clearly, the accuracy of this prediction depends on the extent of linkage of the marker to the allele and therefore the likelihood that it will co-segregate with that allele. A marker that assays the presence of a polymorphism within a gene will always co-segregate with that gene, and is therefore said to be a ‘perfect’ marker, for that gene. Thus, marker technology provides a powerful tool for wheat breeders to detect and select for beneficial alleles in their crosses, a technique known as marker-assisted selection (MAS). In addition, marker technology also allows the analysis of the genetic control of simple and complex traits in wheat through the identification of quantitative trait loci (QTLs), discussed below (Section 1.8.5).

The main problem associated with developing markers in wheat is the relatively low level of DNA sequence polymorphism, especially in commercial wheat varieties (Bryan *et al.*, 1999). Due to this, many potential markers need to be assessed in order to find polymorphisms and in addition, the level of polymorphism is not consistent within each

of the six genomes of wheat and thus, the random development of markers, using the techniques outlined below (Section 1.8.4) will result in the unequal distribution of markers, with particularly poor coverage of the D genome, as it exhibits more inter-cultivar sequence conservation than the A or B genomes. This level of conservation is reflected in the relatively lower number of markers mapped to the D genome in the studies discussed below (Section 1.8.4) and has led to the targeting of markers specific to the D genome (Pestova *et al.*, 2000).

The other problem associated with genetic analyses in wheat is the size and complexity of the genome. The wheat genome is approximately 1.6 x 10^{10} bp, i.e., roughly 40 times the size of the rice genome (Arumuganathan and Earle, 1991), leading to problems with commonly used techniques in marker development such as Southern blotting (Langridge *et al.*, 2001). Further, the hexaploid nature of the genome means that special care needs to be taken while designing molecular markers that assay a single locus, as many markers tend to assay the three homoeologous loci simultaneously, and can lead to problems with analysis of results. However, the hexaploid nature of wheat does provide certain benefits, especially for the physical mapping of genes, discussed below (Section 1.8.3).

1.8.2 Types of markers

1.8.2.1 Morphological and biochemical markers

Morphological (visible) markers are based on morphological, physiological or pigmentation phenotypes. These markers have been used in genetic studies since well before DNA was identified as the hereditary material (Morgan, 1911). Biochemical or protein markers are commonly referred to as isozymes or allozymes. These markers have been in use since the late 1950s (Markert and Moller, 1959) and are visualized by electrophoresis and protein staining methods where differences in their mobility in the gel matrix are easily assayed. The major disadvantages of morphological and biochemical markers are the relatively small numbers available and their tendency to be influenced by environmental factors, tissue type and/or developmental stage of the organism (Winter and Kahl, 1995).

1.8.2.2 RFLP markers

The limited numbers of morphological and biochemical markers, which rely on differences in phenotype conferred by genes, led to suggestions that larger numbers of genetic markers may be available due to polymorphisms in the DNA molecule itself. Thus, molecular markers were developed, that could discriminate single nucleotide changes between individual genotypes.

Restriction fragment length polymorphism (RFLP) markers were the first DNA-based markers used to assay such DNA sequence polymorphisms (Botstein *et al.* 1980) and were found to be extremely effective at identifying polymorphisms between cultivars and as such, most of the genetic maps developed for wheat contain many RFLP-based markers. Paull *et al.* (1998) screened 124 varieties of wheat with 119 RFLP probes, of which 98 varieties (82%) were polymorphic. Similarly, Shah *et al.* (2000) screened the cultivars Cheyenne and Wichita with 52 RFLP probes and found that 41 of these (78.8%) were polymorphic. Further, the nature of the typical assay, involving the screening of Southern blots with cDNA clones, means that many of the identified polymorphisms are at or near expressed genes, and probes from one species can easily be transferred to related species. Despite these benefits, the use of RFLPs is declining due to certain disadvantages, primarily, the requirement for large amounts of DNA and the slow nature of the assay. These have led to the development and widespread use of the, Polymerase Chain Reaction (PCR)-based methods.

1.8.2.3 PCR-based markers

A number of PCR-based methods have been used to assay polymorphism in wheat cultivars, including randomly amplified polymorphic DNA (RAPD) (Hohmann *et al.* 1994; Kojima *et al.* 1998), amplified fragment length polymorphisms (AFLP) (Vos *et al*. 1995; Boyko *et al.* 1999) and assay of simple sequence repeats (SSR) (Röder *et al.* 1998; Shah *et al.* 2000; Parker *et al.* 2002). Each of these methods has its own advantages and disadvantages (Table 1.2), however the reliability, ability to automate, effectiveness at detecting polymorphisms, ability to target individual genomes and primarily co-dominant nature (produces different sized PCR products for each allele, as

opposed to the presence or absence of a PCR product (dominant markers)) has led to popularity of the SSR markers. These markers assay the size of repetitive DNA at a locus (or loci) in different cultivars, usually in the form of tandem dinucleotide repeats, but where possible, the trinucleotide and tetranuclotide repeats due to the higher quality of the assay. The SSR assay requires the design of PCR primers that flank the repetitive loci, which can then be amplified and the size of the products compared between cultivars. Shah *et al.* (2000) found that out of 10 loci screened in the cultivars Cheyenne and Wichita, 6 were polymorphic (60%). Likewise, Parker *et al.* (2002) assayed 31 SSR loci in 101 wheat varieties and found 19 polymorphisms (61%), highlighting the efficiency of SSR-based marker systems at identifying polymorphisms. However, a major problem with SSR markers is that their development is laborious, involving the screening of genomic libraries with repetitive oligomers (e.g. AC or AG), isolation and analysis of genomic clones containing repeats, design of primers flanking the repetitive sequences and screening different cultivars with these primers for polymorphisms in the size of the repeats. The specificity of the primers flanking each SSR marker also means that they generally cannot be used in other species, as was demonstrated during attempts involving the use of wheat SSR primers in barley (Röder *et al.*, 1995).

Table 1.2 Comparison of PCR-based marker systems for wheat

Marker System	Loci detected per assay	Advantages	Disadvantages
SSR	1	Reliable, co-dominant, often genome-specific, target specific regions, amenable to automation	High development cost
AFLP	50	Reliable, detect large number of loci per assay, amenable to automation	Random, dominant
RAPD	10	Cheap to design and assay, technically simple	Unreliable, dominant

An outline of the advantages and disadvantages of the most commonly used PCR-based marker systems in wheat. Adapted from Langridge *et al.* (2001).

1.8.2.4 Gene-specific markers

The markers systems mentioned so far are based on sequence polymorphisms at random positions in the genome, which are extremely beneficial in the construction of genetic maps (Section 1.8.4); however, the functional significance, or tight linkage of these markers to specific alleles, needs to be established for their use in marker-assisted selection of cultivars for a given characteristic. Markers based on polymorphisms within genes ensure that the markers are 'perfect' for the gene to be assayed; however, the screening of genes for polymorphisms can be a laborious and expensive exercise. This problem has been partially addressed by analysis of the vast number of expressed sequence tags (ESTs) for wheat, that have become available over the past few years. The latest release of the dbEST division of GenBank (release# 021805; February 18, 2005) contained 587,650 ESTs from *T. aestivum*, more than any other plant species, which are derived from a number of cultivars and thus can be screened for inter-cultivar polymorphisms. Although valuable, the use of ESTs in this manner poses some problems, (1) the quality of ESTs is occasionally low and can lead to the false identification of polymorphisms; (2) the level of polymorphism within coding sequences is intrinsically low; and (3) distinguishing true inter-cultivar polymorphism of a single gene in one of the three genomes from the homeoalleles in the other two genomes can be a lengthy and tedious process. Despite these potential problems, ESTs as potential genetic markers have been reported, generally through the analysis of 5' and 3' untranslated regions, which generally have a higher level of polymorphism (Clarke *et al.*, 2001). The availability of ESTs also opens up the possibility of using PCR to screen genomic DNA (gDNA) for polymorphisms, which can lead to the identification of introns that tend to be more polymorphic than coding sequences, thus providing an additional source of 'perfect' markers.

1.8.3 Physical mapping of markers in wheat

Although the hexaploid nature of the *T. aestivum* genome poses a number of problems for the genetic analyses of genes and traits (Section 1.8.1) it also offers a number of advantages. One main advantage is the ability of bread wheat to tolerate the loss of a

chromosome pair, which can be compensated functionally by the remaining two homoeologous chromosome pairs (involving the duplication of one of the remaining two pairs). Sears (1954) exploited this feature in the development of nullisomic/tetrasomic (N/T) lines of wheat using the cultivar Chinese Spring, wherein individual pair of chromosomes were deleted and replaced by an additional pair of homoeologous chromosomes from either of the two genomes (Figure 1.10). Such N/T lines can be used to assign a marker to a particular chromosome; the absence of a marker in a N/T line indicates that the marker is present on the chromosome that is absent (nullisomic) in that particular line. The location can be further resolved by screening ditelosomic lines of wheat, wherein each line has a deletion of a particular chromosome arm and a duplication of the other arm. Further, development of deletion lines of Chinese Spring (Endo and Gill, 1996) that have partial deletions of specific chromosome arms has allowed even greater resolution of the physical maps of markers, as these lines allow the markers to be localised to a subsection of a chromosome arm, commonly referred to as a 'bin'. This ability, has allowed the comparison of physical and genetic maps, as demonstrated for the group 5 chromosomes of wheat (Gill *et al.*, 1996). The physical mapping of wheat EST sequences using N/T, ditelosomic and deletion stocks of wheat has been the aim of large scale efforts to develop physical maps of the expressed portion of the wheat genome (i.e., The US Wheat Genome Project; http://wheat.pw.usda.gov/NSF/htmlversion.html) and has led to the recent report of a genome-wide physical map of 16,000 expressed genes in wheat (Qi *et al.* 2004).

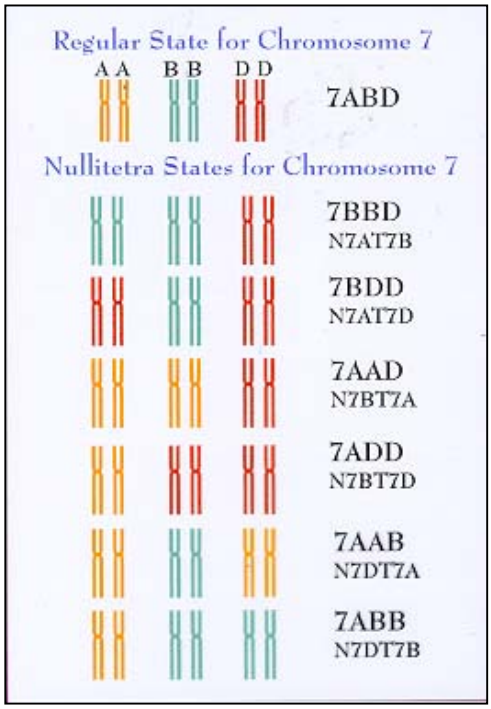


Figure 1.10 The N/T lines of group 7 chromosomes of *T. aestivum*

Two forms of nomenclature are shown, e.g. the line nullisomic for chromosome 7A and tetrasomic for 7B is shown as 7BBD or N7AT7B.

1.8.4 Genetic mapping

A genetic map provides the relative arrangement of genes or markers with regard to one anther and their relative genetic linkage to on another along each of the chromosomes of an organism. Genetic mapping relies on inheritance of parental chromosomes in a cross and subsequent meiotic recombination ('crossing over') in the gametes of the progeny of this cross. Genetic markers that are physically close to one another on the chromosomes are less likely to be recombined and are thus usually inherited together, while those that are further apart, or on separate chromosomes, are often 'recombined' and thus undergo independent assortment and are inherited in all possible combinations in the progeny (much like Mendelian crosses). The relative distance between any two markers can thus be determined by the extent of recombination between them (% recombination), or the percentage of progeny that inherited the two different markers from the two parents. This process can be repeated for many markers, allowing the determination of the relative distances between them and the construction of 'genetic maps', i.e., the linear order of markers on each chromosome. A genetic map can be considered complete when any new markers fall into established linkage groups (Figure 1.11) (Paterson *et al*. 1991).

The parents of a mapping population must have sufficient polymorphism, either at the phenotypic level (e.g. morphological and biochemical markers) and/or DNA level (molecular markers) to permit the tracing of this variation in recombination events. To reduce the complexity of analysing the progeny of crosses involving such polymorphic parents (mapping populations), the parents are generally highly inbred so as to be homozygous themselves at as many loci as possible (Young, 1994; Paterson, 1996; Liu, 1998). The simplest mapping populations are F_2 progeny derived by selfing of the F_1 hybrid. Each of the F_2 progeny contains recombinant chromosomes from the parents and thus all possible combinations of the parental alleles (i.e., AA, Aa, aa). Another simple mapping population is a backcross, derived by crossing the F_1 hybrid to one of the parents, producing a single recombinant chromosome from each homologous pair. Yet another type of mapping population consists of doubled haploid (DH) progeny that have undergone artificial chromosome doubling from individual pollen grains. Although much more difficult to produce than F_2 or backcross populations, individuals

in DH populations have the advantage of being homozygous at all loci, reducing the complexity of analyses and permitting the transfer of DH populations between different laboratories and environments (Young, 1994). Recombinant inbred lines (RILs) represent another type of mapping population, derived by inbreeding individual F_2 progeny, producing homozygous populations (Young, 1994; Paterson, 1996).

Although the assembly of genetic maps can be time- and labour-intensive, there are a number of benefits that can be derived from their construction, including, (1) allowing the localisation of genes of interest; (2) facilitating marker-assisted breeding and map-based cloning; and (3) providing the framework for understanding the biological basis of complex traits (Vuylsteke *et al.* 1999).

1.8.4.1 Genetic maps in wheat

A number of genetic maps have been assembled for different wheat crosses (Chalmers *et al.*, 2001; Kammholz *et al.*, 2001; Langridge *et al.*, 2001). The most comprehensive single population wheat map was assembled from W7984 x Opata 85 cross, W7984 being a synthetic hexaploid, containing a tetraploid (AABB) and diploid (DD) genome and Opata 85 being a cultivar of *T. aestivum* selected to increase the chance of polymorphism in the cross. This map is widely regarded as the international reference mapping population for wheat and, in 2001, contained 1074 loci, primarily consisting of RFLP and SSR markers (Figure 1.12) (Langridge *et al.*, 2001). This map has good genome coverage, however, as mentioned previously, the D genome contains fewer markers than the A or B genomes, reflecting a lower level of polymorphism in it. In Australia, five genetic maps are being constructed using the crosses Cranbrook x Halberd, CD87 x Katepwa, Sunco x Tasman, Egret x Sunstar and Kukri x Janz, based on the contrasting traits of the parental cultivars (Table 1.3) to allow dissection of the genetic basis of these traits through QTL analysis (Section 1.8.5). When published, the maps assembled from these crosses contained 902 (Cranbrook x Halberd), 505 (CD87 x Katepwa) and 355 (Sunco x Tasman) mapped markers consisting of RFLP, AFLP, SSR, gene-specific and protein markers. The coverage of these maps varies, with the D genome having the poorest coverage (Chalmers *et al.*, 2001).

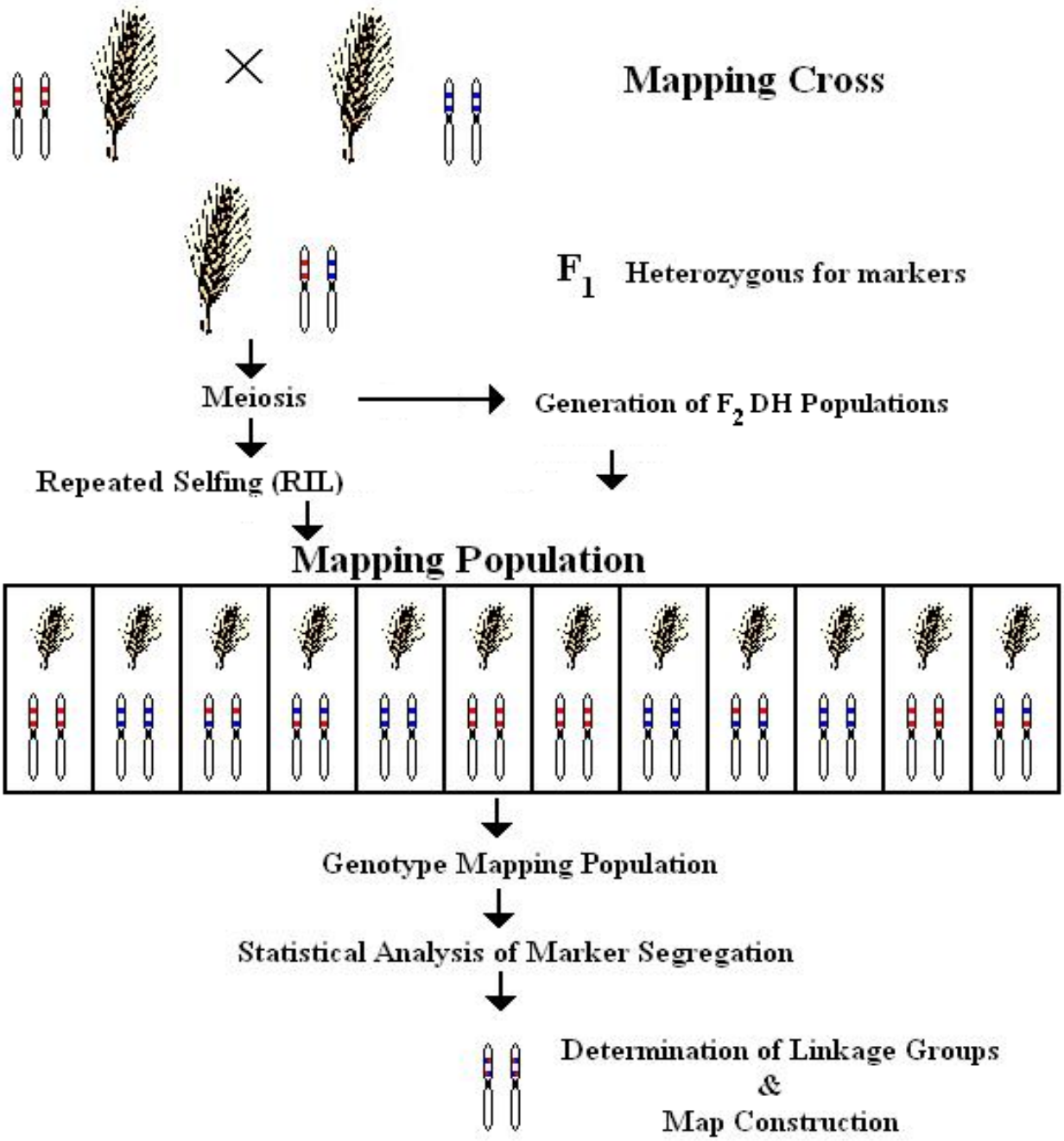


Figure 1.11 Basic steps in genetic map construction.

A basic depiction of the steps involved in genetic map construction highlighting two genetically linked polymorphic loci in the mapping parents. Hundreds or thousands of such markers are genotyped to provide marker coverage over the whole genome. Only the homozygous RIL and DH populations are depicted, however other mapping populations are used (i.e., F_2 and backcross) RIL – Recombinant Inbred Lines, DH – Doubled Haploid.

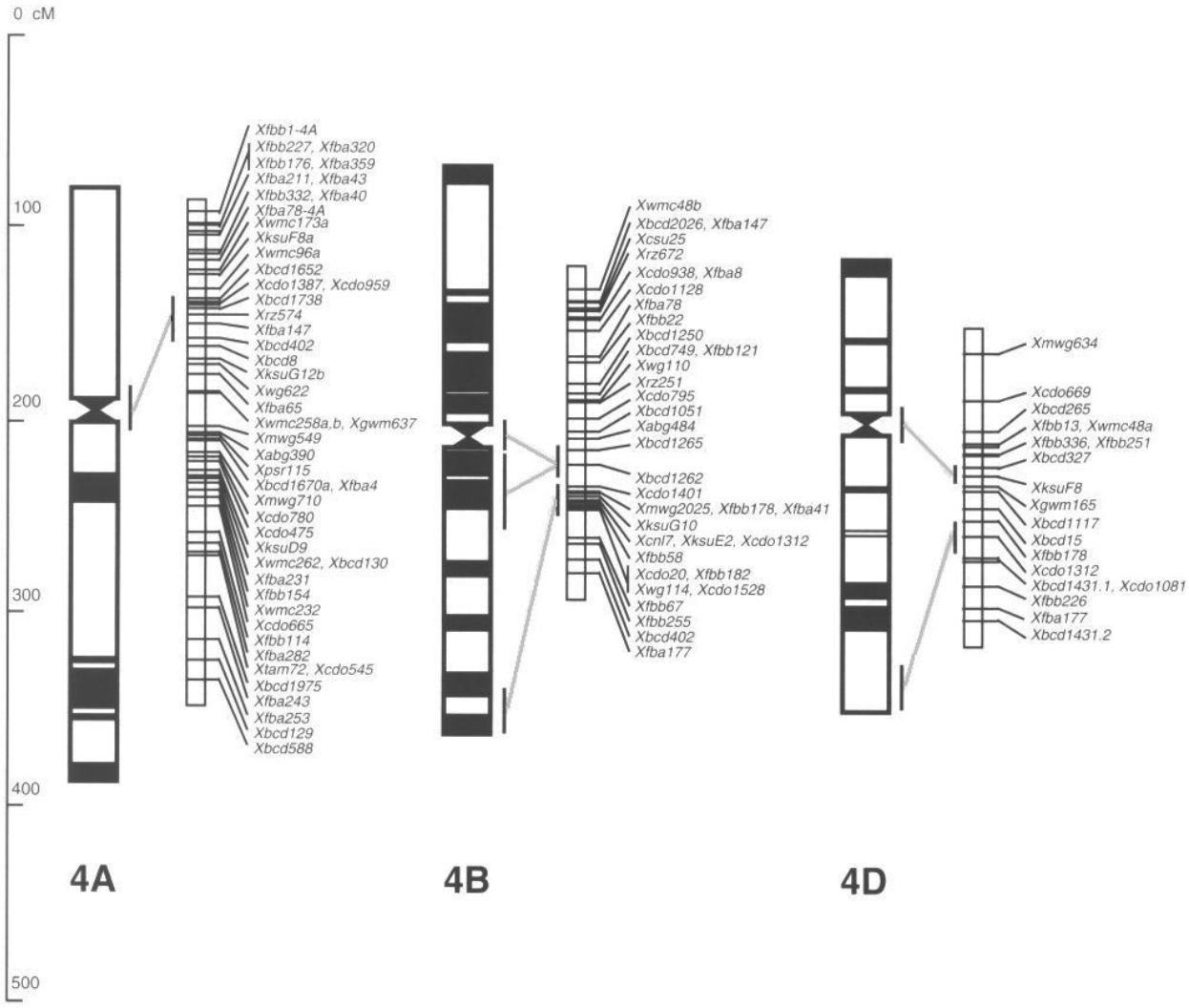


Figure 1.12 Genetic map of the group 4 chromosomes of Synthetic x Opata

A comparison of the physical and genetic maps is shown, with light grey lines highlighting the 'bins' that some markers map to. Note the presence of fewer markers mapped to chromosome 4D, due to the lower level of polymorphism observed in the D genome. Scale for genetic distance (cM) is shown on left (figure from GrainGenes website; http://wheat.pw.usda.gov).

Table 1.3 Segregating traits in the NWMMP mapping crosses

Cross	Segregating Traits
Cranbrook x Halberd	• Flour mixing characteristics • Starch pasting characteristics • Grain protein • Dough strength • Extensibility
CD87 x Katepwa	• Dough extensibility
Sunco x Tasman	• Flour colour • Flour water absorption • Extensibility • Dwarfing genes • Glume colour • Rust genes • Black point resistance • Grain shattering
Egret x Sunstar	• Dough strength and extensibility in a genetic background with similar storage protein alleles
Kukri x Janz	• Dough strength • Leaf rust resistance • Stem rust resistance • Stripe rust resistance • Yellow spot resitance • Flour water absorption

Details of the five crosses established as genetic mapping populations as part of the Australian National Wheat Molecular Marker Program (NWMMP) (Kammholz *et al.*, 2001).

1.8.5 QTL analysis

As mentioned above, one of the main benefits of genetic map construction is the dissection of complex traits. Some genetic traits demonstrate a simple 'plus or minus' (e.g. resistance to a pathogen) or alternative (e.g. black or white) phenotype and are generally under the control of alternative alleles of a single gene with markers linked to these genes generally segregating perfectly with the traits concerned in a mapping population. However, the majority of traits observed in wheat tend to show a linear

gradient between two extremes of phenotype, e.g. plant height. Such 'quantitative' traits tend to arise due to either (i) the cumulative effect of many alleles, each of a small influence on the final phenotypes and/or, (ii) a trait significantly influenced by environmental factors, the ultimate phenotype determined through a combination of genetic and environmental effects. The development of molecular markers and genetic maps has provided a powerful method for dissecting the genetic components of such quantitative traits, through the identification of 'quantitative trait loci' (QTL).

The identification of QTL is a much more complex and lengthy process than the identification of a locus affecting simple traits. Most QTL analyses require the construction of a genetic map in a cross that the trait of interest is segregating in. The selection of the mapping populations used as part of the National Wheat Molecular Marker Program (NWMMP) was based on the segregation of agronomically important quantitative traits in these populations (Table 1.3) (Kammholz *et al.*, 2001). The basic method of identifying QTL involves quantifying the trait in the individual members of the mapping population, followed by analyses to identify any significant linkage between the trait and the markers present on the genetic map, using statistical tools such as those reviewed in Liu (1998) and Eckermann *et al.* (2001). However, the quality and density of the markers on the genetic map are central in determining the resolution of the QTL, hence the importance of many markers covering the entire genome. As mentioned above, many quantitative traits are strongly affected by various environmental factors and therefore the phenotype is usually measured after growing the mapping population under varied environmental conditions, so as to separate the effects of the genetic component from these environmental effects.

QTL analysis has a number of benefits in gaining insights into both gene function and genetic control of quantitative traits:

- The putative role of a candidate gene can be strengthened through identifying linkage of that gene with its putative phenotype e.g. mapping of the storage proteins and their genetic association with processing traits of dough has been demonstrated (Appels *et al.*, 2001).
- Segregation of a gene-specific marker with a QTL may reveal a previously unknown role for that gene.

- QTL analyses can often identify a locus containing no mapped genes and its subsequent characterisation may identify new genes potentially involved in the trait.
- QTL analyses can also elucidate metabolic pathways and their regulation, as demonstrated in identifying the regulatory and structural genes involved in the corn earworm resistance pathway in maize (McMullen *et al*. 1998) and boron tolerance in wheat (Jefferies *et al.* 2000).

1.8.5.1 QTLs identified in wheat

The benefits of identifying QTLs to breeding programs and to elucidation of pathways affecting complex traits has led to a number of reports of QTL analysis in wheat, including, quality traits, physiological traits, disease resistance and abiotic stress tolerance (Table 1.4).

Table 1.4 A selection of QTL studies in wheat

Trait	Reference(s)
Pre-harvest sprouting	Zanetti *et al.* 2000; Mares and Mrva, 2001; Mares *et al.* 2002
Grain size	Mares and Campbell, 2001
Grain hardness	Osborne *et al*. 2001
Kernel morphology and texture	Campbell *et al.* 1999
Milling yield	Parker *et al.* 1999
Flour colour	Parker *et al.* 1998; Mares and Campbell, 2001
Starch characteristics	Batey *et al*. 2001
Bread making quality	Perretant *et al.* 2000; Zanetti *et al.* 2001
Ear compactness	Sourdille *et al.* 2000b
Free threshing habits	Simonetti *et al.* 1999
Lodging resistance	Keller *et al.* 1999b
Crossability	Tixier *et al.* 1998
Heading time	Sourdille *et al.* 2000a
Plant height	Batey *et al*. 2001; Mares and Campbell, 2001; Rebetzke *et al.* 2001
Fusarium head blight resistance	Waldron *et al.* 1999; Bai *et al.* 1999
Powdery mildew resistance	Keller *et al.* 1999a
Resistance to chlorosis	Effertz *et al.* 1998
Boron tolerance	Jefferies *et al.* 2000

1.9 ORTHOLOGY BETWEEN GRASS GENOMES

1.9.1 Evidence supporting orthology between grass species

The development of RFLP-based genetic maps during the late-1980's and early 1990's revealed that some probes were able to cross-hybridise between different species (Bonierbale *et al*. 1988; Ahn and Tanksley, 1993; Causse *et al.* 1994). Analysis of these genetic maps revealed a quite unexpected phenomenon, of large sections of 'marker colinearity', i.e., the same linear order of RFLP markers on a chromosome of different species that had evolved separately for millions of years, such as rice and maize, or tomato and potato (Bonierbale *et al*. 1988; Ahn and Tanksley, 1993). The field of comparative genetics in plants, and grasses in particular, has been the subject of a number of reviews (Gale and Devos, 1998; Keller and Feuillet, 2000; Paterson *et al.*, 2000); this section will give a brief overview of investigating orthology and gene colinearity in grasses.

Comparative genetic maps were constructed to determine the extent of orthology between species using RFLP 'anchor' probes, and revealed a striking level of colinearity of molecular markers across seven grass species, with some clear exceptions where genome rearrangements, such as inversions, translocations and deletions, probably responsible for at least part of the speciation process, had taken place (Ahn and Tanksley, 1993; Kurata *et al.* 1994; Wilson *et al.*, 1999). The large sections of orthologous regions led to the establishment of 25 rice linkage blocks, representing the entire rice genome, that exhibit colinearity amongst these seven grass species, with some small areas of uncertainty (Figure 1.13) (Moore *et al.* 1995; Gale and Devos, 1998). More recently, research efforts have aimed to assess colinearity at the sub-megabase or gene level, also referred to as microcolinearity, often using large insert genomic clones such as bacterial artificial chromosomes (BACs) (Chen *et al.* 1997; Tikhonov *et al.* 1999). Chen *et al.* (1997) sequenced genomic fragments from two orthologous loci of rice, maize and sorghum revealing good conservation of gene order at the *Sh2/a1* locus, although the size of intergenic spaces varied between species. Analysis of BACs containing genomic inserts encompassing the orthologous *Adh1* loci from maize and sorghum revealed an exception to colinearity with the presence of three additional genes in sorghum (Tikhonov *et al.*, 1999). Analysis of the *Rpg1* locus in

barley and it orthologous region in rice (Kilian *et al.*, 1997) also revealed that 10-15 kb of DNA flanking the *Rpg1* locus in rice was present 2.5 cM proximal to the orthologous locus in barley, in a non-orthologous region, suggesting that micro-rearrangements e.g. deletions and gene duplications, can interrupt gene co-linearity. These observations have been supported in more recent comparative genomic studies on grass species (Bennetzen and Ramakrishna 2002; Ramakrishna *et al.* 2002; Bennetzen and Ma, 2003) and in a large-scale comparison between the rice genome sequence and the chromosomal 'bin' positions of wheat genes (Section 1.8.3) (LaRota and Sorrells, 2004).

1.9.2 Benefits of investigations into gene orthology

The findings of colinearity and limited microcolinearity of markers have allowed development of new strategies to investigate genes and/or traits in species with large, complex genomes, using a species with a simpler genome as a model organism. The publications of the draft sequences of the genomes of two subspecies of rice (Goff *et al.*, 2002; Yu *et al.*, 2002) has provided the framework for a grass species with a relatively small genome (4.3×10^8 bp), approximately 40 times smaller than wheat (Arumuganathan and Earle, 1991). Investigations into the gene orthology have the potential to allow the isolation of genes of agronomic importance based on their locus in another, related species. Another exciting prospect is that QTLs for important traits such as vernalisation, flowering time, plant height, dwarfism and grain shattering, tend to have an orthologous relationship between barley, wheat, rice and maize (Paterson *et al.*, 1995; Pereira and Lee, 1995; Sarma *et al.*, 1998; Bailey *et al.*, 1999). This phenomenon has significant implications; to quote Gale and Devos (1998), "Thus, many of the major gene mutants mapped in barley, maize, and rice may be used as pointers to homoeogenes with more subtle, exploitable effects in the same or other genomes".

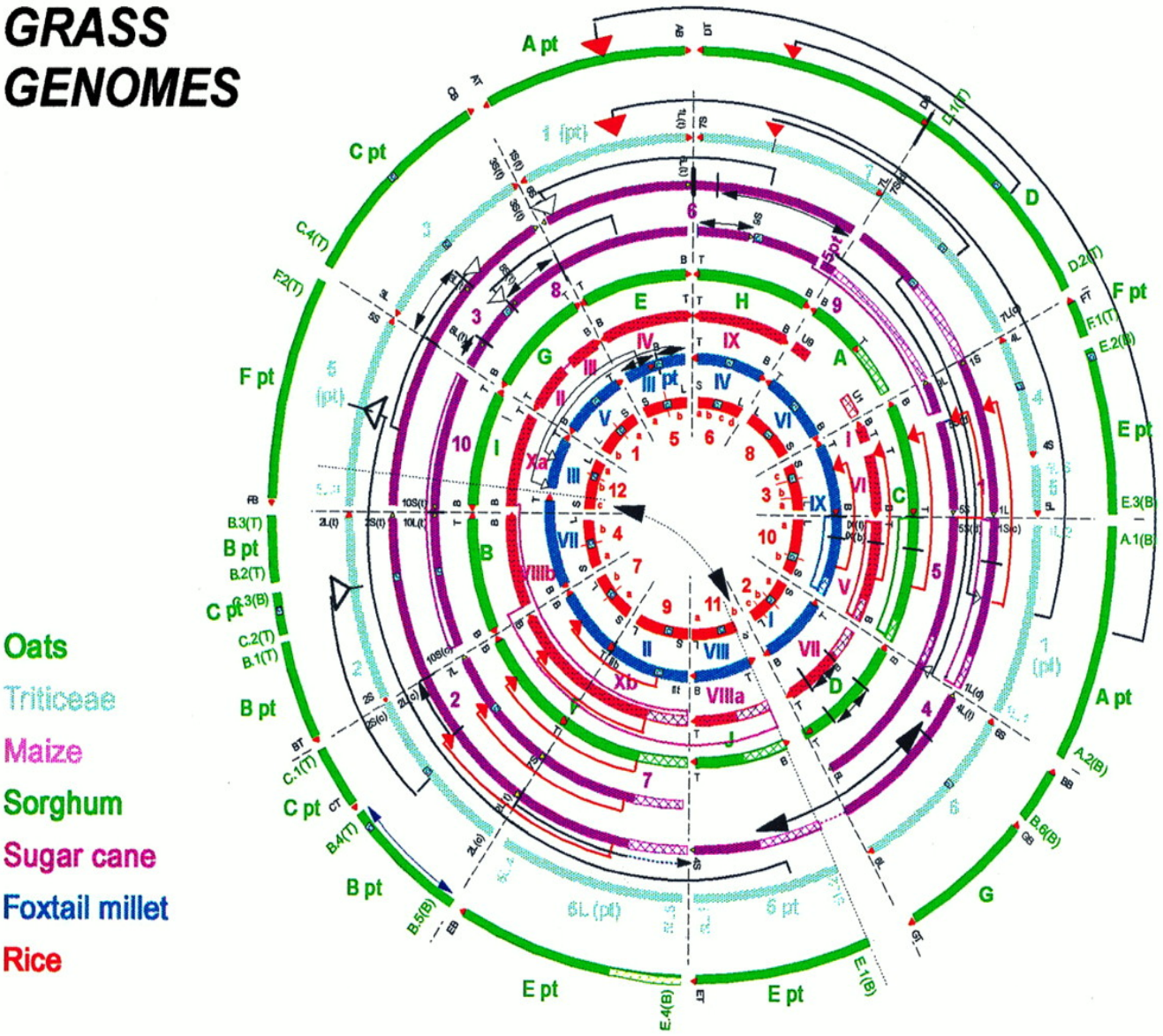


Figure 1.13 A consensus grass comparative map

The relationship between the chromosomes of oats, a wheat consensus (Triticeae), maize, sorghum, sugar cane, foxtail millet and the 25 linkage blocks on the 12 chromosomes of rice. Arrows indicate inversions and transpositions, required to display the present day chromosomes. The locations of known telomeres (△) and centromeres (□) are shown. Hatched areas are unclear due to lack of comparative data. Chromosome numbers for each species are shown (Gale and Devos, 1998).

1.10 AIMS OF THE PROJECT

As discussed earlier there is clear evidence to suggest that the foldase enzymes PDI and cyclophilin may play important roles during the development of the wheat endosperm (in the association and deposition of the storage proteins). However, to date little is known about the genes encoding these enzymes in wheat. The isolation and molecular characterisation of gene families encoding these enzymes, with an aim to develop 'perfect' molecular markers for these genes, will permit their genetic mapping and allow any association with quantitative traits to be recognized through QTL analysis.

1.10.1 Major aim

To investigate the organisation, gene copy number and sequence diversity of the *PDI* and *cyclophilin* gene families in wheat, to develop molecular markers for these genes, where possible, to permit their integration into genetic maps, and to assess their linkage with known QTLs as well as investigate orthology with loci of potential agronomic importance in rice.

1.10.2 Specific aims

The specific aims of this project were:

- To clone and sequence the *PDI* genes from the D genome of *Ae. Tauschii* and the A and B genomes of *T. turgidum*.
- To use the sequences of the *PDI* genes characterised from *Ae. tauschii* and *T. turgidum* to isolate and characterise the corresponding genes from the A, B and D genomes of *T. aestivum*.
- To sequence the *PDI* genes of eight cultivars of *T. aestivum* used in the establishment of genetic mapping populations, identify any polymorphism(s) and then use them (if found), to develop 'perfect' *PDI* molecular markers.
- To integrate any *PDI* molecular markers into the genetic maps that they are found to be polymorphic in and assess their map position for QTLs previously mapped to this locus (or loci).

- To use gene-specific sequences to physically map the *PDI* genes of *T. aestivum* and use this data in conjunction with publicly available wheat and rice genomic data to investigate orthology between the *esp2* locus of rice and the *PDI* loci of wheat.
- To amplify, clone and sequence the cyclophilin genes from the eight cultivars of *T. aestivum* previously used in the establishment of genetic mapping populations to assess these genes for the presence of any inter-cultivar polymorphism.
- To use gene-specific sequences to physically map the cyclophilin genes of *T. aestivum*.
- To screen a cDNA library database and publicly available EST data for novel cyclophilin-like sequences.

The achievement of these aims will provide a strong basis for elucidating a role for both of these gene families and their association with quantitative traits in wheat. The following chapters describe the methods used to address these aims and the results obtained during the course of this investigation.

CHAPTER 2

MATERIALS AND METHODS

List of Figures and Tables in Chapter 2

Chapter 2 Materials and Methods

MATERIALS

2.1 CHEMICALS, REAGENTS AND KITS

The materials used in the course of this research are listed below. Full details on the preparation and sterilisation (where appropriate) of these reagents are provided in Appendix I.

2.1.1 Commercial kits

A number of commercially available kits were used during the course of this research. The names of the kits and their suppliers are summarised in Table 1.1.

Table 2.1 Commercial kits used during this project

Name of Kit	Supplier	Purpose
Plant DNAzol	Invitrogen	Preparation of gDNA (Section 2.2.2)
Platinum Taq Polymerase	Invitrogen	"Hot-Start" PCR (Section 2.3.2)
HotStarTaq DNA Polymerase	Qiagen	"Hot Start" PCR (Section 2.3.2)
Concert Rapid Gel Extraction Kit	Invitrogen	Purification of DNA from agarose gels (Section 2.4.1)
Concert Rapid PCR Purification Kit	Invitrogen	Purification of PCR products (Section 2.4.2)
ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit v.3	Applied Biosystems	DNA Sequencing (Section 2.6.2)
Prime-a-Gene Labeling System	Promega	Radiolabeling of probes (Section 2.7.3.2)
Ready-To-Go DNA Labeling Beads	Pharmacia	Radiolabeling of probes (Sections 2.7.3.2)
pGEM-T-Easy Cloning Kit	Promega	Cloning of PCR products (Section 2.9.2)
Wizard *Plus* SV Minipreps DNA Purification System	Promega	Plasmid purification (Section 2.9.3)

2.1.2 Enzymes

A number of enzymes were used in this project, many of which were supplied in the kits used for PCR, cloning and sequencing reactions. These enzymes included:

- Platinum *Taq* polymerase (Invitrogen) and HotStarTaq (Qiagen);
- T4 DNA Ligase (Promega);
- Various restriction endonucleases (Promega, MBI Fermentas and New England Biolabs);
- The Klenow fragment of *E. coli* DNA polymerase I (Promega and Pharmacia);
- Alkaline Protease solution (Promega);
- RNase A (10μg/μl) (Progen).

2.1.3 Buffers

A number of buffers were supplied with kits or enzymes for use in PCR reactions, DNA purifications, restriction digestions, ligation and radiolabelling reactions. In addition, the following buffers were prepared:

- TE Buffer;
- TAE electrophoresis buffer;
- SM Buffer;
- BigDyeSequencing dilution buffer (1.4mM $MgCl_2$, 60mM Tris-Cl (pH 7.5), 3M sodium acetate, pH 5.2);
- SSC Buffer;
- Denaturation solution (1.5M NaCl, 0.5M NaOH);
- Neutralisation solution (1.5M NaCl, 0.5M Tris-Cl (pH 7.2));
- Prehybridisation Buffer I (0.2% Ficoll, 0.2% bovine serum albumen (BSA), 0.2% polyvinyl pyrrolidone (PVP), 1% SDS, 1.2% HEPES, 3x SSC, 0.002% sheared herring sperm DNA (Sigma)).

2.1.4 General solutions and stocks

- 20% Maltose (Aldrich);
- X-Gal (20mg/mL) (Progen);

- 20% IPTG (Progen);
- Ethidium Bromide (10mg/mL) (Sigma);
- 10% SDS (electrophoresis grade, BioRad);
- 1M $MgSO_4$ (BDH);
- 2M Mg^{2+} stock (BDH);
- 0.5M EDTA (Ajax Chemicals);
- 3M Sodium Acetate (Ajax Chemicals);
- Ampicillin (50mg/mL) (Sigma).

2.1.5 Microbiological media

- LB broth (Luria-Bertani broth);
- SOC media.

2.2 PLANT MATERIAL AND GENOMIC DNA

Plant material from hexaploid, tetraploid and diploid wheat was used to investigate the organization of the *cyclophilin* and *PDI* gene families. The sources of this plant material are outlined in Table 2.1. This material included:

1. *Ae. tauschii*, the proposed donor of the D genome to common wheat;
2. *T. turgidum* ssp. *durum*, the putative tetraploid donor of the A and B genomes to common wheat;
3. Common wheat, *T. aestivum* (AABBDD). A number of different cultivars were investigated, including the experimental wheat variety Chinese Spring, the biscuit wheat variety used previously to construct a cDNA library from developing endosperm tissue, Wyuna (Clarke *et al.*, 2000); and seven cultivars used previously to establish genetic maps in common wheat, Cranbrook, Halberd, Egret, Sunstar, Sunco, Tasman and Katepwa (Section 1.8.4.1).

2.2.1 Growth of seedlings

Seeds were grown under controlled conditions with 16-hour day lengths at a constant temperature of 18°C. Two-week old seedlings were collected and snap-frozen in liquid nitrogen and stored at -80°C for future gDNA extractions.

2.2.2 Preparation of genomic DNA

Genomic DNA (gDNA) was extracted from plant material using the Plant DNAzol reagent (Invitrogen) according to the supplied protocol. This involved the grinding of frozen leaf tissue under liquid nitrogen to obtain a fine powder. The cells were then lysed and the RNA hydrolysed by mixing with Plant DNAzol containing RNase A (100µg/mL of Plant DNAzol), followed by extraction with chloroform and centrifugation to separate the aqueous and organic phases. The gDNA was then precipitated from the aqueous phase with ethanol, the DNA pellet washed with DNAzol/ethanol (1:0.75) and a final wash with 70% ethanol. The gDNA pellet was allowed to dry briefly at room temperature and resuspended in dH_2O overnight at 4°C. Insoluble material was removed from the DNA solution by centrifugation as described in the supplied protocol. The integrity of the purified gDNA was visually assessed to confirm the absence of RNA and gDNA degradation by agarose gel electrophoresis and the concentration was determined by UV spectrophotometry (Section 2.15.1).

2.2.3 DNA samples from other sources

A number of gDNA samples used in this project were the generous gift of Lynette Rampling of CSIRO – Plant Industry, Canberra, Australia. These samples included:

1. gDNA extracted from common wheat, *T. aestivum* (AABBDD); eight cultivars used previously to establish genetic maps in common wheat, Cranbrook, Halberd, Egret, Sunstar, Sunco, Tasman, CD87 and Katepwa (Section 1.8.4).
2. gDNA extracted from aneuploid lines of *T. aestivum* cv. Chinese Spring including, 42 nullisomic/tetrasomic lines, 6 ditelosomic lines and 26 deletion lines (Table 2.2).

3. gDNA extracted from doubled haploid (DH) mapping populations derived from crosses of Cranbrook x Halberd (160 DH lines), Sunco x Tasman (180 DH lines) and CD87 x Katepwa (180 DH lines) (Section 1.8.4).

Table 2.2 Sources of gDNA used in this study

Species	Variety	Ploidy of genomes	Genome	Source
Ae. tauschii	Aus# 21712	2n	DD	AWCC
T. turgidum ssp. *durum*	Aus# 11438	4n	AABB	AWCC
T. aestivum	Chinese Spring, Wyuna, Cranbrook, Halberd, Egret, Sunstar, Sunco, Tasman, CD87[a], Katepwa	6n	AABBDD	AWCC and CSIRO[b]
T. aestivum (aneuploids)	42 nullisomic/tetrasomic lines, 6 ditelosomic lines and 26 deletion stocks derived from variety Chinese Spring[c]	6n	AABBDD	CSIRO[b]
T. aestivum (doubled haploid[d])	160 DH lines generated from Cranbrook x Halberd; 180 DH lines generated from Sunco x Tasman; 180 DH lines generated from CD87 x Katepwa	6n (DH)	AABBDD	CSIRO[b]

AWCC: Australia Winter Cereals Collection, Tamworth, NSW. CSIRO: Commonwealth Scientific and Industrial Research Organization, Plant Industry, Canberra, ACT, Australia.

[a] CD87 was only received from CSIRO, not from AWCC.

[b] Plant material obtained from L. Rampling of CSIRO – Plant Industry was received in the form of purified gDNA.

[c] Refer to Table 2.2 for details of these lines.

[d] The doubled haploid (DH) mapping populations were generated as described in Kammholz *et al.* (2001) (Section 1.8.4.1).

Table 2.3 The aneuploid lines of *T. aestivum* used in this study.

Type of aneuploid stock	Aneuploid stock used
Nullisomic/Tetrasomic Stocks	1BBD, 1BDD, 2BBD, 2BDD, 3BBD, 3BDD, 4BBD, 4BDD, 5BBD, 5BDD, 6BBD, 6BDD, 7BBD, 7BDD. 1AAD, 1ADD, 2AAD, 2ADD, 3AAD, 3ADD, 4AAD, 4ADD, 5AAD, 5ADD, 6AAD, 6ADD, 7AAD, 7ADD. 1AAB, 1ABB, 2AAB, 2ABB, 3AAB, 3ABB, 4AAB, 4ABB, 5AAB, 5ABB, 6AAB, 6ABB, 7AAB, 7ABB.
Ditelosomic Stocks	DT4AL, DT4BS, DT4DL. DT6AS, DT6BS, DT6DS.
Deletion Stocks[a]	4AS1, 4AS2, 4AS4, 4AL4, 4AL5, 4AL13, 4BS1, 4BS4, 4BS8, 4BL1, 4BL5, 4DS1, 4DS3, 4DL9, 4DL13. 6AS1, 6AL8, 6BS2, 6BS3, 6BS5, 6BL5, 6BL6, 6DS6, 6DS4, 6DS2, 6DL6.

[a] Positions of breakpoints in these deletion stocks are highlighted in Figure 2.1.

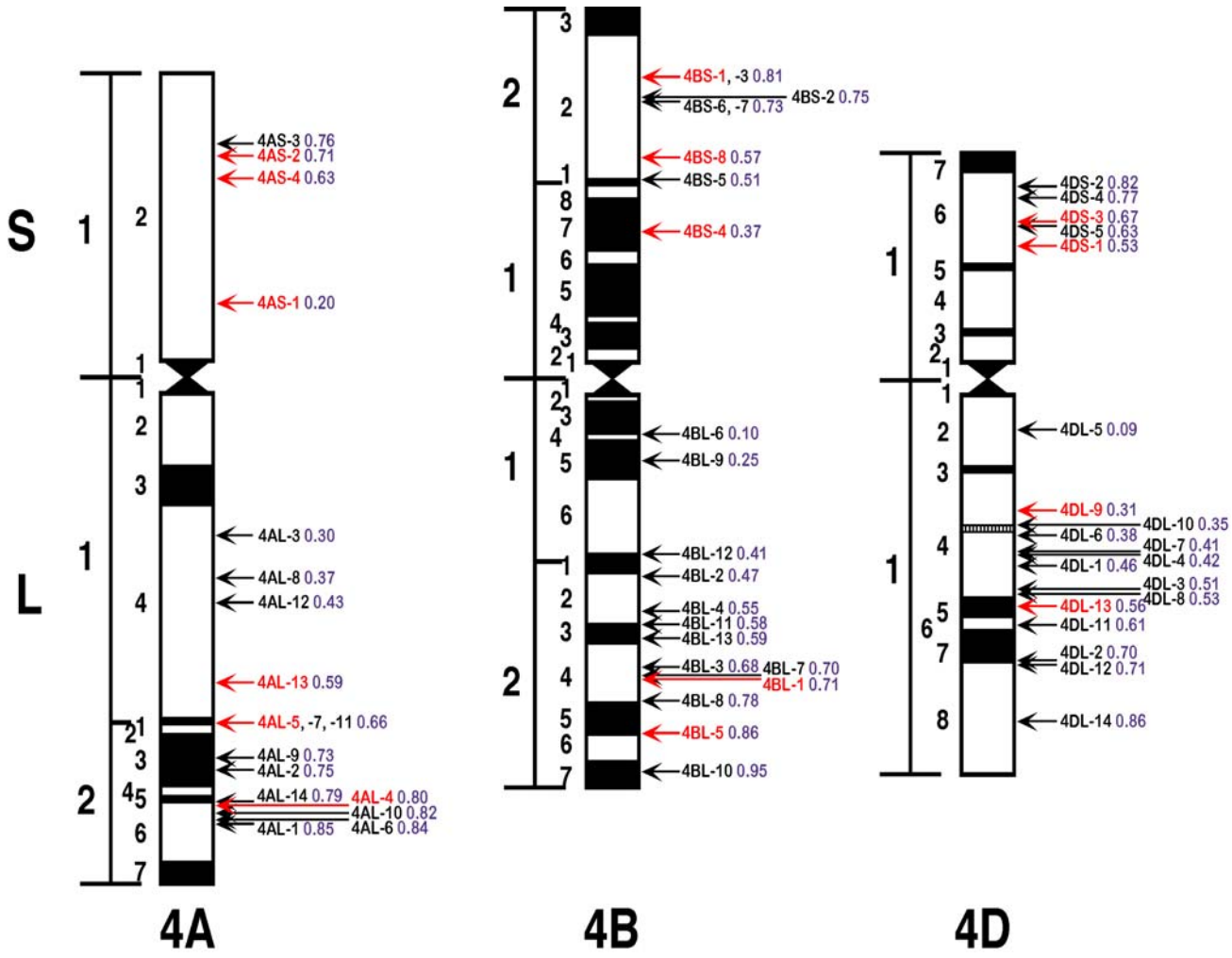


Figure 2.1 Continued.

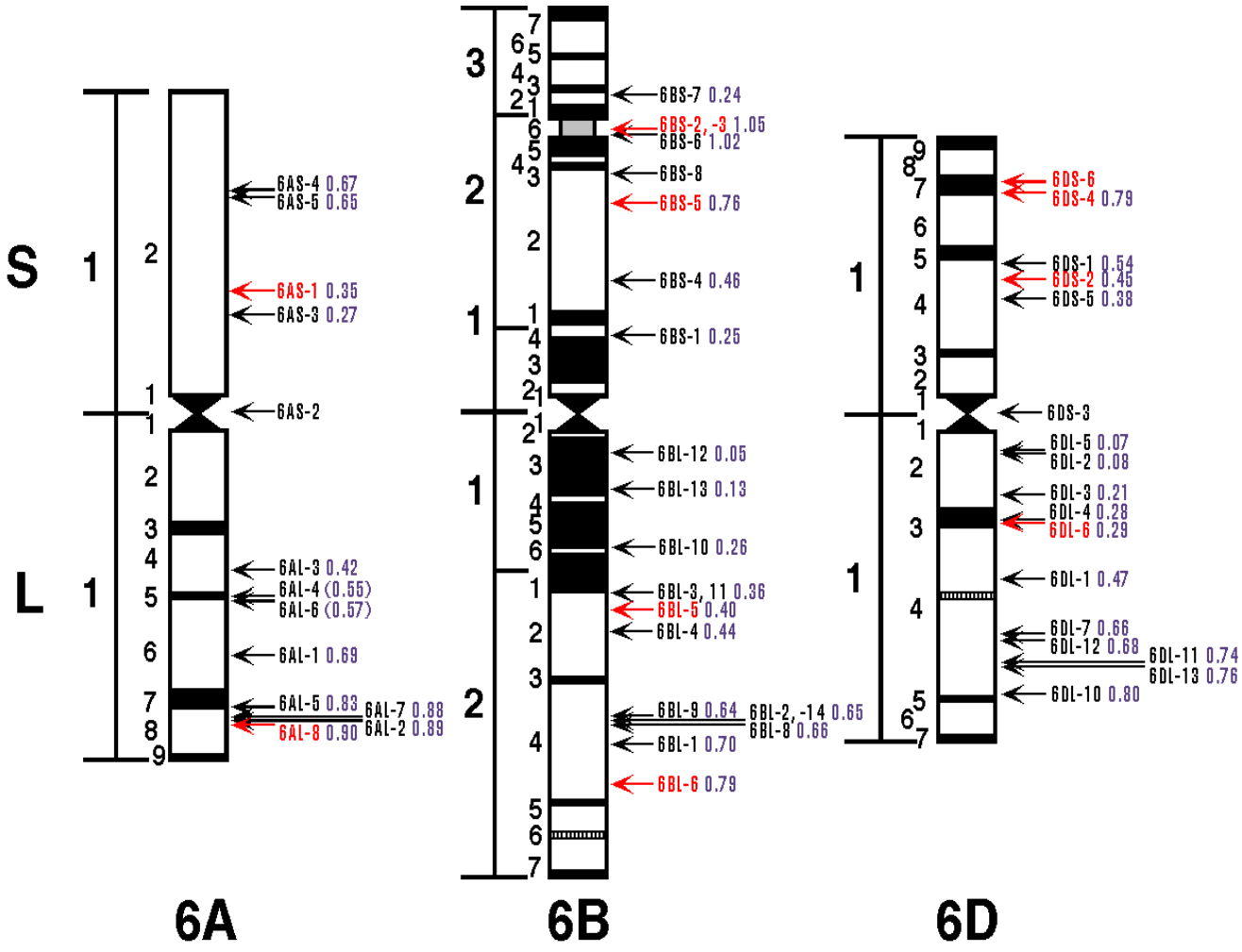


Figure 2.1 The deletion stocks of *T. aestivum* used in this study.

The breakpoints of the deletions used in this work are highlighted in red. (Original figures from The Wheat Genetics Resource Page: http://www.k-state.edu/wgrc/).

GENERAL MOLECULAR METHODS

2.3 THE POLYMERASE CHAIN REACTION

2.3.1 Synthesis of oligonucleotide primers

All primers were synthesised commercially by Sigma Genosys with cartridge purification. Primers were received as a dried precipitate and resuspended in ddH_2O to a final stock concentration of 1μg/μL before storage at -20°C.

2.3.2 Typical PCR conditions

Following the principles of the Polymerase Chain Reaction (PCR) (Mullis and Faloona, 1987), this technique was carried out typically in 25μL reactions containing 1x PCR Buffer, 0.2mM dNTPs, 1.5mM $MgCl_2$, 0.5uM primer (each), 2U *Taq* Polymerase and 50ng template DNA. PCR reactions were prepared in 10-50μL final volumes, the template DNA being kept at 50ng in all reactions. "Hot start" PCR was used routinely through the use of the Platinum Taq Polymerase (Invitrogen) or HotStarTaq (Qiagen). These polymerases are inactive at room temperature and only become active during the high temperatures (95°C) of initial denaturation when the antibody blocking their active site is denatured (Platinum Taq) or they are chemically modified (HotStarTaq). This "hot start" is suggested to reduce the production of primer-dimers and other non-specific priming events that can interfere with the amplification of the section of DNA targeted in the design of the experiment (Qiagen and Invitrogen product literature). This was especially important during the allele-specific amplification of each of the *PDI* and *cyclophilin* genes, where the primers had to discriminate between single polymorphic nucleotides at each priming site (Sections 2.10.1 and 2.16.1). Thermal cycling was performed in the PTC-100 or PTC-200 Peltier Thermal Cycler (MJ Research) or PCR Sprint thermal cycler (ThermoHybaid) with cycling conditions generally consisting of an initial denaturation at 95°C for 5 minutes (10 minutes for HotStarTaq), followed by 35 cycles of denaturation (94°C) for 30 seconds, annealing at primer specific temperatures (Tables 2.5, 2.6 and 2.7) for 30 seconds and elongation (72°C) for 2 minutes, then a final elongation step at 72°C for 7 minutes. Negative controls, containing no gDNA, were used in all experiments. PCR was carried out in a in a number of areas of this study, including the amplification of the *PDI* genes

(Sections 2.9.1, 2.10.3, 2.13 and 2.14), the *cyclophilin* genes (Sections 2.16.3, 2.17, 2.18.3 and 2.19.1) and production of probes (Sections 2.8.2, 2.11.2 and 2.18.2). The specifics and any variations to the above-described typical conditions are provided in the appropriate sections.

2.4 AGAROSE GEL ELECTROPHORESIS AND PCR PURIFICATIONS

The "Concert" line of products (Invitrogen) was initially used for both agarose gel and PCR purifications. These products were discontinued by the company during this project necessitating an alternate product. As such, the Wizard SV Gel and PCR Clean-Up System (Promega) was used for both PCR and agarose gel purifications during the later stages of this project.

2.4.1 Concert Rapid Gel Extraction Kit (Invitrogen)

This kit was used for the purification of DNA from agarose gels. Briefly, this involved:

- 30µL of 'Gel Solubilization Buffer' was added for every 10mg of agarose gel followed by incubation at 50°C for 15 minutes, with occasional vortexing.
- The eluted DNA was then bound to the supplied column by centrifugation at 12,000g for one minute.
- The column was washed with 500µl of 'Gel Solubilization Buffer' and centrifuged at 12,000g for one minute to remove any residual agarose.
- The column was then washed with 700µl of 'Wash Buffer' and centrifuged at 12,000g for one minute. The 'Wash Buffer' was discarded and the column was centrifuged again at 12,000g for one minute.
- The bound DNA was then eluted from the column by the addition of 50µl of TE buffer at 65°C, incubation at room temperature for 1 minute and centrifugation at 12,000g for 2 minutes.
- Purified DNA was used as a template for PCR as described below (Section 2.13)

2.4.2 Concert Rapid PCR Purification Kit (Invitrogen)

Where required, PCR reactions were purified with the "Concert Rapid PCR Purification System" (Invitrogen). Briefly, this method involved:

- The PCR mix, after amplification was combined with 400µl of 'Binding Solution' and bound to the supplied column by centrifugation at 12,000g for 1 minute.
- The column was washed with 700µl of 'Wash Buffer' and centrifuged at 12,000g for 1 minute, the wash solution discarded and the column was centrifuged again at 12,000g for 1 minute to remove any residual buffer.
- The DNA was eluted by the addition of 50µl of TE buffer at 65°C, incubation at room temperature for 1 minute and centrifugation at 12,000g for 2 minutes.
- The purified DNA was subsequently roughly quantified by agarose gel electrophoresis (Section 2.5.2).

2.4.3 Wizard SV Gel and PCR Clean-Up System (Promega)

This kit was used in the purification of PCR products and the purification of DNA from agarose gels as per the supplier's instructions. Briefly, this involved:

- For PCR purifications, an equal amount of 'Membrane Binding Solution' was added to each PCR reaction.
- Or, for agarose gels, 10µl of 'Membrane Binding Solution' was added per 10mg of agarose gel, followed by melting of the agarose at 65°C for 10 minutes.
- The DNA was bound to an SV minicolumn by centrifugation at 10,000g for 1 minute.
- The column was washed with 700µl of 'Membrane Wash Solution' followed by centrifugation at 10,000g for 1 minute and a second wash with 500 µl of 'Membrane Wash Solution', followed by centrifugation at 10,000g for 5 minutes.
- The DNA was then eluted into 50µl of dH_2O by centrifugation at 10,000g for 1 minute.
- Purified gel fractions and PCR products were then used as described in Sections 2.13 and 2.6.2, respectively.

2.5 QUANTIFICATION AND ELECTROPHORESIS OF DNA

2.5.1 UV Spectrophotometric analysis of purified DNA

Purified gDNA samples were quantified by UV spectrophotometry on an Ultrospec III UV/Vis (Pharmacia) or DU Series 500 UV/Vis (Beckman, Fullerton CA, USA) spectrophotometer. Concentration of double stranded DNA was determined according to the formula: 1 A260 unit = 50mg DNA/mL (Sambrook *et al.*, 1989). Oligonucleotide concentrations were determined according to the formula provided by the supplier (Sigma Genosys) based on the base composition of each oligonucleotide. DNA was considered pure with a 260nm/280nm ratio of 1.75-1.80 (Sambrook *et al.*, 1989). The appropriate quantity of purified gDNA was then used as a template in PCR reactions (Sections 2.9.1 and 2.16.3) or digested for Southern blots (Sections 2.8.1 and 2.18.1).

2.5.2 Estimation of DNA concentration by agarose gel electrophoresis

When determining the exact concentration of DNA was not critical, or the when the amount of DNA was limited, the concentration of PCR products were estimated by comparison of the intensity of the DNA sample on an agarose gel with that of the DNA standards of known concentration. The appropriate concentration of PCR product was then used directly in sequencing reactions (Sections 2.6.2), radiolabeled as a probe (Sections 2.8.2, 2.11.2 and 2.18.2) or cloned into pGEM-T-Easy (Section 2.9.2).

2.5.3 Agarose gel electrophoresis

Agarose gel electrophoresis was used for a number of purposes throughout this research, including:

- Visual assessment of the results of PCR reactions (Sections 2.8.2, 2.9.1, 2.10.3, 2.15.3, 2.16.3, 2.17, 2.18.2 and 2.18.3) and restriction digestions (Sections 2.9.3, 2.11.4, 2.14.1, 2.14.2 and 2.19.1);
- Quantification of PCR products (Section 2.5.2);
- Visual confirmation of the success and quality of genomic and plasmid DNA preparations (Sections 2.2.2, 2.9.3 and 2.11.4);

- Separation of restriction digested gDNA for Southern blotting (Sections 2.8.1 and 2.18.1);
- Size fractionation of gDNA for the PCR isolation of a *PDI* gene (Section 2.13).

Aliquots to be separated by agarose gel electrophoresis were mixed with 6x Loading Dye Solution (MBI Fermentas) to a final loading concentration of 1X and electrophoresed on agarose gels (1% - 2%) at 35-100V for times ranging from 45 minutes to 16 hours (details in relevant sections, below). All gels were prepared with 1X TAE and ethidium bromide (0.6μg/ml) (Sambrook *et al.*, 1989). Appropriate molecular weight markers (MBI Fermentas), eg. λDNA-*Eco*RI/*Hind*III (0.564kb-21.22kb), GeneRuler 100bp DNA Ladder (100bp – 1000bp) or GeneRuler 1kb DNA Ladder (0.25kb-10kb) were used for estimating the size of DNA samples, as per Sambrook *et al*. (1989). The gels were viewed and photographed on a UV/White Darkroom (UVP, Upland CA, USA) and analysed with Labworks Analysis Software (UVP).

2.6 DETERMINATION OF DNA SEQUENCES

2.6.1 Primers for sequencing reactions

The primers used to initially sequence cDNA clones were designed on vector-based sequences (Table 2.4). For full-length sequencing of clones primers were designed progressively based on the results of previous sequencing reactions, thereby effectively ‘primer-walking’ down each DNA strand of the clone. In sequencing reactions where the template was a purified PCR product, the primers used in the amplification of the product were used to sequence each DNA strand.

Table 2.4 Vector-based primers used in this study

Primer	Sequence 5'→ 3'	Length (bp)	Vector(s)
T3	AATTAACCCTCACTAAAGGG	20	pBluescript SK$^+$ (Stratagene)[a]
T7	GTAATACGACTCACTATAGGGC	22	pBluescript SK$^+$ (Stratagene)[a] pGEM-T-Easy (Promega)[a]
SP6	ATTTAGGTGACAGTATAGAATAC	23	pGEM-T-Easy (Promega)[a]

[a] Full vectors maps are given in Appendix II.

2.6.2 Sequencing reactions

DNA sequencing was performed according to the method of Sanger *et al.* (1977) and was conducted using the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit v.3 (Applied Biosystems). Sequencing reactions were typically conducted in 20μL final volumes containing 2μL of the commercial BigDye solution, 1.5mM MgCl2, 37.5mM Tris-HCl (pH 9.0), 3.2 pmoles of primer and ~50ng of PCR product or ~200ng of plasmid DNA. Thermal cycling was undertaken in the PTC-100 or PTC-200 Peltier Thermal Cycler (MJ Research) according to a modified version of the standard protocol (Applied Biosystems), consisting of 45 cycles, each of 96°C for 30 seconds, 55°C for 15 seconds and 60°C for 4 minutes, followed by five cycles of 96°C for 30 seconds and 60°C for 4 minutes. The sequencing reactions were ethanol-precipitated according to the supplied protocol (Applied Biosystems) and air-dried prior to being analysed on an ABI 373A automated sequencer at the Micromon DNA Sequencing Facility, Department of Microbiology, Monash University, Clayton 3168, Australia. To ensure the accuracy of the sequences, both strands of all clones and PCR products were sequenced and regions of low-quality or ambiguous sequence were clarified through replicate sequencing reactions.

2.6.3 DNA sequence analysis

Sequence editing and alignments were conducted using the BioEdit software package v5.0.9 (www.mbio.ncsu.edu/BioEdit/bioedit.html; Hall, 1999). Unless otherwise stated, the software utilised for sequence analysis was accessed via the BioManager and WebANGIS interfaces of the Australian National Genomic Information Service (ANGIS; www.angis.org).

2.6.3.1 Quality assessment of DNA sequences

The quality of DNA sequence was assessed by visual analysis of the trace file using the software package Chromas v1.45 (http://www.technelysium.com.au/chromas.html). Areas of ambiguity ('N' bases) could often be resolved by visual checking of the chromatogram; where this was not possible, replicate sequencing reactions were conducted, as described above (Section 2.6.2).

2.6.3.2 Alignment of multiple sequences

Multiple sequence alignments of DNA and amino acid sequences were performed with the "Eclustalw" program of ANGIS or the "clustalw" program in BioEdit. The conventional conditions were a gap opening penalty of 10.0 and a gap extension penalty of 0.05. Multiple sequence alignments were prepared for publication and areas of sequence conservation or heterogeneity were highlighted using the "Graphic View" option in BioEdit.

2.6.3.3 Comparison of two sequences

Pairwise sequence alignments were conducted with the "Gap" program in ANGIS. The conventional conditions for these comparisons were a gap creation penalty of 5.0 and a gap extension penalty of 0.3. This program also provided the level of sequence identity and similarity (in amino acid alignments).

2.6.3.4 Identification of restriction enzyme recognition sites

The DNA sequences of the PDI and *cyclophilin* genes were analysed for any restriction enzyme recognition sites within, using the "Map" program of ANGIS or "Restriction Map" program of BioEdit. These programs were also used in the selection of suitable restriction enzymes to exploit the polymorphic sites in the *PDI* alleles and design a PCR-RFLP assay to identify the molecular markers for these (Section 2.14.1).

2.6.3.5 BLAST analyses of sequences

The BLAST suite of programs were accessed from the National Centre for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov/BLAST/). These programs were used to identify similar DNA (blastn) or protein (blastx or blastp) sequences in the non-redundant ("nr") databases. In all cases, the default program settings were used.

2.6.3.6 Identification of open reading frames and translation of sequences

The intron/exon structure of the *PDI* genes from *Ae. tauschii* and *T. turgidum* were identified by their alignment with the *PDI* cDNA sequences (Johnson *et al.*, 2001), the *PDI* genes reported previously (Ciaffi *et al.*, 2001) and identification of the conserved GT/AG intron/exon junctions. Open reading frames (ORFs) within the exon sequences of these genes as well as the cDNAs *TaCYP20-1*, *TaCYP20-2* and *TaCYP70* cDNAs were identified using the "Flip 6 Frames" program of ANGIS. Translation of the ORFs was conducted using the "Translate" program of ANGIS or the "Translate or Reverse Translate" command in BioEdit. The start of translation was assumed to be the first ATG codon in the ORF and the amino acid sequences were trimmed to reflect the putative protein product.

2.6.3.7 Prediction of protein physical properties and subcellular localisation

The molecular weight (MW) and isoelectric points (pI) of the putative proteins encoded by *TaCYP20-1*, *TaCYP20-2* and *TaCYP70* were estimated using the "PepStats" program of ANGIS and functional motifs within the protein sequences were identified by

searching the PROSITE database (www.expasy.ch/prosite) (Falquet *et al.*, 2002). The subcellular localisation of *TaCYP20-1*, *TaCYP20-2* and *TaCYP70* as well as the *PDI* and *cyclophilin* genes identified in the TIGR TaGI database was predicted using the "TargetP" (www.cbs.dtu.dk/services/TargetP; Emanuelsson *et al.*, 2000) and/or "PSORT" (http://psort.nibb.ac.jp; Nakai and Kanehisa, 1992) programs.

2.7 SOUTHERN BLOTTING

2.7.1 Electrophoresis conditions

Determination of the copy number of the *PDI* and *cyclophilin* gene families (Sections 2.8 and 2.18) was conducted using the technique of Southern blotting (Southern, 1975). This involved the digestion of gDNA isolated from a number of wheat species as described in the relevant sections below (Sections 2.8.1 and 2.18.1). Digested DNA was separated by agarose gel electrophoresis in 1% TAE gels containing ethidium bromide (0.6μg/mL). The gels were run at 35V for ~16 hours and the DNA was visualised in the UV/White Darkroom (UVP Laboratory Products) to confirm digestion of the DNA and record the distance that the molecular weight markers migrated in the gel.

2.7.2 Southern blotting of gels

The DNA was denatured and prepared for blotting according to the methods outlined by the Hybond N+ and Hybond XL supplier (Amersham). Briefly, this involved depurination of the DNA in 0.125M HCl with gentle agitation for 10 minutes, separation of the DNA strands in denaturation buffer (0.5M NaOH, 1.5M NaCl) for 30 minutes followed by gentle agitation in neutralisation buffer (0.5M Tris-HCl pH 7.2, 1.5M NaCl) for 30 minutes. The gel was briefly washed with distilled water between each of these treatments.

The gels were blotted onto Hybond N+ or Hybond XL positively charged nylon membranes according to the method of capillary transfer (Southern, 1975) as outlined in Sambrook *et al*. (1989). This involved transfer of the DNA from the gel to the nylon membrane by the capillary action of the transfer buffer (10x SSC) from a reservoir, via

a wick of Whatman 3M paper, through the gel/membrane, to paper towels with a moderate weight on top. Transfer was allowed to proceed for 16-24 hours, after which the blot was disassembled and the membrane allowed to air dry. Crosslinking of the DNA to the membrane was carried out by exposure to UV at 302nm on a 2011 Macrovue Transilluminator (LKB, Bromma, Sweden) for 4 minutes.

2.7.3 Probing of Southern blots

2.7.3.1 Design of probes

The Southern blots were probed with either *PDI* (Section 2.8.2) or *cyclophilin* (Section 2.18.2) probes. These probes were amplified under standard PCR reaction conditions (Section 2.3.2) using primers outlined in the appropriate sections below (Sections 2.8.2 and 2.18.2). The amplified PCR products were purified, visually assessed for size and roughly quantified by electrophoresis prior to labelling (Sections 2.4 and 2.5).

2.7.3.2 Radiolabeling of probes

Radiolabeling of the probes with [α^{32}P]-dCTP was conducted using the Prime-a-Gene Labeling System (Promega) or Ready-To-Go DNA Labeling Beads (Pharmacia), according to the suppliers' protocols. Both methods are based on the protocols established by Feinberg and Vogelstein (1983) and involve the extension of random hexamers by the Klenow fragment of *E. coli* DNA Polymerase I using the denatured DNA fragments as templates. The Prime-a-Gene Labeling System (Promega) method involved the initial denaturation of ~50ng of amplified probe at 95°C-100°C for 5 minutes followed by chilling on ice. This was followed by the addition of dATP, dGTP, dTTP to a final concentration of 25mM each,[α^{32}P]-dCTP 50µCi, 3,000Ci/mmol to a final concentration of 333nM, 1X of the supplied labeling buffer and Klenow fragment (100u/mL) and incubation at 37°C for 1 hour (Promega) before stopping the reaction by the addition of 10mM EDTA. The Pharmacia method was the same, with the exception that the dATP, dGTP, dTTP and Klenow fragment were provided in a dehydrated bead-form and the denatured probe and [α^{32}P]-dCTP were added and incubated at 37°C for 30 minutes before stopping the reaction by the addition of EDTA, as described above.

Unincorporated nucleotides were removed from labelling reactions using G-50 ProbeQuant Microspin Columns (Pharmacia) according to the supplied protocol. Briefly, this involved packing the column by centrifugation at 3000g for one minute, loading of the labelled probe to the column and centrifugation at 3000g for two minutes to elute the labelled probe, while the small MW unincorporated dNTPs and small fragments (<20bp) were retained in the column (Pharmacia protocol). The probe was then denatured in a boiling water bath for 5 minutes, briefly chilled on ice and applied to the membrane.

2.7.3.3 Hybridisation and washing of the blots

Prehybridisation of Southern blots was undertaken to minimise non-specific binding of the probe to the membranes. The Southern blots were incubated in roller tubes at 15 rpm in prehybridisation buffer I (Section 2.1.3) at 65°C for at least two hours prior to probing. Alternatively, DIG Easy Hyb (Roche) supplemented with sheared herring sperm DNA (0.1mg/mL) was used in the place of prehybridisation buffer I at 50°C. The denatured radiolabeled probes (Section 2.2.7.2) were added to the prehybridisation buffer and allowed to hybridise overnight at 65°C or 50°C (same as the respective prehybridisation temperature). The blots were then washed twice in a low stringency wash solution (2xSSC, 0.1% SDS) at 65°C for 10 minutes, followed by 20 minutes in a high stringency wash solution (0.1X SSC, 0.1% SDS) at 65°C to remove any unhybridised probe.

2.7.3.4 Detection of hybridised probe

Hybridised probe was detected by exposure of the washed Southern blot to a BAS-MS 2340 imaging plate (FujiFilm) for 6 –12 hours. Scanning of the imaging plate was conducted in an FLA-3000 phosphorimager (FujiFilm) and analysis of results was conducted with Image Gauge v3.12 (FujiFilm) software. In some cases the blots were exposed to BioMax X-ray film (Kodak) for 48-96 hours at -80°C and developed by gentle agitation in G150 Developer (Agfa) for 4 minutes, washing in water for 2 minutes and gentle agitation in G354 Manual Fixing Bath (Agfa) for 4 minutes.

SPECIFIC MOLECULAR METHODS

2.8 DETERMINATION OF COPY NUMBER OF THE *PDI* GENE FAMILY

2.8.1 Southern blots of *T. turgidum*, Ae. *tauschii* and *T. aestivum* gDNA

As an initial step in characterising the *PDI* gene families of *T. turgidum* and *Ae. tauschii* Southern blots were performed to establish the gene copy number in these species. This was of particular importance for *T. turgidum*, as this species has been reported to contain either two or three genes in different cultivars, while only two partial gene sequences have been characterised in this species (Section 1.5.5). Likewise, the *PDI* gene family of *T. aestivum* has been reported to contain three or four genes, in different cultivars (Section 1.5.5), while only three cDNAs encoding *PDI* have been characterised from this *T. aestivum* cv. Wyuna (Section 1.5.5), thus, Southern blots were performed to clarify the copy number in this cultivar. 10µg of *T. turgidum*, *Ae. tauschii* or *T. aestivum* cv. Wyuna gDNA was digested with 30U each of *Eco*RI, *Bam*HI or *Hind*III and double digests containing two of these enzymes, in 1X of the supplied buffer at 37°C for 4 hours. The restriction digested gDNA was electrophoresed on 1% agarose gels and transferred to Hybond N^+ or Hybond XL membrane, as described in Section 2.7.2 and probed with the 838bp *PDI* probe, described below (Section 2.8.2).

Southern blots were also conducted on the eight parental cultivars used in the establishment of the mapping populations (Section 1.8.4) analysed in this study to establish the *PDI* gene copy number in these cultivars. As a positive control, *T. aestivum* cv. Chinese Spring was also included, as this cultivar has been reported to contain the additional fourth hybridising fragment, which is especially clear in the *Eco*RI digests of the earlier Southern blots (Section 1.5.5). 10µg of gDNA from *T. aestivum* cvs. Chinese Spring, Cranbrook, Halberd, Egret, Sunstar, Sunco, Tasman, CD87 and Katepwa was digested with 30U of *Eco*RI at 37°C for 4 hours. The digested DNA was then electrophoresed on 1% agarose gels at 35V for 16 hours and transferred to Hybond XL as described in Section 2.5.3. These Southern blots in were subsequently probed with the 838bp *PDI* probe described below (Section 2.8.2).

Further, the identification of three hybridising fragments in each of the single digests on the Southern blots of *T. aestivum* cv. Wyuna and Chinese Spring Southern blots

(Chapter 4), when four genes had been reported in this species (Section 1.5.5), raised the question as to whether this result was due to the actual absence of a fourth gene in our samples of these cultivars or due to differences in the experimental conditions in these two studies. To address this question, 10μg of gDNA from *T. aestivum* cvs. Wyuna and Chinese Spring was digested with *Eco*RI (30U) for 4 hours at 37°C, separated by electrophoresis in 1% agarose gels at 35V for 16 hours, the digests transferred to Hybond XL membrane (Amersham) according to the methods outlined above (Section 2.7.2) and the blot was probed with the 1466bp PDI probe described below (Section 2.8.2).

2.8.2 Amplification and radiolabelling of *PDI* gene fragments

An 838bp product was generated from the *wPDI1* cDNA clone (Section 1.5.5; Figure 2.2) by PCR amplification with the PDISF and PDI2R primers (Figure 2.2; Table 2.5), using the conditions outlined in Section 2.3.2. Alternatively, a 1466bp cDNA fragment, containing the entire coding sequence except for the carboxyl end of the protein, and similar to that used by Ciaffi *et al.* (1999), was amplified from the *wPDI1* cDNA clone with the primers PDI1F and PDILR (Figure 2.2; Table 2.5).

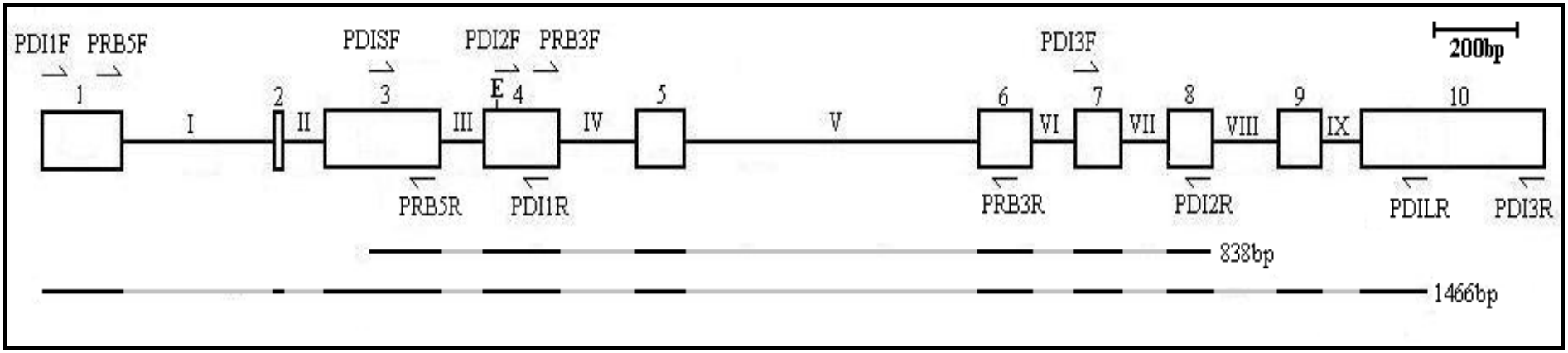


Figure 2.2 Consensus primer locations and probes used in the characterisation of the *PDI* gene family

The positions of the primers used to amplify the *PDI* genes from *T. turgidum* and *Ae. tauschii* based on consensus sequences in the previously characterised *PDI* cDNA clones (Johnson *et al.*, 2001). In addition, the locations of primers used in the amplification of probes and characterisation of the *PDI* gene on chromosome 1B are shown. The intron/exon organization of the *PDI* gene is based on that reported in the literature (Ciaffi *et al.*, 2001) and the results generated in the current study, with exons being numbered boxes and introns shown as lines between them. The direction of arrows indicates primers in the sense (arrow pointing right) and antisense (arrow pointing left) direction. The conserved *Eco*RI restriction site in the three cDNA clones (Johnson *et al.*, 2001) is indicated (E). Details of the primer sequences are listed in Table 2.3. The 838bp and 1466bp cDNA probes (amplified from the *wPDI1* cDNA clone) used in the *PDI* Southern blot experiments are shown below, highlighting the areas of homology (in black) between the probes and the exons of the three characterised *PDI* genes.

Table 2.5 PDI primers based on consensus sequences

Primer pairs and sequences (5'-3')[1]	Annealing temperature used for PCR (°C)	Expected size of PCR product (bp)[2]	Purpose
PDI1F: TCACTGCTCCCCAGTCCCTTC PDI1R: AACCAATGGCCTCTCCACTGC	64	1221-1248	To amplify the *PDI* genes from *T. turgidum* and *Ae. tauschii*
PDI2F: GGTCTGATTATGACTTTGGCC PDI2R: CCAGATTTGAAGACCACGTCG	60	1612-1671	
PDI3F: CTGAAAGAGGATCAGGCACC PDI3R: GATAATTTAACTAGGTACACCAC	50	1034-1045	
PRB5F: CTTCATCCTCGTCGAGTTCTA PRB5R: AATGTGGATCTTGCCGTCTT	59	786-802	To investigate the presence of a fourth *PDI* gene in *T. aestivum* cv. Chinese Spring
PRB3F: TGAGCTCGTTGTTGACAGCAA PRB3R: CTGAACTCCTCTACAGCACCA	59	1018-1085	
PRB5F: CTTCATCCTCGTCGAGTTCTA PDI1R: AACCAATGGCCTCTCCACTGC	59	1026-1053	
PDISF: CCTCAAGATCTTCAGGAACGG PDI2R: CCAGATTTGAAGACCACGTCG	60	846	To amplify the probes from the cDNA clone *wPDI1* for the Southern blot and cDNA library secondary screen experiments
PDI1F: TCACTGCTCCCCAGTCCCTTC PDILR: GTAGTCGACAATCTCGTCGGCCG	62	1466	

[1] The positions of these primers are shown in Figure 2.2.

[2] The expected size range is that predicted from the *PDI* genes from *T. turgidum* (Ciaffi *et al.*, 2001 and current study) and *Ae. tauschii* (current study), with the exception of the last two pairs, which were based on the sequence of the cDNA clone *wPDI1* (Johnson *et al.*, 2001).

2.8.3 Probing and analysis of Southern blots

The above-amplified *PDI* gene fragments were radiolabeled and used to probe the blots as described in Section 2.7.3. Standard hybridisation and washing methods (Section 2.7.3) were used and hybridised probes were detected on phosphorimager plates and autoradiography (Section 2.7.3.4). The sizes of the hybridising fragments were estimated by comparing the distance that the hybridising fragments migrated to molecular weight standards on the original agarose gel.

2.9 THE *PDI* GENES OF *T. turgidum* subsp. *durum* AND *Ae. tauschii*

2.9.1 Amplification of the *PDI* genes of *T. turgidum* and *Ae. tauschii*

Partial genomic sequences have been reported for the *PDI* genes from the A and B genomes of *T. turgidum* (Section 1.5.5), however, the 3' end of the coding sequence and 3' untranslated regions, as well as entire gene from the D genome of *Ae. tauschii*, which may be useful the isolating the corresponding genes from *T. aestivum*, have yet to be elucidated. Therefore, based on the analysis of the three *T. aestivum* cDNA clones, *wPDI1*, *wPDI2* and *wPDI3* (Section 1.5.5), consensus primers (PD1F/PD1R, PD2F/PD2R and PD3F/PD3R) were designed to amplify the genomic copies of these *PDI* genes from the proposed D genome progenitor, *Ae. tauschii* and A and B genome tetraploid progenitor, *T. turgidum*. All three primer pairs were used to amplify the *PDI* gene(s) from *Ae. tauschii* as three overlapping fragments, while one pair (PD3F/PD3R) was used to amplify the uncharacterised 3' end of the *PDI* genes from the A and B genomes of *T. turgidum.*

PCR amplifications were carried out as described in Section 2.3.2. Aliquots of the PCR products were visualised by agarose gel electrophoresis in 1% gels (Section 2.5.3) and PCR products were purified with the Concert Rapid PCR Purification System (Invitrogen) according to the supplied protocol (Section 2.4.2). Purified PCR products were then cloned and sequenced as described below.

2.9.2 Cloning of *PDI* PCR products

The pGEM-T-Easy vector (Promega), containing terminal deoxythymidines at the 3' ends of both strands of the linearised plasmid to increase the efficiency of cloning PCR products generated using a DNA polymerase that leaves an overhanging deoxyadenosine in a template-independent manner at the 3' ends of amplified fragments, was utilised for the cloning of PCR products. For ease of insert removal, this vector contains *Eco*RI restriction sites flanking the cloning site. In addition, this vector contains the coding region of the α-peptide of β-galactosidase (*LacZ*'), allowing α-complementation of this subunit with that encoded by the appropriate host cells and thus, differentiation of colonies containing recombinant plasmids (white/colourless) from non-recombinant plasmids (blue) on media containing IPTG and X-gal. The pGEM-T-Easy vector map is provided in Appendix II. The methods used were supplied with the vector; briefly, a typical 10μl ligation reaction contained 1μL pGEM-T-Easy vector (50ng/μl), 3μL purified PCR products (10-20ng/μL), 5μL T_4 DNA ligase 2x buffer and 1μL T_4 DNA ligase (5U/μL). Ligations were allowed to proceed overnight (~16 hours) at 4°C and subsequently transformed into the supplied competent *E. coli* JM109 (e14-(McrA-) *recA1 endA1 gyrA96 thi-1 hsdR17*(r_K- m_K+) *supE44 relA1* Δ(*lac-proAB*) [F' *traD36 proAB lacI*^{q}*ZΔM15*]) cells. This was achieved by thawing the cells on ice for 5 minutes, mixing with the ligation reactions and storing on ice for 20 minutes. The cells were then heat-shocked at 42°C for 50 seconds, allowed to recover on ice for 2 minutes before the addition of 950μl of SOC medium. The cells were then incubated at 37°C for 1 hour with agitation (150 rpm) and 100μl was plated onto duplicate LB-colour selection plates. These plates contained 50μg/ml ampicillin, 0.5mM isopropylthio-β-D-galactoside (IPTG) and 80μg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal). The duplicate plates were incubated overnight (~16 hours) at 37°C to allow colony formation. Six white colonies (ie. containing recombinant plasmids) were selected from each PCR reaction for plasmid purification and sequencing.

2.9.3 Preparation and restriction analysis of plasmid DNA

Selected recombinant colonies were inoculated into 4mL LB broth containing ampicillin (50μg/mL) and incubated overnight in a shaking incubator at 37°C, 200rpm. Plasmid DNA was purified from 3mL of culture with the "Wizard *Plus* SV Minipreps DNA Purification System" (Promega). Briefly, this involved centrifugation of 1.5mL of overnight culture at 10,000g for 5 minutes, discarding the supernatant and repeating the centrifugation with another 1.5mL of culture. Plasmid DNA was purified as follows:

- The cell pellet was thoroughly resuspended in 250μL of "Cell Resuspension Solution".
- Cell lysis was achieved by the addition of 250μL of "Cell Lysis Solution" and incubation at room temperature for 5 minutes.
- 20μL of "Alkaline Protease Solution" was added and incubated at room temperature for 5 minutes to inactivate any endonucleases.
- 350μL of "Wizard *Plus* SV Neutralization Solution" was added and gently mixed.
- The cell lysate was centrifuged at 14,000g for 10 minutes and the supernatant added to a supplied spin column.
- The column was centrifuged at 14,000g for 1 minute and the flowthrough was discarded.
- The column was washed with 750μL of "Column Wash Solution" and centrifuged at 14,000g for 1 minute.
- The column was washed again with 250μL of "Column Wash Solution", followed by centrifugation at 14,000g for 2 minutes.
- The plasmid DNA was eluted from the column by the addition of 100μL dH_2O and centrifugation at 14,000g for 1 minute.

An aliquot of the purified plasmid was digested with 5U of *Eco*RI to remove the cloned insert in 1X of the supplied buffer at 37°C for 1 hour. The presence of the appropriate sized insert was confirmed by agarose gel electrophoresis (Section 2.5.3) and the inserts of selected plasmids were then sequenced as follows.

2.9.4 Sequencing and analysis of *PDI* clones

Sequencing of the cloned *PDI* inserts in pGEM-T-Easy was conducted according to the method described in Section 2.6.2 using the vector-based primers T7 and SP6 (Table 2.4) followed by 'primer walking' wherein internal primers were designed based on sequence data as it was generated for subsequent sequencing reactions.

Multiple sequence alignment was conducted as described in Section 2.6.3.2 to identify overlapping regions of contiguous clones and sequences unique to each of the homeoalleles of *T. turgidum PDI* to allow the construction of contiguous sequences for each gene.

2.10 AMPLIFICATION OF *PDI* GENE SECTIONS FROM *T. aestivum*

2.10.1 Allele-specific primer design

The sequences of the three *PDI* cDNAs, *wPDI1*, *wPDI2* and *wPDI3* (Section 1.5.5), the published genomic sequences from *T. turgidum* (Section 1.5.5) and the sequences from *T. turgidum* and *Ae. tauschii* generated during this study (Section 2.9 and Chapter 3) were used to identify nucleotide positions unique to each of the three *PDI* homeoalleles. These positions were used to design primer sequences that were able to discriminate between the three *PDI* genes of *T. aestivum*, wherein the terminal 3' base of both the forward and reverse primers was unique to the homeoalleles to be amplified (an example is shown in Figure 2.3). Wherever possible, these primers were designed to flank non-coding sequences, eg. introns and 5' and 3' UTRs. The details of these allele specific (AS)- primers are shown in Table 2.6 and their positions are illustrated in Figure 2.4.

2.10.2 Confirming the allele-specificity of primers

The primers that were based solely on cDNA sequences were first used to amplify the cDNA clones *wPDI1* (PA1F/PA1R, PA3F/PA3RB and PA4F/PA4R), *wPDI2* (PB1F/PB1R, PB2F/PB2R and PB3F/PB3R) and *wPDI3* (PD2F/PD2R, PD4F/PD4R, PD5F/PD5R and PD6F/PD6R) to ensure that they amplified only the clone targeted by the PCR. As this strategy would not work for the primers based on intron sequences in

the *PDI* genes of *T. turgidum* and *Ae. tauschii* (PA2F/PA2R, PA3F/PA3R, PD1F/PD1R and PD3F/PD3R; Table 2.6), specificity of these primers pairs was confirmed by direct sequencing of these PCR products. As the different *PDI* genes have characteristic sequences (Johnson *et al.* 2001, Ciaffi *et al*. 2001 and Section 3.3.3), mixed populations of PCR products amplified from multiple loci, due to non-specificity of AS-primer annealing, would appear as areas of sequence ambiguity due to SNPs or overlapping, nonsense sequences due to insertion/deletions (indels) between the homeoalleles (Section 3.3.3). Thus, allele specificity was checked by ensuring unambiguous sequence data was generated from all AS-primer pairs during the initial amplification and sequencing of *T. aestivum* cv. Katepwa gDNA. Initial annealing temperatures (Table 2.6) were determined using the web-based software NetPrimer (www.premierbiosoft.com) and increased in 1°C increments, until allele specificity was achieved.

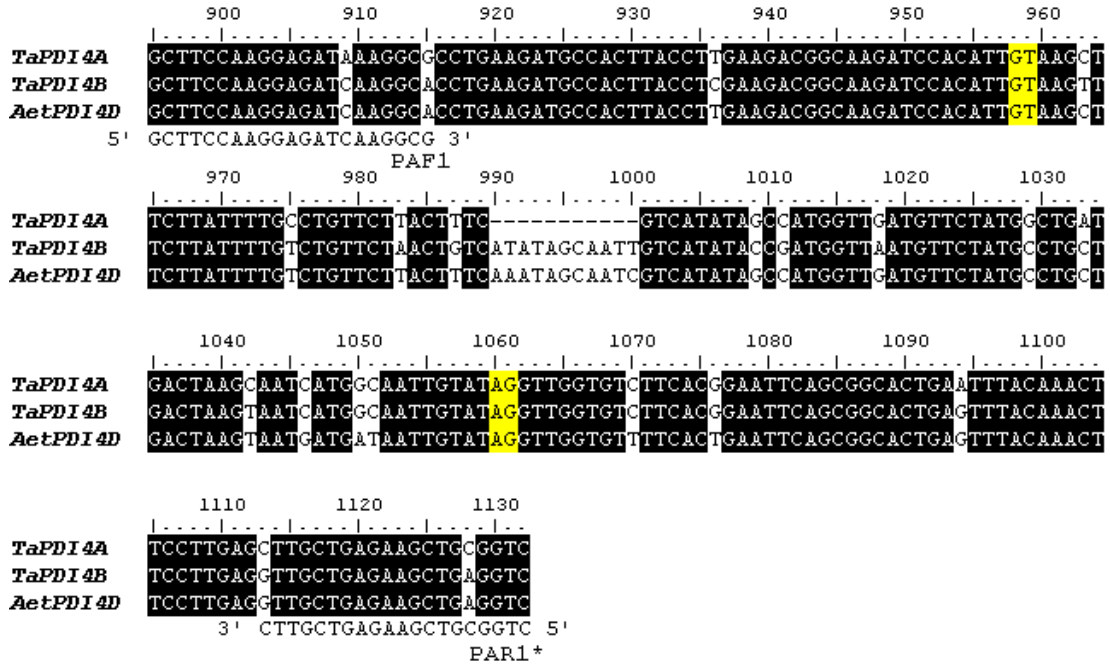


Figure 2.3 An example of the design of allele specific primers

The sequences of the homeoalleles from *T. turgidum* (*TaPDI4A* and *TaPDI4B*) and *Ae. tauschii* (*AePDI4D*) (Chapter 3) flanked by the allele specific primer pair PAF1/PAR1. This alignment illustrates the basis for primer allele-specificity, in this case for *TtPDI4A*. *The complementary sequence of PAR1 is shown. The GT/AG at the beginning and end of intron 3 are highlighted. The numbering system is based on *TtPDI4A* (Chapter 3).

Table 2.6 The allele-specific primers used to amplify the *PDI* genes of wheat

	Primer pairs and sequences (5'-3')[a]	Expected gDNA PCR product[b] (bp)	Annealing temperatures Calculated[c]/Actual (°C)	Expected cDNA PCR product[d] (if applicable) (bp)
A Genome	PA1F: GCTTCCAAGGAGATAAAGGCG PA1R: GACCGCAGCTTCTCAGCAAG	227	61/64	134
	PA2F: GAATTTACAAACTTCCTTGAGC PA2R[e]: AAAAATATACAGCAGTAATCATCTG	650	54/57	N/A
	PA3F: CAACCATCCTTACCTCTTGAAATA PA3R[e]: TAAAGAAATAGAAAGGGAAACAAT	900	55/58	N/A
	PA3F: CAACCATCCTTACCTCTTGAAATA PA3RB: GCCTGATCCTCTTTCAGGCCA	987	58/60	210
	PA4F: GAGGCCAACAATGAGCCTGTG PA4R: ATACGAGACCTTCTTCCCGCTA	511	60/62	282
B Genome	PB1F: CCTGAAGATGCCACTTACCTC PB1R: CCAATGGCCTCTCCACTGCG	291	57/61	185
	PB2F: CTGCTTTGGAGAAATTCATTGAG PB2R: AGGAACTTGACATCCTTGCCG	888	58/63	214
	PB3F: TTGCTTGGTTGAAGGATTACTTC PB3R: CCCTCGTAGGAGACCTTCTTT	653	57/62	352
D Genome	PD1F[e]: CCCTCGCCTCTGGTGTCT PD1R[e]: CATGGCTATATGACGATTGCTATT	824	58/62	N/A
	PD2F: GGGAGGCTGAGGGAATTGTC PD2R: AGTGCCGCTGAATTCAGTGAAA	244	61/65	140
	PD3F[e]: TTGTATAGGTTGGTGTTTTCACT PD3R: CATTGGTCTGGAAGAATTTCAG	488	57/58	N/A
	PD4F: CTGACAACCATCCTTACCTCC PD4R: GTCACTGTCTTGAATGAGGATC	960	53/60	242
	PD5F: AAAGAGGATCAGGCACCACTG PD5R: GCGATCACAACGTCCTCTTCA	553	59/62	332
	PD6F: GCTGCCACCCTTCAAAGT PD6R: TACAGTATTTCTCGCAACGGGA	519	55/56	434

[a] The positions of these primers are shown in Figure 2.4. Nucleotides unique to the amplified gene are highlighted in yellow.

[b] The expected size is predicted from the *PDI* genes from *T. turgidum* (A and B genomes) (Ciaffi *et al.*, 2001 and Section 3.3) and *Ae. tauschii* (D genome) (Section 3.3).

[c] Calculated annealing temperatures were determined with NetPrimer (www.premierbiosoft.com).

[d] The expected sizes are from *wPDI1* (A genome), *wPDI2* (B genome) or *wPDI3* (D genome) (Section 1.5.5).

[e] Primers based on intron sequences in the *PDI* genes of *T. turgidum* or *Ae. tauschii* (Chapter 3).

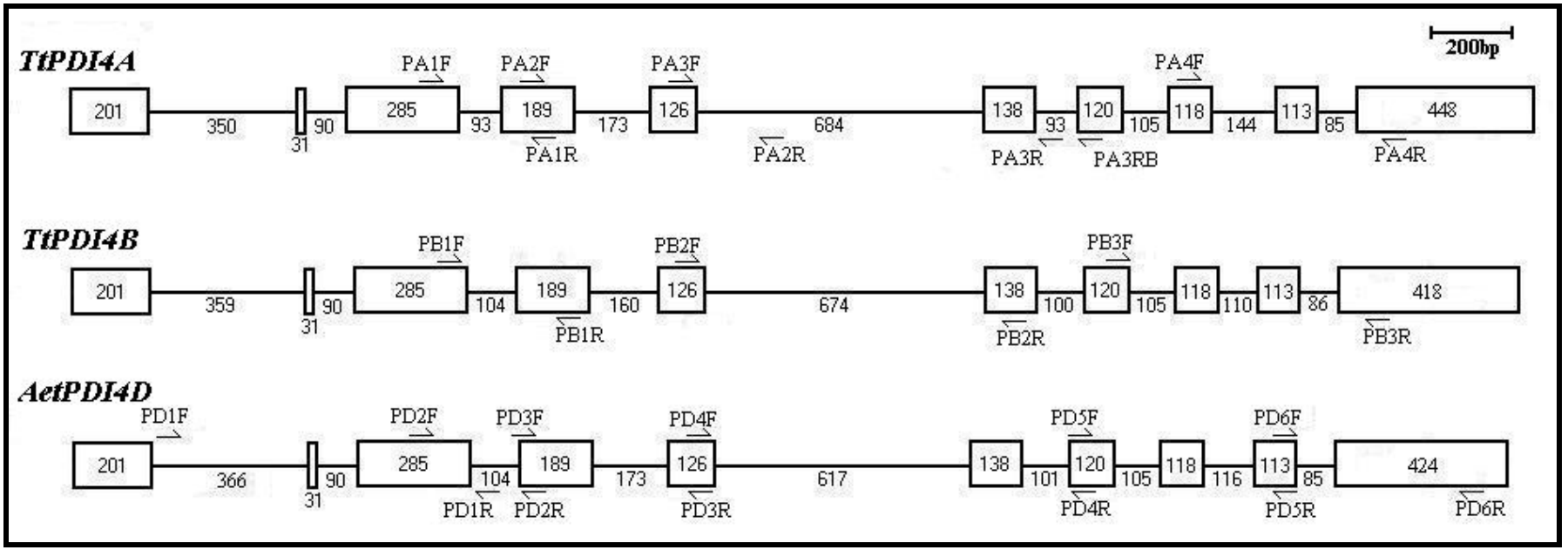


Figure 2.4 Allele-specific primer locations in the *PDI* genes of wheat.

The genomes (A, B and D), exons (boxes) and introns (lines joining boxes) are shown and are based on Ciaffi *et al*. (2001) and work generated in this study (Chapter 3). The sizes (bp) of the introns and exons are shown. Arrows pointing right or left indicate forward or reverse primers, respectively. Details of these primers are given in Table 2.6.

2.10.3 AS-PCR amplification and sequencing of *T. aestivum PDI* genes

Sections of the three PDI homeoalleles of *T. aestivum* were amplified from gDNA isolated from the parental cultivars of the mapping populations (Section 1.8.4.1), Cranbrook, Halberd, Egret, Sunstar, Sunco, Tasman, CD87 and Katepwa using standard PCR conditions (Section 2.3.2) with annealing temperatures specific for each primer pair (Table 2.6). Aliquots of the PCR reactions were electrophoresed on 1.5% agarose gels (Section 2.5.3) to test for successful amplifications. The PCR products were then purified using the Concert Rapid PCR Purification Kit (Invitrogen) (Section 2.4.2) and directly sequenced (Section 2.6.2) using the forward and reverse primers used in the amplification of the PCR product according. DNA sequences were aligned as described in Section 2.6.3.2 to identify any inter-cultivar polymorphisms, which were then confirmed by repetition of the PCR and sequencing of the appropriate section of the *PDI* gene from the appropriate cultivar(s).

2.11 CHARACTERISATION OF FURTHER PUTATIVE *PDI* cDNA PLASMIDS

As four *PDI* loci were reported in *T. aestivum* (Ciaffi *et al*. 1999; Section 1.5.5) while only three sequences represented by the cDNA clones *wPDI1*, *wPDI2* and *wPDI3* were discovered in the early phases of this work (J. Johnson, Honours Thesis, VU, 1999; Johnson *et al.* 2001; Section 1.5.5) further efforts were undertaken to identify any novel *PDI* genes. Since these three cDNA classes were originally isolated from a *T. aestivum* cv. Wyuna 10-12 days post-anthesis (DPA) cDNA library, 14 additional clones that had been co-isolated with the *wPDI1*, *wPDI2* and *wPDI3* sequence types in the primary screen of the library were characterised further and a secondary screen was conducted to ensure their homogeneity. The plating and screening was carried out essentially as per the supplied protocol (Stratagene) and Johnson *et al*. (2001); the methods are briefly as follows.

2.11.1 Plating of putative *PDI* Lambda Zap II clones and transfer to membranes

The plating cells *E. coli* XL-1 Blue (*recA1 endA1 gyrA96 thi-1 hsdR17 (rk-, mk-) supE44 relA1 lac* {F' *proAB lacI*^{q}ZΔ*M15* Tn*10* (Tet^{r})}) were grown in LB broth

supplemented with maltose (0.2%) and $MgSO_4$ (10mM) at 37°C for 6 hours. The cells were centrifuged at 1000g for 10 minutes and the cell pellet was resuspended in 10mM $MgSO_4$ to an OD_{600} of 1.0. The 14 previously isolated *PDI* clone phage stocks (Section 2.11) were diluted 1:100 with SM buffer and plated out on two plates for each clone, one containing 10µl and the other 100µl of the diluted phage stock, each with 200µl of *E. coli* XL-1 Blue. The cells were incubated at 37°C for 15 minutes to allow the phage to bind to the *E. coli* host cell surface after which the infected cells were mixed with LB top agar and plated out on LB agar plates, prewarmed to 37°C. The plates were incubated overnight (ON) at 37°C to allow plaque formation. The plaques were transferred to membranes and probed as described below.

2.11.2 Amplification and radiolabeling of the *PDI* probe

The 838bp *PDI* probe (Figure 2.2) used for the secondary screen was identical to that used in the primary screen of the cDNA library (Johnson *et al.*, 2001). The probe was amplified from the *wPDI1* cDNA clone using the PDISF and PDI2R primers (Table 2.5) and the standard PCR conditions (Section 2.3.2) with an annealing temperature of 60°C. The PCR product was purified and radiolabeled as described above (Sections 2.4.2 and 2.7.3.2) and applied to the membranes containing the putative *PDI* cDNA clones (Section 2.11.1).

2.11.3 Preparation and probing of the library lifts

The plaques were transferred to Hybond-N+ (Amersham) positively charged nylon membranes which were marked with a needle to enable the alignment of the membranes on the plates. The filters were incubated at room temperature in denaturation solution (1.5M NaCl, 0.5M NaOH) to separate the strands of λ cloning vector containing the putative *PDI* cDNA insert, followed by incubation in neutralisation solution (1.5M NaCl, 0.5M Tris-HCl pH 7.2, 1mM EDTA). The filters were then washed in room tempearture 2x SSC, followed by another wash in 2x SSC 0.1% SDS at 65°C.

The filters were placed in glass incubation bottles with prehybridisation buffer I at 65°C and incubated for 2 hours rotating at 10 rpm. The DNA probe (Section 2.11.2) was denatured in a boiling water bath for 10 minutes and then added to the hybridisation buffer, (prehybridisation buffer I + 10% dextran sulfate). The probe was incubated with the membranes overnight at 65°C, rotating at 10 rpm. Following hybridization, the membranes were washed twice in 2xSSC, 0.1%SDS at 65°C for 30 minutes, followed by a more stringent wash (0.1xSSC, 0.1% SDS) at 65°C for a further 30 minutes.

Detection of membrane-bound probe was achieved by exposing the membranes to BioMax Film (Kodak) for 12-24 hours at -80°C. The film was developed in G150 Developer (Agfa) for 4 minutes, followed by rinsing in water for 2 minutes, then incubation in G354 Manual Fixing Bath (Agfa) for 4 minutes and a final thorough rinse under water. The hybridisation signals on the X-ray film were realigned with the library plates using the alignment marks on the filters and plates described above to identify the plaques that corresponded to these hybridisation signals. The selected plaques were excised from the plate and stored in 500µl SM buffer with chloroform (1%).

2.11.4 *In vivo* excision and purification of plasmid DNA

The pBluescript plasmid DNA (Appendix II) containing the putative *PDI* cDNA inserts was 'rescued' from the Lambda Zap II vector by *in vivo* excision as described in the supplied protocols (Stratagene). The method used for the *in vivo* excision of the clones involved:

- Growing two separate cultures of *E. coli* XL-1 Blue and *E. coli* SOLR (e14$^-$ (mcr A) Δ(*mcr CB-hsd SMR-mrr*) *171 sbcC recB recJ uvrC umuC*::Tn*5* (Kan^r) *lac gyrA96 relA1 thi-1 endA1* λ^r {F' *proAB lacI*^{q}*ZΔM15* Tn*10* (Tet^r)} Su$^-$) in LB broth containing maltose (0.2%) and $MgSO_4$ (10mM) overnight at 37°C and 150rpm.
- Centrifugation of both cultures at 1000g for 10 minutes and resuspension of each in 10mM $MgSO_4$ until the absorbance at 600nm was 1.0.

- Incubation of 200μL of the *E. coli* XL-1 Blue with 250μL of the isolated Lambda Zap II clone and 10μL of the ExAssist helper phage (Stratagene) at 37°C for 15 minutes.
- Addition of 3ml of LB broth to the cells and incubation at 37°C, 150rpm for 3 hours to allow excision and packaging of the pBluescript plasmid from the λ cloning.
- Further incubation for 20 minutes at 67°C to lyse the *E. coli* XL-1 Blue and then centrifugation at 1000g for 15 minutes to remove the cell debris.
- The supernatant containing the excised, packaged plasmid was then transferred to a new tube and mixed with 200μL of *E. coli* SOLR, which is immune to λ-phage infection, prepared as above.
- The cells were incubated at 37°C for 15 minutes and 10μL of this transfected cell preparation was plated out on LB agar plates containing ampicillin (100μg/ml) and colony formation was allowed to occur overnight at 37°C.

Individual colonies were grown overnight in 3ml cultures of LB broth containing ampicillin (100μg/ml) at 37°C, 150 rpm. Plasmid DNA was purified from these cultures using the Promega kit (Section 2.9.3) and digested with 5U each of *Eco*RI and *Xho*I in 1X of the supplied buffer at 37°C for one hour to remove the cDNA inserts. The results of restriction digestion were visualised in 1% agarose gels (Section 2.5.3). Plasmid DNA was then sequenced (Section 2.6.2) using the vector-based primers T3 and T7 (Table 2.4) and sequence quality assessment and analysis was undertaken as described in Section 2.6.3.

2.12 IDENTIFICATION OF *PDI* SEQUENCES IN THE TIGR TaGI DATABASE

To further investigate whether a novel PDI gene is expressed in *T. aestivum*, The Institute for Genomic Research (TIGR) wheat gene index (TaGI, version 7.0) (The TIGR Gene Index Databases, The Institute for Genomic Research, Rockville, MD 20850 (URL: http://www.tigr.org/tdb/tgi)), containing consensus sequences generated from 492,694 publicly available ESTs (as at 8/2003), was searched. The TaGI database was queried with the amino acid sequences “LVEFYAPWCGHCKSLAPEY” and

"LIEFYAPWCGHCKKLAPIL**"** using a tBLASTn search. These peptide represent conserved sequences in the translation of the *wPDI1*, *wPDI2* and *wPDI3* cDNA sequences and were selected as they fit the thioredoxin-like active site motif (Section 1.5) outlined in the ExPASy PROSITE database (http://us.expasy.org/prosite/). Thus, other PDI family members would be expected to contain this or a similar motif in their sequence. Only tentative consensus (TC) sequences derived from at least 10 ESTs were selected for further analysis. Identification of similar sequences to the *PDI*-like TIGR TC sequences in the non-redundant (nr) protein databases was conducted using the BLASTx program (Section 2.6.3.5).

2.13 INVESTIGATION OF THE PUTATIVE *PDI* 1B GENE OF *T. aestivum*

The results of the Southern blots of *T. aestivum* cv. Chinese Spring (Section 3.6.1) were utilised to investigate the nature of the *PDI* gene reported to be present on chromosome 1B (Ciaffi *et al.*, 1999; Section 1.5.5). 10µg of gDNA each from *T. aestivum* cv. Chinese Spring and Wyuna was digested with 30U of *Eco*RI overnight at 37°C. The digest was electrophoresed on a 1% agarose gel and DNA from the 4.2-4.9kb section of the gel, containing the hybridising fragment, localised to the chromosome 1B (Section 1.5.5), was purified with the "Concert Rapid Gel Purification Kit" (Invitrogen) (Section 2.4.1). The primer pairs PRB5F/PRB5R, PRB3F/PRB3R and PRB5F/PDI1R (Table 2.5; Figure 2.2) and the PCR conditions described in Section 2.3.2 were used to amplify the *PDI* gene from this gel- purified DNA as three fragments, representing the gene section 5' to the 'conserved' *Eco*RI site (observed in our three cDNA sequences and the sequence data in Ciaffi *et al.,* 1999), 3' to it, or encompassing it, respectively. To test as to whether or not this site was conserved also in this tentatively isolated gene, 10µl aliquots of the PCR products generated with the PRB5F/PDI1R primer pair were digested with 5U of *Eco*RI for 2 hours at 37°C and the results visualised on 2% agarose gels (Section 2.5.3).

2.14 GENETIC MAPPING OF THE *PDI* GENES IN *T. aestivum*

2.14.1 Design of AS-PCR RFLP molecular markers for the *PDI* genes

The identified inter-cultivar single nucleotide polymorphisms (SNPs) in the *PDI* genes on chromosomes 4A and 4B (Chapter 4) were assessed for RFLPs using the BioEdit software package as described in Section 2.6.3.4. As outlined in Chapter 4, a RFLP for *Sml*I was identified in the PCR product generated with the PA3F/PA3R primer pair and a RFLP for *Bsu*36I was identified in the PCR product generated with the PB3F/PB3R primer pair. To confirm the genotype of each of the parental cultivars the relevant sections of the PDI genes were amplified with these aforementioned primer pairs, with the exception that the reverse primer PA3RB was used in place of PA3R (Figure 2.4), using appropriate PCR conditions for the individual primer pairs (Section 2.3.2). The PCR products generated with the PA3F/PA3RB and PB3F/PB3R primer pairs were digested with 2U each of *Sml*I or *Bsu*36I, respectively. Digestions were performed without prior purification of the PCR products in 10µl volumes in 1X of the supplied buffer at 37°C (*Bsu*36I) or 55°C (*Sml*I) for 4 hours. Aliquots of the digested PCR products were electrophoresed on 1.5% agarose gels as described in Section 2.5.3, to identify the genotype of each of the parental cultivars.

2.14.2 Scoring of markers and linkage analysis

Scoring of the markers in the 160 doubled haploid (DH) progeny of Cr x Hb, 180 DH progeny of CD x Ka and 180 DH progeny of Sc x Ta (Section 1.8.4.1) was conducted by PCR amplification of the markers with the PA3F/PA3RB primers in Cr x Hb, CD x Ka and Sc x Ta or the PB3F/PB3R primers in CD x Ka. These reactions were performed in a 96-well plate format in a PCR Sprint thermal cycler (ThermoHybaid) in 10µl final volumes as described in Section 2.3.2. Digestion of PCR products with *Sml*I (PA3F/PA3RB) or *Bsu*36I (PB3F/PB3R) and visualisation of markers was conducted as described in Section 2.14.1. Determination of linkage groups was assessed using the Map Manager QT software package (Manly and Olson, 1999) and genetic/comparative maps were assembled with cMAP (http://www.gmod.org/cmap/).

2.15 ORTHOLOGY BETWEEN THE RICE *esp2* AND THE WHEAT *PDI* LOCI

2.15.1 Identification of a rice BAC containing the putative *PDI* orthologue

The BAC containing the putative *PDI* gene of rice was identified by BLASTn search of the Gramene TIGR genome assembly 2004 available in the "Rice_genome_japonica_TIGR" database on the Gramene website (www.gramene.org; Ware *et al*. 2002) using the wheat *PDI* cDNA *wPDI1* as a query sequence. The structure of this putative rice *PDI* was then analysed by its alignment with the *PDI* gene from A genome of *T. turgidum* (Section 3.3.3), using the BioEdit software package (Hall, 1999).

2.15.2 Comparison of the genetic positions of the rice BAC and the *esp2* marker

An investigation to determine whether the *PDI* gene of rice is present at the *esp2* locus (Section 1.5.6) was conducted by search of the marker database on the Gramene website (www.gramene.org/cmap/feature_search, accessed 10/04) using the query "esp2" and limiting the species to "rice" and feature type to "phenotype", to identify the genetic position of the latter. The construction of comparative genetic/sequence maps of the putative *PDI* locus on the BAC of interest (Section 2.15.1) and the position of the *esp2* marker (from the Morph 2000 map, identified as described above) was then conducted using the CMap interface of the Gramene website (www.gramene.org/cmap/), as both were found to be located on the same chromosome (Section 5.2.2). However, due to this analysis showing no common markers between the two maps (Section 5.2.2), a third map, the JRGP Nipponbare/Kasalath RFLP 2000 (JRGP RFLP 2000) chromosome 11 genetic map (www.grameme.org) was then used for comparison of these two maps, as it contained markers common to the other two.

2.15.3 Identification of putative wheat *PDI* orthologues at the rice *PDI* locus and their physical loci in the wheat genome

The "Genome Browser" interface of the Gramene website (version 16; www.gramene.org; accessed 1/05), which contains data from the TIGR pseudomolecule assembly release 2 of the International Rice Genome Sequencing Project (IRGSP)

finished sequence (www.tigr.org) was used to investigate putative wheat orthologues at the *PDI* locus of rice. This analysis involved a multistep-process to: (i) identify the 1Mb region of the rice genome encompassing the *PDI* gene (by centring the genome browser on the rice PDI gene on chromosome 11 and zooming to view a 1 Mb section); (ii) identify the wheat tentative consensus (TC) sequences in the TIGR TAGI database found to have significant sequence homology to this region of the rice genome, by selecting the "Wheat_ESTCluster_TGI" feature track; and (iii) determine whether any of these putatively orthologous TC sequences were assembled with ESTs that had been physically mapped to a wheat chromosomal 'bin', by comparing the ESTs used to assemble these TC sequences (www.tigr.org) with those sequenced from probes used in wheat physical mapping experiments available from the wEST database (wheat.pw.usda.gov/wEST/). As some of this mapping data appears to not have been evaluated rigorously (as per the website), it was visually assessed for accuracy.

2.15.4 Physical mapping of *PDI* genes to chromosomal 'bins'

As discrepancies were identified for the mapping data for the *PDI* genes of wheat in the GrainGenes-SQL database (Section 5.3.2) the *PDI* genes of wheat were physically mapped using an AS-PCR approach to clarify their chromosomal loci. The physical chromosomal loci of the *PDI* genes of wheat have previously been reported to be on chromosomes 4AL, 4BS, 4DS and 1BS (Section 1.5.5). To further refine the chromosomal loci of the genes encoding the cDNA clones *wPDI1*, *wPDI2* and *wPDI3*, AS-primers described above (Section 2.10) were used to assay aneuploid lines of *T. aestivum* cv. Chinese Spring, including nullisomic/tetrasomic lines, ditelosomic lines and deletion stocks, to permit the localization of each of these genes to a section of a chromosome arm, or chromosomal 'bin'.

Allele-specific (AS) PCR was utilised to physically map the *PDI* genes with the nullisomic/tetrasomic, ditelosomic and deletion lines of *T. aestivum* cv. Chinese Spring. The primers used for discriminating the 3 genes in the physical mapping experiments were PA1F/PA1R, PB3F/PB3R and PD5F/PD5R (Figure 2.4; Table 2.6). Initial experiments using gDNA isolated from 42 nullisomic/tetrasomic lines were conducted to determine the chromosomal location of each *PDI* gene. Absence of a PCR product from two lines nullisomic for a particular chromosome was interpreted as a positive

identification of the chromosomal location of each gene. Repetition of the experiment on the nullisomic/tetrasomic lines and the inclusion gDNA from the ditelosomic and deletion lines from the identified chromosome were undertaken to provide finer mapping details.

PCR reactions were performed with the HotStarTaq DNA Polymerase kit (Qiagen) in 10µl total reaction volumes containing 1x reaction buffer, 1.5 mM $MgCl_2$, 200 µM dNTPs, 0.5 µM of each primer, 50ng template DNA and 1U of HotStarTaq. Thermal cycling was conducted in a PCR Sprint thermal cycler (ThermoHybaid) and consisted of 5 minutes at 95°C followed by 35 cycles of 30 seconds at 94°C, 30 seconds at annealing temperature (Table 2.5) and 1 minute at 72°C. A final extension step at 72°C was conducted for 7 minutes, followed by storage at 4°C. All reactions included a negative control containing no template DNA. PCR products were visualised by electrophoresis in 1.2% agarose gels as described in Section 2.5.3.

2.16 CHARACTERISATION OF *CYCLOPHILIN A* GENES

As the literature search indicated that the genomic organization and intron/structure of *cyclophilin* genes in wheat was unknown, it was undertaken to design allele-specific primers to specifically amplify these genes from the gDNA of cultivars used previously in the establishment of genetic maps in wheat (Section 1.8.4.1). This approach would serve a two-fold purpose, in revealing the intron/exon structure of the *cyclophilin* genes and also identifying any inter-cultivar polymorphism that could be exploited later as molecular markers for the genetic analysis of the *cyclophilin* gene family.

2.16.1 Design of allele-specific primers

The literature review indicated that information on the *cyclophilin* genes in wheat was limited to three *cyclophilin* cDNA sequences, *wCyp1*, *wCyp2* and *wCyp3*, isolated by us earlier (J. Johnson, Honours Thesis, VU, 1999; Johnson *et al.*, 2001; Section 1.5.5); these were used to design allele-specific primers to individually amplify each of the homeoalleles from the gDNA. As described for the *PDI* AS-primers (Section 2.10.1), the *cyclophilin* cDNA sequences were aligned and primer sequences were designed

such that the 3' terminal nucleotide in the forward and reverse primer was unique to the homeoallele to be amplified (Figure 2.3). Details of these primers, their relative sizes and positions on the cDNA sequences are outlined in Table 2.7 and Figure 2.5.

2.16.2 Confirming the allele-specificity of primers

A similar method was used to confirm the specificity of the AS-primers as that used for the *PDI* AS-primers (Section 2.10.2), specifically, the amplification of the corresponding sections from the cDNA clones *wCyp1*, *wCyp2* and *wCyp3* to test whether or not a product was only amplified from the target clone. PCR conditions for these amplifications were as described above (Section 2.3.2) and further confirmation of the primer specificity of the was achieved through analysis of the sequencing data of these PCR products ensuring that unambiguous data was obtained, as for the *PDI* primers (Section 2.10.2).

2.16.3 Amplification and sequencing of partial *cyclophilin A* genes

The AS-primers described in Section 2.16.1 were used to amplify the *cyclophilin* genes from the gDNA of *T. aestivum* cvs. Cranbrook, Halberd, Egret, Sunstar, Sunco, Tasman, CD87 and Katepwa. PCR conditions were as described in Section 2.3.2, which, in the case of the C2F1/C2R1 primer pair, led to a 2-step PCR, as the ideal annealing temperature was found to be 72°C (Table 2.7), with a cycle consisting of denaturation at 94°C for 30 seconds and annealing/extension at 72°C for 2.5 minutes. The products were checked by agarose gel electrophoresis (Section 2.5.3), the PCR products were purified with the Promega kit (Section 2.4.3) and purified PCR products were directly sequenced with the primers used to amplify each according to the reaction conditions outlined in Section 2.6.2. The sequences generated from different cultivars were aligned and compared as described in Section 2.6.3.

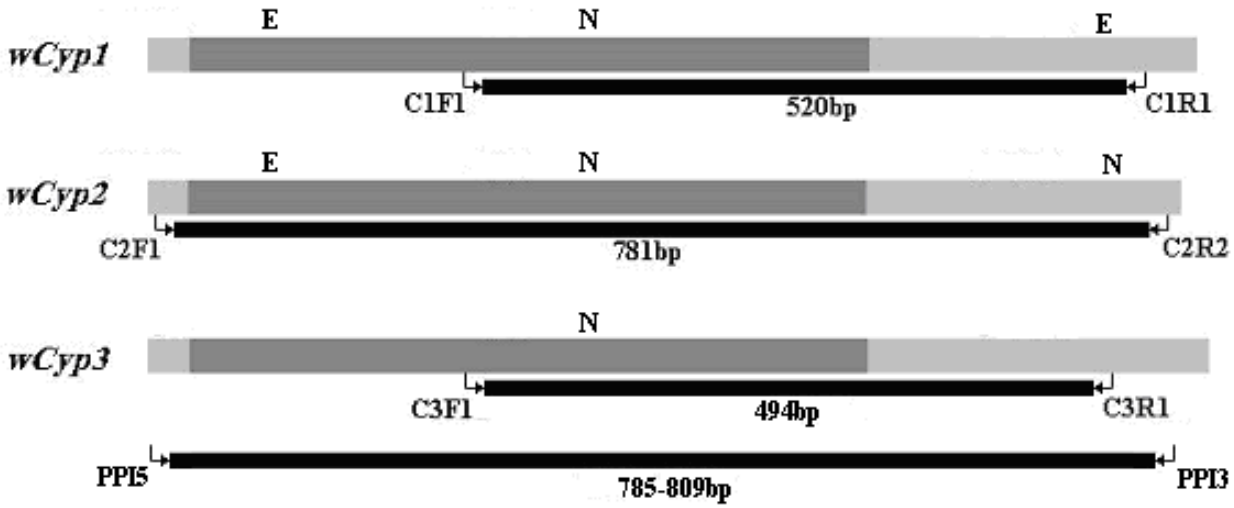


Figure 2.5 Primer locations in the *cyclophilin* genes of wheat.

Positions of the allele-specific and consensus primer pairs used in this study and expected sizes of the PCR products after amplification of the previously characterized cDNA sequences (Johnson *et al.*, 2001). The 5' and 3' untranslated regions are shown in light grey, preceding and following the coding sequences (in dark grey), respectively. The *Eag*I (E) and *Nco*I (N) restriction sites characteristic of these clones are shown.

Table 2.7 Primers used to characterise the *cyclophilin* genes of wheat

Primer and its 5'→3' sequence[a]	Annealing temperatures Calculated/Actual (°C)	Expected Length of PCR product (bp)[b]	Purpose
C1F1: AGGGCGGCGACTTCACCC C1R1: CACAGGAAGATAACCAAGGCG	61/70	520	Specific for *wCyp1* sequence class of genes
C2F1: ATCTCAAAACCCTAGCCCG C2R1: GGAGTTCATCTCATTCAAACA	52/64	781	Specific for *wCyp2* sequence class of genes
C3F1: GGGCGGCGACTTCACCAG C3R1: AACCAAGGCCGCCGGACG	64/72	494	Specific for *wCyp3* sequence class of genes
PPI5: CCCATCTCAAAACCCTAGCCC PPI3: CGGAGTTCATCTCATTCAAAC	54/52	785 to 809	Consensus for all wheat *cyclophilin A*'s

[a] Nucleotides unique to the target gene are highlighted in yellow.
[b] Sizes are based on cDNA products and gene copies with no introns.

2.17 PHYSICAL MAPPING OF THE *CYCLOPHILIN A* GENES OF *T. aestivum*

The gDNAs of all (42) nullisomic/tetrasomic were used to determine the chromosomal loci of each of the *cyclophilin A* homeoalleles using the allele-specific primers C1F1/C1R1, C2F1/C2R1 and C3F1/C3R1 (Figure 2.5; Table 2.7). A reproducible absence of the PCR product in a certain reaction was interpreted as absence of that particular homeoallele in that particular aneuploid line, and therefore, an indication of its chromosomal locus. This was followed by amplification of the gDNAs from a subset of ditelosomic and deletion lines of *T. aestivum* cv. Chinese Spring (Table 2.3) to further refine their loci to chromosomal 'bins'. The PCRs were conducted in 10µL volumes using conditions described in Section 2.3.2, except for use of the HotStar *Taq* Kit (Qiagen), and the products were analysed on 1% agarose gels (Section 2.5.3).

2.18 THE COPY NUMBER OF *CYCLOPHILIN A* GENES IN *T. aestivum*

2.18.1 Southern blot of *T. aestivum* gDNA

As *cyclophilins* have been reported to be members of small multigene families in other plant species (Section 1.6.3) it was undertaken to determine the size and diversity of the *cyclophilin A* gene family in *T. aestivum* cv. Chinese Spring. Purified gDNA (10μg) from *T. aestivum* (cv. Chinese Spring), was restriction digested with 30U of *Bam*HI or *Hind*III for 12 hours at 37°C, separated by electrophoresis and transferred to Hybond XL membrane according to the methods outlined in Section 2.7.2.

2.18.2 Probing of the Southern blot with a *cyclophilin A* probe

Probing of the Southern blots was conducted with an 796bp PCR product generated from the *wCyp3* cDNA clone (Johnson *et al.*, 2001) by amplification with the PPI5 and PPI3 primers (Table 2.7) using the PCR conditions described in Section 2.14 and an annealing temperature of 52°C. The probe was radiolabeled as described in Section 2.7.3.2, hybridised to the Southern blot (Section 2.7.3.3), followed by washing of the blots and detection of hybridised probe using the Fuji phosphorimager according to the methods outlined in Section 2.7.3.4.

2.18.3 Isolation of genome-specific *cyclophilin A* genes

The results of the Southern blots were used to characterise the *cyclophilin A* genes originating from the A, B or D genomes of wheat. Using these results (Section 6.4) a *Hind*III digest of 10μg of gDNA of *T. aestivum* cv. Chinese Spring was then separated on a 1% agarose gel and sections of the gel corresponding to the four major bands seen in the *Hind*III digest (Section 6.4) were cut out. DNA was purified from each of these using the Wizard® SV Gel and PCR Clean-Up System (Promega) (Section 2.4.3) and used as template to amplify any *cyclophilin A* genes contained therein, using the consensus primers PPI5 and PPI3 (Table 2.7) and the PCR conditions given in Section 2.3.2, with an annealing temperature of 52°C. The identity of the amplified genes was determined by testing for restriction fragment length polymorphisms (RFLPs) for the

enzymes *Eag*I and *Nco*I as predicted from the three cDNA sequences of *cyclophilin A* (Figure 2.5). Aliquots of the PCR products were digested with 5U each of *Eag*I, *Nco*I or both enzymes for 2 hours at 37°C and analysed by electrophoresis in 1.4% agarose gels as described in Section 2.5.3.

2.19 IDENTIFICATION OF NOVEL WHEAT *CYCLOPHILIN* GENES

2.19.1 Characterisation of *cyclophilin* genes from an endosperm cDNA library

The sequence database derived from cDNA sequences of 8-12 DPA endosperm tissue from cv. Wyuna (Clarke *et al.*, 2000) was searched for ESTs with a most significant BLASTn hit containing the term "cyclophilin", to identify any novel *cyclophilin*-like sequences. Corresponding candidate cDNA clones were excised as plasmids, as described above (Section 2.11.4) and purified with the Wizard® Plus SV Minipreps DNA Purification System (Promega) (Section 2.9.3). Plasmids were digested with *Xho*I and *Eco*RI in 1X of the supplied buffer for 1 hour at 37°C to excise the cDNA insert. Aliquots of digested plasmids were analysed by electrophoresis on 1% agarose gels as described in Section 2.5.3.

Plasmids were initially sequenced with the vector-based primers T3 and T7 (Table 2.4) before elucidating the sequence of both strands of each clone by 'primer-walking' down each as described in Section 2.6.2.

2.19.2 Screening of the TIGR wheat gene index

The Institute for Genomic Research (TIGR) wheat gene index (TaGI, version 6.0) (The TIGR Gene Index Databases, The Institute for Genomic Research, Rockville, MD 20850; URL: http://www.tigr.org/tdb/tgi), containing 413,955 ESTs (as at 1/2003), was searched using BLASTn and tBLASTx with *wCyp1* as the query sequence. Only the candidate contigs containing at least 10 ESTs were evaluated further.

2.19.3 Analysis of novel *cyclophilin* gene sequences

Novel *cyclophilin* sequences identified in the preceding sections (Sections 2.19.1 and 2.19.2) were analysed in the BioEdit software package and through the use of specific software including the identification of homologous sequences by BLAST analyses, open reading frames were identified and putative proteins analysed for physical properties and putative subcellular localisation according to the methods described above (Section 2.6.3).

CHAPTER 3

MOLECULAR CHARACTERISATION OF THE PROTEIN DISULFIDE ISOMERASE GENES OF WHEAT

List of Figures and Tables in Chapter 3

Chapter 3 Molecular characterisation of the protein disulfide isomerase genes of wheat

ABSTRACT

The protein disulfide isomerases (PDI), enzymes that catalyze the formation, cleavage and isomerisation of disulfide bonds, are suggested to be involved in regulating the folding and deposition of storage protein bodies in the wheat endosperm, thus potentially playing an important role in influencing grain quality. However, limited information exists on the *PDI* genes in wheat and their possible roles in this process. Via innovative combinations of various experimental approaches such as exploitation of sequence variations between alleles, RFLPs in cDNAs and genomic copies, and direct isolation of gene fragments, we have characterised the individual PDI genes from the common wheat, *T. aestivum*, the diploid progenitor of its D genome, *Ae. tauschii*, and the tetraploid progenitor of its A and B genomes, *T. turgidum*. *Ae. tauschii*, durum wheat and common wheat exhibit one, two and three *PDI* gene(s), respectively, and the additional *PDI* gene suggested to reside on chromosome 1B, if at all present, appears to be a partial, nonexpressed copy. All genes consist of ten exons and nine introns, with the *Ae. tauschii* PDI gene showing near complete identity to the corresponding one in common wheat but more limited identity to the A and B genome PDI genes of common wheat and *T. turgidum,* and the two genes of *T. turgidum* showing higher degrees of conservation with their counterparts in common wheat rather than between themselves.

3.1 INTRODUCTION

As discussed in Section 1.5.6 investigations into protein disulfide isomerase (PDI) and it roles in the developing wheat endosperm are of interest due to the potential role for this enzyme in regulating the folding and deposition of the storage proteins of wheat. As these proteins are critical in conferring the visco-elastic properties unique to wheat dough PDI may play an important role in determining quality characteristics of wheat cultivars. Despite such potential importance, only limited evidence is available to support such a role, based on the upregulation of *PDI* genes in the wheat endosperm and abberant storage protein deposition in a rice mutant lacking *PDI* expression (Section 1.5.6). Therefore, there is a need to develop other strategies to investigate the involvement of these enzymes in these processes. One such strategy is the development of molecular markers for the *PDI* genes, which may provide a means for testing for any genetic association between the *PDI* genes and QTLs involved in dough processing traits and a method for selecting for beneficial alleles in breeding programs. However, firstly a better understanding of this gene family in hexaploid wheat, both the number of genes present and the sequence of such genes, was required.

Our current knowledge of the *PDI* gene family in hexaploid wheat and its progenitor species was summarised in Section 1.5. Briefly, Ciaffi *et al.* (1999) reported two *PDI* hybridising bands in some cultivars of *T. turgidum*, while some cultivars were found to contain a third polymorphic fragment. In contrast, RFLP analysis revealed the genome of *Ae. tauschii* contains a single, highly conserved hybridising fragment. Current sequence data on the *PDI* genes of *T. turgidum* is limited to partial gene sequences from chromosome 4A and one of the two genes localised to the B genome of *T. turgidum* and no sequence data is currently available for the *PDI* gene(s) of *Ae. tauschii*. In *T. aestivum* our knowledge of the *PDI* gene family is limited to the presence of three hybridising bands on a Southern blot, localised to chromosomes 4AL, 4BS and 4DS and a fourth hybridising band in some cultivars localised to chromosome 1B, and we have reported the characterisation of three *PDI* cDNAs isolated from a *T. aestivum* cv. Wyuna developing endosperm library (Section 1.5.5). Thus, we have used the conserved sequences in these *T. aestivum* cDNAs to design three pairs of consensus primers to allow the amplification and cloning of the uncharacterised 3' sections of the two *PDI* genes of *T. turgidum* (*TtPDI4A* and *TtPDIB*) and the entire *PDI* gene from *Ae.*

tauschii (*AetPDI4D*). Further, the sequence data generated from the *PDI* gene families in *T. turgidum* and *Ae. tauschii* was used to design allele-specific primers (Section 2.10.1) to permit the targeted amplification and characterisation of the corresponding homeoalleles in hexaploid wheat.

3.2 COPY NUMBER OF *PDI* GENES IN *T. aestivum* AND ITS PROGENITORS

In order to determine the copy number of the *PDI* gene families in *T. aestivum* and its diploid (*Ae. tauschii)* and tetraploid (*T. turgidum* subsp. *durum*) progenitors, Southern blot analyses were performed as described in Section 2.8.

3.2.2 Southern blot of *PDI* genes in *T. aestivum* cv. Wyuna

The Southern blot of *T. aestivum* cv. Wyuna probed with the 838bp *PDI* cDNA fragment, revealed three strongly hybridising fragments in each of the single digests (Figure 3.1A, Table 3.1) with the estimated sizes in the *Eco*RI (9.3kb, 5.3kb and 1.9kb) and *Bam*HI (8.9kb, 5.4kb and 4.1kb) digests being similar to those reported (Section 1.5.5; Ciaffi *et al.* 1999). The estimated sizes of the bands in the *Hind*III digest were consistently slightly larger (5.1kb, 4.3kb and 2.9kb) than the corresponding reported sizes (4.5kb, 3.7kb and 2.5kb) and the fourth band, localised to chromosome 1B (Section 1.5.5), was absent in all digests (Table 3.1). Three strongly hybridising bands were also observed in the *Eco*RI/*Bam*HI (5.2kb, 5.0kb and 1.9kb), *Bam*HI/*Hind*III (4.3kb, 3.4kb and 2.9kb) and *Eco*RI/*Hind*III (4.1kb, 2.7kb and 1.8kb) double digests, in all cases their sizes being either smaller than (as would be expected in some cases, due to internal restriction sites), or the same as, those in the corresponding single digests, and some lighter bands. It was thus not possible to identify a major band as originating specifically from the reported gene on chromosome 1B in these digests too. The fragments sizes for all digests were utilized in conjunction with the gene sequencing results (Sections 3.3.3 and 3.4) to construct restriction maps for the individual *PDI* genes of the hexaploid wheat (Figure 3.6).

3.2.2 Southern blot of *PDI* genes in *Ae. tauschii* and *T. turgidum*

The Southern blot of *T. turgidum* gDNA probed with the 838bp cDNA fragment revealed two strongly hybridising fragments approximately 4.9kb and 1.8kb in the *Eco*RI digest, 6.7kb and 3.3kb in the *Bam*HI digest and 4.0 and 3.4kb in the *Hind*III digest (Figure 3.1B), similar to the *PDI* genes reported from chromosomes 4A and 4B of *T. turgidum* and *T. aestivum* cv. Chinese Spring (Table 3.1). Likewise, the double digests contained two strongly hybridising bands in each case, approximately 4.5kb and 1.8kb in the *Eco*RI/*Bam*HI digest, 3.4kb and 1.7kb in the *Eco*RI/*Hind*III digest and 3.4kb and 2.9kb in the *Bam*HI/*Hind*III digest. Thus, the double digests contained fragments equal in size to, or smaller than, those observed in the single digests, as would be expected.

The *Ae. tauschii* Southern blot probed with the 838bp cDNA fragment revealed single, intensely hybridising fragments, as well as one or two weakly hybridising fragments. The sizes of the strongly hybridising fragments were approximately 6.6kb, 5kb and 2.7kb in the *Eco*RI, *Bam*HI and *Hind*III digests, respectively (Figure 3.1C, Table 3.1), compared to the reported 9.3kb, 5.5kb and 2.5kb fragments from chromosome 4D in *T. aestivum* cv. Chinese Spring, for the respective digests (Table 3.1). The *Eco*RI/*Bam*HI, *Eco*RI/*Hind*III and *Bam*HI/*Hind*III double digests yielded single intensely hybridising fragments of approximately 5.5kb, 2.3kb and 2.5kb, again, similar in size to, or smaller than, those in the corresponding single digests.

These Southern blot results were utilised in conjunction with the respective sequences for these genes (Sections 3.3.3) to construct the respective restriction maps for the *PDI* genes in these two genotypes (Figure 3.6).

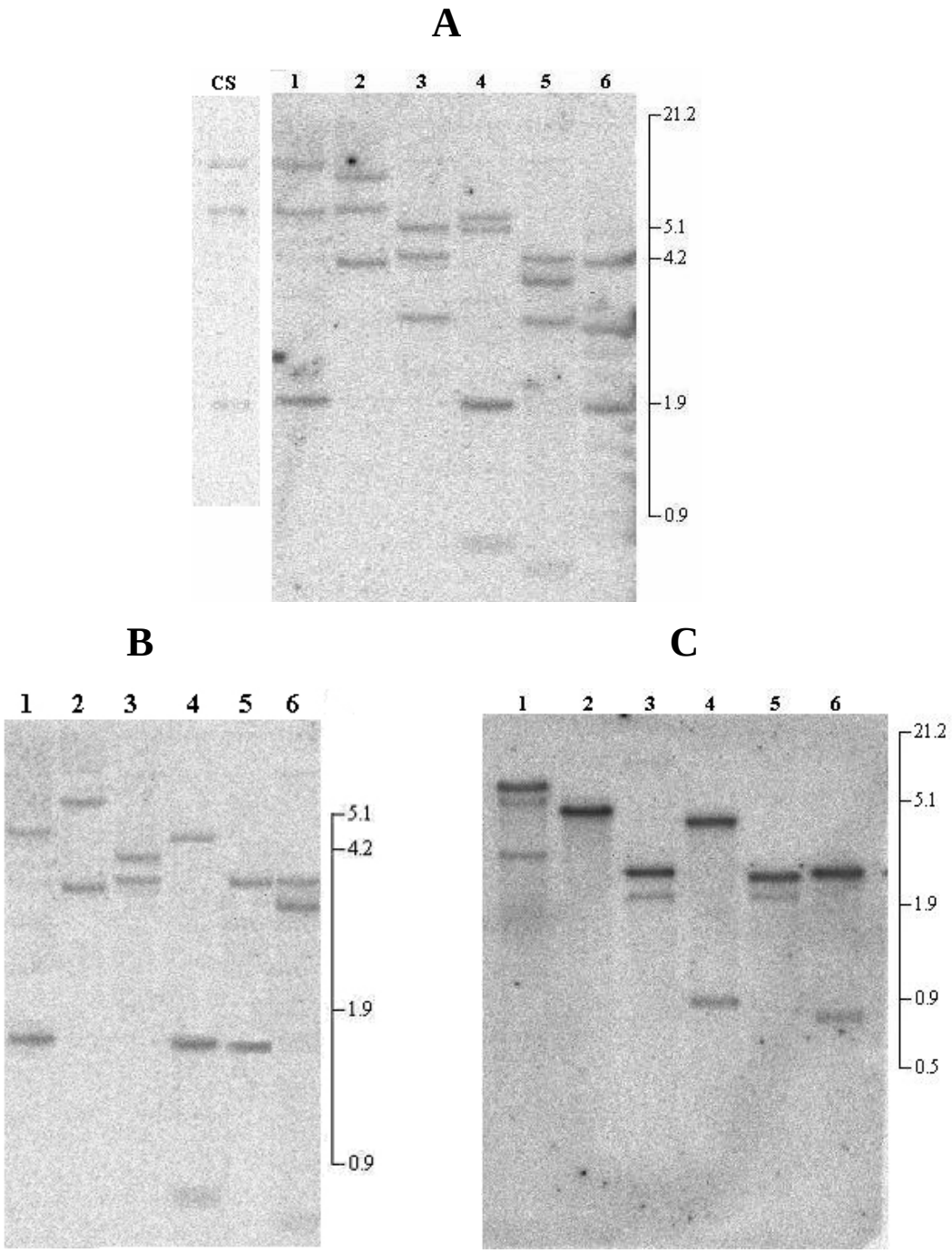


Figure 3.1 *PDI* Southern blots of *T. aestivum, T. turgidum* and *Ae. tauschii*
The *PDI* Southern blots of *T. aestivum* cv. Wyuna (A) *T. turgidum* (B) and *Ae. tauschii* (C). gDNA was digested with *Eco*RI (1), *Bam*HI (2), *Hind*III (3), *Eco*RI/*Bam*HI (4), *Bam*HI/*Hind*III (5) and *Eco*RI/*Hind*III (6) and probed with the 838bp cDNA fragment from *wPDI1* (Figure 2.2). The Southern blot of *Eco*RI digested *T. aestivum* cv. Chinese Spring (CS) gDNA is also shown (A). The molecular weight marker (kb) is given on the right.

Table 3.1 Sizes of *PDI* hybridising fragments (kb) on Southern blots

	Ciaffi *et al*. (1999)		838bp cDNA probe			1466 bp cDNA probe
	T. aestivum cv. Chinese Spring[a]	*T. turgidum*	*T. aestivum*[b] (AABBDD)	*T. turgidum* (AABB)	*Ae. tauschii* (DD)	*T. aestivum* cvs. Wyuna and Chinese Spring
*Eco*RI	9.3 (4DS) 5.3 (4AL) 4.6 (1BS)[c] 1.8 (4BS)	5.3 4.6[c] 1.8	9.3 5.3 1.9	4.9 1.8	6.6	9.3 5.7 5.3 4.4 3.6 1.9
*Bam*HI	23.0 (1BS) 8.9 (4AL) 5.5 (4DS) 4.2 (4BS)	Not reported	8.9 5.4 4.1	6.7 3.3	5.0	Not conducted
*Hind*III	23.0 (1BS) 4.5 (4BS) 3.7 (4AL) 2.5 (4DS)	Not reported	5.1 4.3 2.9	4.0 3.4	2.7	Not conducted

[a] Chromosome location of each hybridising fragment shown as described by Ciaffi *et al*. (1999).
[b] The cultivars analysed were Wyuna and Chinese Spring. The *Bam*HI and *Hind*III data listed was from Wyuna only.
[c] This fragment has been reported to be highly polymorphic, varying in presence/absence and size, in different genotypes (Ciaffi *et al.* 1999).

3.3 CHARACTERISATION OF *PDI* GENES IN *T. turgidum* AND *Ae. tauschii*

3.3.1 PCR products of *PDI* genes from *T. turgidum* and *Ae. tauschii*

Two partial *PDI* gene sequences, a 3476bp section of the gene localised to chromosome 4A and a 2721bp section of the gene localised to the B genome (4B or 1B) were reported in *T. turgidum* while this study was underway (Ciaffi *et al.* 2001). Thus, to utilise this data, minimise duplication and extend the sequence information further the 3' sections of the *PDI* genes of *T. turgidum* containing uncharacterised sections were amplified using the consensus primers (PDI3F/PDI3R) flanking 759-766bp of the *PDI* cDNA sequences (Table 2.5, Figure 2.2), resulting in products of approximately 1.1kb, approximately 350bp larger than that predicted from the cDNA sequences (Figure 3.2).

As no sequence data for the *PDI* gene of *Ae. tauschii* was available, the gene encompassing the entire coding sequence and any intervening introns was amplified as

three overlapping fragments using the primers PDI1F/PDI1R, PDI2F/PDI2R and PDI3F/PDI3R (Table 2.5, Figure 2.2), resulting in products of approximately 1.2, 1.6 and 1.1kb, respectively, again, larger than the sizes predicted from the cDNA sequences (Figure 3.2).

3.3.2 Cloning of the *PDI* genes of *T. turgidum* and *Ae. tauschii*

The *PDI* Southern blot data for *T. turgidum* generated previously (Section 1.5.5) and above (Section 3.2.2) suggested the presence of at least two *PDI* genes in this species. To ensure that all members of the *PDI* genes were isolated the PCR products generated from the gDNA of *T. turgidum* and *Ae. tauschii* (Section 3.3.1) were cloned into pGEM-T-Easy (Sections 2.9.2). Six recombinant (white) colonies were chosen at random for each cloned PCR product and their plasmids were purified and digested with *Eco*RI (Sections 2.9.3) revealing appropriate sized inserts in all cases (Figure 3.3) and a potential RFLP for *Eco*RI in the cloned *T. turgidum* PCR products amplified with PD3F/PD3R (Figure 3.3D).

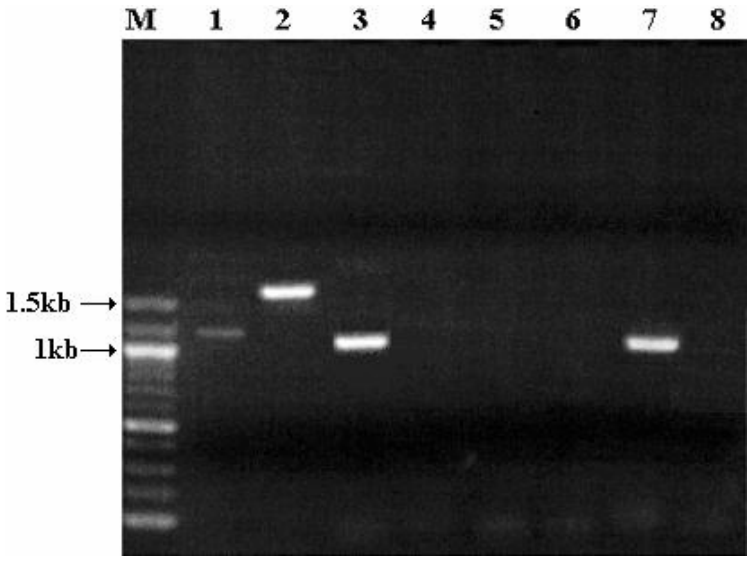


Figure 3.2 PCR products of the *PDI* genes from *Ae. tauschii* and *T. turgidum*
The *PDI* PCR products amplified from the gDNA of *Ae. tauschii* (1: PD1F/PD1R; 2: PD2F/PD2R; 3: PDI3F/PD3R) and their corresponding water-only negative controls (4: PD1F/PD1R; 5: PD2F/PD2R; 6: PD3F/PD3R). The PCR product encompassing the 3'end of the *PDI* gene of *T. turgidum* (7: PD3F/PD3R) and the corresponding water-only negative control (8: PD3F/PD3R). M: 100bp molecular weight marker.

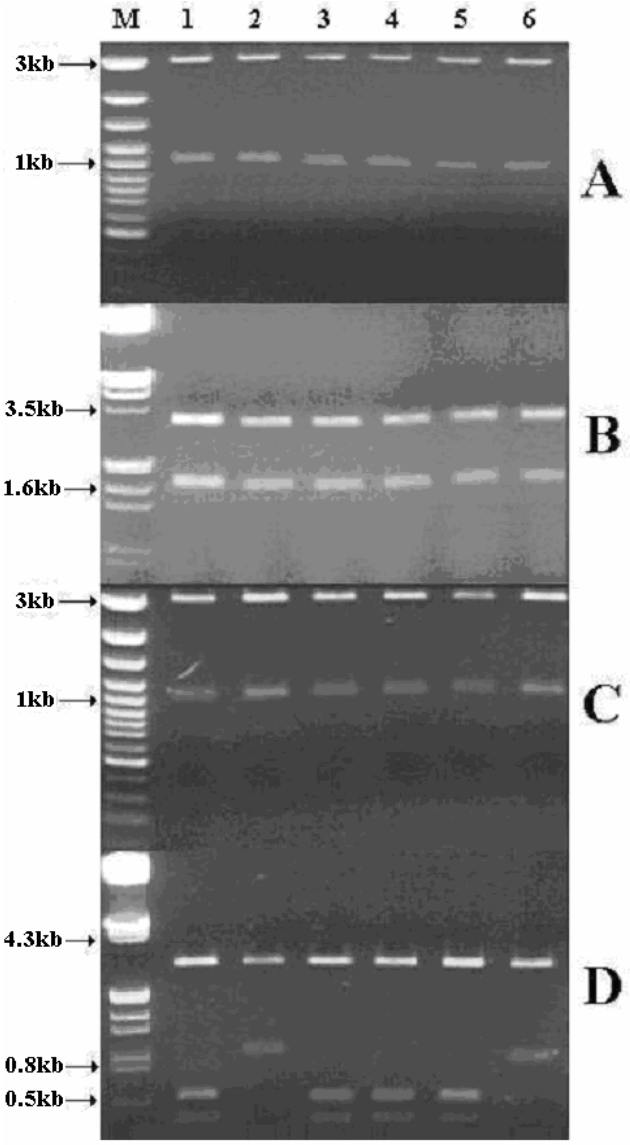


Figure 3.3 The *Ae. tauschii* and *T. turgidum PDI* clones

The *Eco*RI digested clones (A) PD1-1, PD1-2, PD1-3, PD1-4, PD1-5 and PD1-6 (Lanes 1-6, respectively) containing the cloned PD1F/PD1R PCR product from *Ae. tauschii*; (B) PD2-1, PD2-2, PD2-3, PD2-4, PD2-5 and PD2-6 (Lanes 1-6, respectively) containing the cloned PD2F/PD2R PCR product from *Ae. tauschii*; (C) PD3-1, PD3-2, PD3-3, PD3-4, PD3-5 and PD3-6 (Lanes 1-6, respectively) containing the cloned PD3F/PD3R PCR product from *Ae. tauschii* and (D) PAB3-1, PAB3-2, PAB3-3, PAB3-4, PAB3-5 and PAB3-6 (Lanes 1-6, respectively) containg the cloned PD3F/PD3R PCR product from *T. turgidum*. M: molecular weight markers, 100bp ladder (A and C) or λ *Eco*RI/*Hind*III (B and D).

3.3.3 Sequencing and analysis of the *PDI* genes of *T. turgidum* and *Ae. tauschii*

Partial sequencing of the six *T. turgidum* clones revealed two distinct sequence classes (partial sequence data not shown); two clones (PAB3-2 and PAB3-6) contained sequences overlapping the 3' end of the reported *PDI* gene on chromosome 4A and four (PAB3-1, PAB3-3, PAB3-4 and PAB3-5) contained sequences overlapping the 3' end of the reported *PDI* gene from the B genome of *T. turgidum* (Ciaffi *et al.* 2001). Clones PAB3-1 and PAB3-2 were selected as representative clones from each class and were sequenced in full.

Clone PAB3-2 had an insert of 1100bp (Genbank accession# AY544170), comprised of 1014bp that was 100% identical to the 3' section of the previously reported *PDI* gene from chromosome 4A of *T. turgidum* (Ciaffi *et al.* 2001), and an additional 3' 86bp. This allowed the construction of a 3562bp contig for the *PDI* gene of the A genome of *T. turgidum*, (*TtPDI4A*) (Figures 3.6 and 3.7). Alignment of *TtPDI4A* with the three *PDI* cDNA clones isolated from *T. aestivum* (Section 1.5.5) revealed a 10 exon structure (Figure 3.6) and highest exon identity (99.7%) with the *wPDI1* cDNA sequence (Table 3.2), the differences consisting of 4 SNPs, 1 each in exons 1, 5, 6 and 10, all of which were silent, three being at the wobble position of codons and the fourth being in the 3' UTR.

Clone PAB3-1 had an insert of 1061bp (Genbank accession# AY544171) of which 255bp was identical to the 3' end of the previously reported partial *PDI* gene from either chromosome 4B or 1B of *T. turgidum* (Ciaffi *et al.* 2001) and an additional 3' 806bp of data. The published data and our data together allowed the creation of a 3527bp contig of the B genome *PDI* gene (*TtPDIB*) (Figures 3.6 and 3.7). Alignment of *TtPDIB* with the *PDI* cDNAs (Section 1.5.5) revealed that the extra 806bp of sequence generated in this study encompassed exons 7 through 10 (Figure 3.6). The exon sequences of *TtPDIB* were the most similar to the *wPDI2* cDNA sequence, with 99.8% identity, the differences consisting of 2 SNPs in exon 1, one of which was a silent mutation and the other would lead to a Cys4→Ser4 mutation in the putative protein encoded by *TtPDIB*. Interestingly, while all of the wheat *PDI* genes and cDNAs characterised thus far (current study, Section 1.5.5) were found to contain a conserved

*Eco*RI site in exon 4, both *TtPDIB* and *wPDI2* contained a second *Eco*RI site in exon 9, approximately 1830bp downstream from the first one.

Partial sequencing of the six clones from each of the three PCR reactions with the primer pairs used for the initial amplifications, spanning the length of the *PDI* gene of *Ae. tauschii* (Figure 2.2), revealed identical sequences within each class of clones (partial sequence data not shown); thus the clones PD1-1, PD2-1 and PD3-1 were selected as representative clones from each type of PCR and sequenced in full. The lengths of the cloned sequences were determined to be 1248bp, 1612bp and 1072bp, respectively. Their overlapping sections were identified by multiple sequence alignment to produce a single 3526bp contig (Genbank accession# AY544169) for the *PDI* gene of *Ae. tauschii* (*AetPDI4D*) (Figure 3.6 and 3.7). A comparison of *AetPDI4D* with the wheat cDNA sequences revealed the exon sequences of *AetPDI4D* exhibited the highest sequence identity (99.8%) to the *wPDI3* cDNA sequence (Table 3.2), differing only at two positions in the 3' untranslated region. This comparison also revealed that this gene was distinct from the two *PDI* genes of *T. turgidum* (Table 3.2).

Table 3.2 Conservation of *PDI* genes and transcripts of wheat

wPDI1	*wPDI2*	*wPDI3*	*TtPDI4A*[2]	*TtPDIB*[2]	*AetPDI4D*	
100[1] n/a[1]	95.7 n/a	96.7 n/a	99.7 n/a	95.8 n/a	96.7 n/a	*wPDI1*
	100 n/a	96.1 n/a	95.7 n/a	99.8 n/a	96.1 n/a	*wPDI2*
		100 n/a	96.7 n/a	96.1 n/a	99.8 n/a	*wPDI3*
			100 100	95.8 85.4	96.7 80.5	*TtPDI4A*[2]
				100 100	96.1 83.9	*TtPDIB*[2]
					100 100	*AetPDI4D*

[1] Sequence identities are shown for exon sequences (upper number %) and intron sequences (lower number %).
[2] These sequences represent a contig of the reported sequences (Ciaffi *et al.* 2001) with the additional 3' sequence generated in the current study.

3.4 CHARACTERISATION OF *PDI* GENES IN *T. aestivum*

The sequence data of *TtPDI4A*, *TtPDI4B* and *AetPDI4D* (Section 3.3) were used to identify single nucleotide polymorphisms that were unique to each of these genes, which were subsequently exploited to design four pairs of allele-specific primers for the A genome, three pairs for the B genome and six pairs for the D genome (Table 2.6) to allow the allele-specific amplification and characterisation of the corresponding genes in *T. aestivum* cv. Katepwa. This cultivar was selected for sequencing as the data generated would prove useful in the development of molecular markers for these genes (Chapter 4), as it was one of the parental cultivars previously used in the generation of genetic maps in wheat (Section 1.8.4.1)

To ensure the allele-specific (AS) primers were specific, those primer pairs solely based on cDNA sequences were used to amplify all three *PDI* cDNA clones, revealing amplification only of the appropriate cDNA clone when the optimised annealing temperatures were used (Figure 3.4). The allele specificity of primers based on intron sequences of the genes of *T. turgidum* or *Ae. tauschii* was confirmed by analysing the sequencing results to assess whether or not they contained a single sequence type, containing no ambiguity (Sections 3.4.1, 3.4.2 and 3.4.3).

3.4.1 The *PDI* gene in the A genome of *T. aestivum*

Amplification of sections of the *PDI* gene of *T. aestivum* cv. Katepwa corresponding to the *TaPDI4A* gene with four pairs of allele-specific primers (Table 2.6, Figure 2.4) produced products similar in size to those predicted from the *TtPDI4A* gene (Figure 3.5A). Direct sequencing of these products allowed the assembly of two contiguous regions of 1391bp (GenBank accession# AY544172) and 458bp (Genbank accession# AY544173), together comprising 51.9% of the *TtPDI4A* gene, 38.4% of exon sequences and 66.5% of intron sequences (Figures 3.6 and 3.7). A comparison of the exon sequences in these two sections with the corresponding regions of *TtPDI4A* revealed one SNP at the wobble position of a codon in exon 5, and a complete conservation with the *wPDI1* cDNA. The comparable intron sequences in these two gene sections and *TtPDI4A* revealed 99.8% conservation, with only two SNPs in intron 5 (Figure 3.7). However, none of the SNPs explained the RFLPs observed on the Southern blots of *T. turgidum* and *T. aestivum* cv Wyuna (Table 3.1).

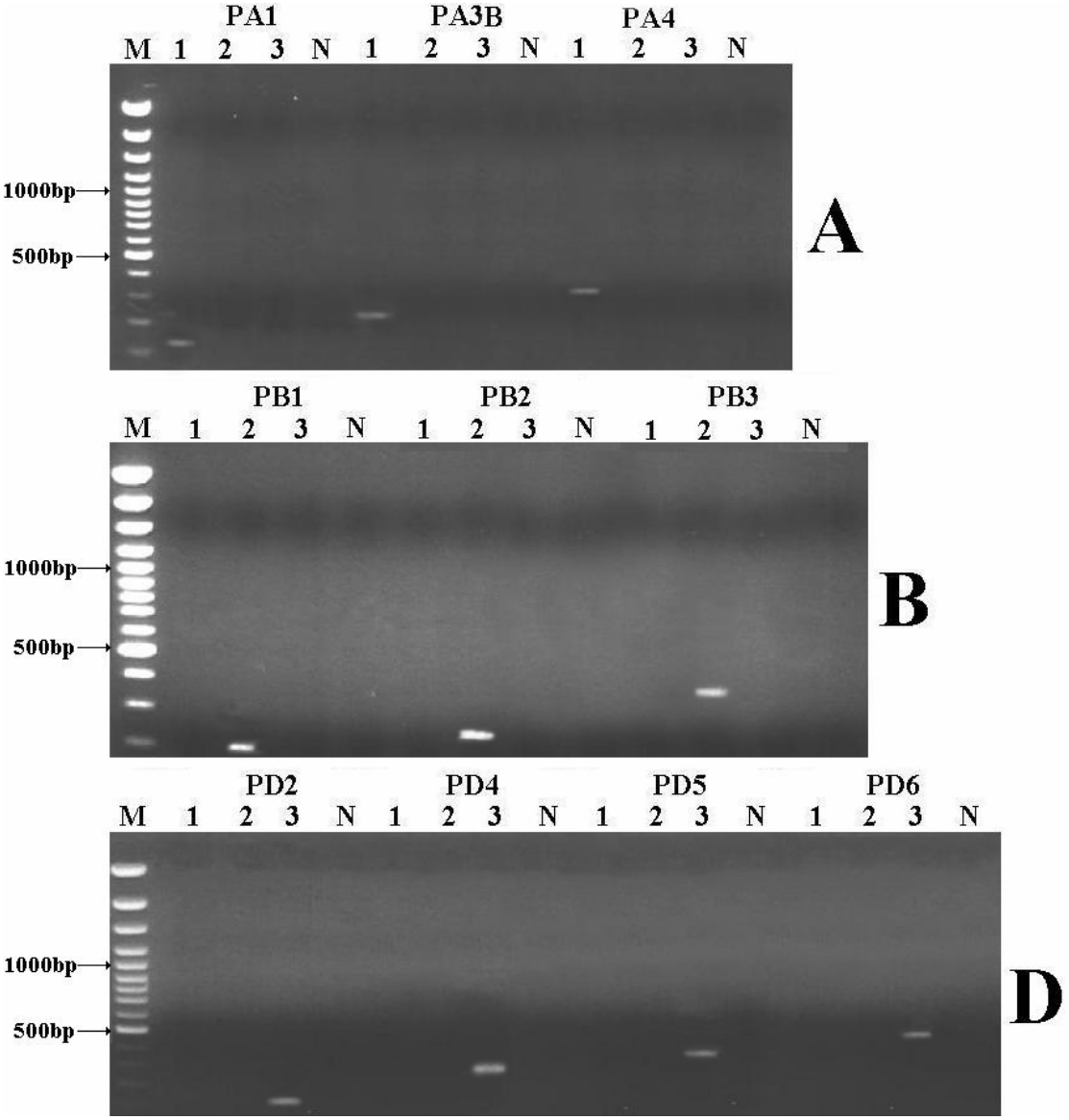


Figure 3.4 Confirmation of the allele-specificity of *PDI* primers

The results of amplifying sections of the *PDI* genes from *wPDI1* (1), *wPDI2* (2) and *wPDI3* (3) with the primers specific for the A genome (A), B genome (B) and D genome (D) of wheat. N: No template negative control, M: 100bp DNA marker.

3.4.2 The *PDI* gene in the B genome of *T. aestivum*

Amplification of sections of the *PDI* gene of *T. aestivum* cv. Katepwa corresponding to the *TaPDIB* gene with three pairs of allele-specific primers (Table 2.6, Figure 2.4) produced products similar in size to those predicted from the *TtPDIB* gene (Figure 3.5B). Direct sequencing of these PCR products allowed the assembly of three contiguous regions of 291bp (Genbank accession# AY544174), 824bp (Genbank accession# AY544175) and 638bp (Genbank accession# AY544176) (Figures 3.6 and 3.7). These contiguous sequences together represent 49.7% of the overall length of the *TtPDI4B* contig, 38.8% and 60.3% of its exon and intron lengths, respectively. Sequence alignments revealed a 100% identity between these sections and the corresponding exons in *TtPDIB* and the cDNA clone *wPDI2,* including the second *Eco*RI site (Figure 3.7), and 99.5% conservation between the appropriate introns, the only differences consisting of three SNPs in intron 5 and one each in introns 8 and 9 (Figure 3.7).

3.4.3 The *PDI* gene in the D genome of *T. aestivum*

Six allele-specific primer pairs based on the sequence of *AetPDI4D* (Table 2.6, Figure 2.4) were used to specifically amplify sections of the *PDI* gene from the D genome of *T. aestivum* cv. Katepwa, revealing sizes similar to those predicted from *AetPDI4D* (Figure 3.5D). Direct sequencing of these products allowed the construction of four contiguous sections, in total covering approximately 84.5% (78.0% of exon sequences and 91.0% of intron sequences) of *AetPDI4D,* sequentially spanning, (i) 797bp (AY544177); (ii) 441bp (AY544178); (iii) 774bp (AY544179); and (iv) 965bp (AY544180) (Figures 3.6 and 3.7). A comparison of exon sequences covered in these sections with the corresponding ones of *AetPDI4D* revealed a single nucleotide difference in exon 9 (G→A) (Figure 3.7), which would lead to a Glu433→Lys433 mutation in *T. aestivum* cv. Katepwa PDI. The intron sequences in the above gene sections revealed a slightly lower degree of conservation (97.7% identity), the differences consisting of a 34bp deletion at the end of intron 1 in *T. aestivum* cv. Katepwa, two extra bases in *AetPDI4D*, and one SNP. Interestingly, none of these

explained the fragment size differences observed between the Southern blots of *Ae. tauschii* and *T. aestivum* cv. Wyuna (Table 3.1).

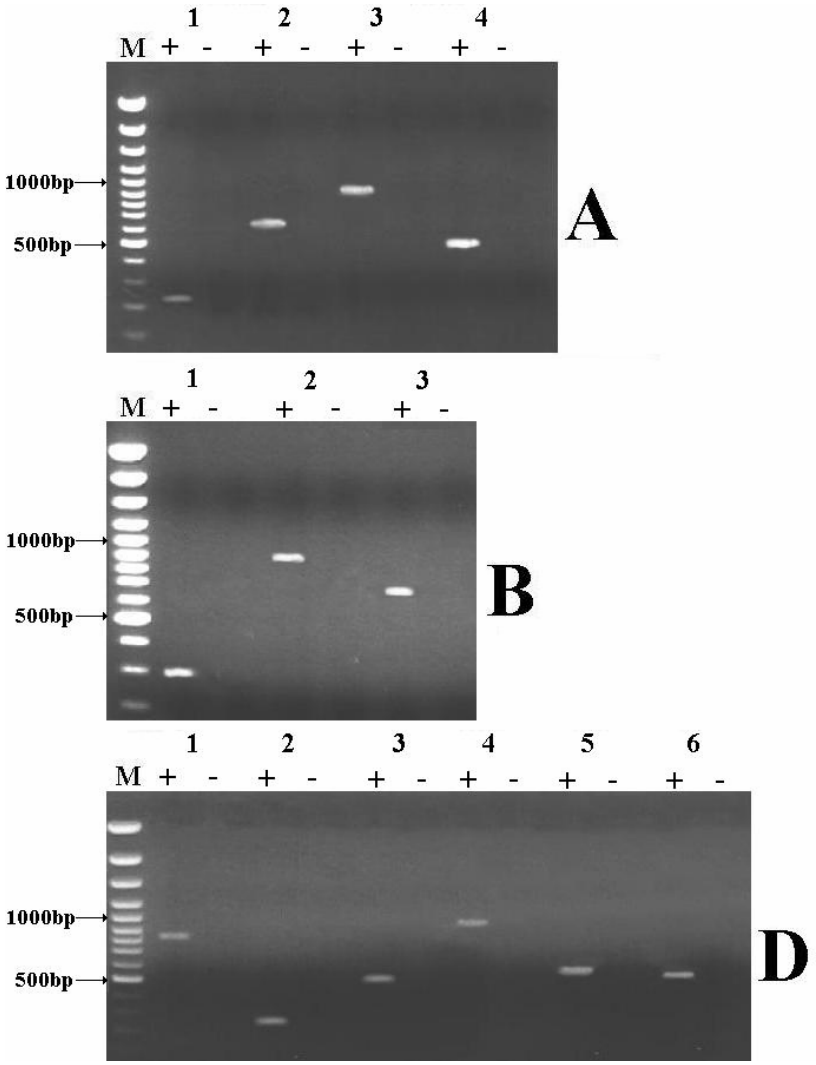


Figure 3.5 *PDI* AS-PCR products from *T. aestivum* cv. Katepwa

The *PDI* PCR products generated from *T. aestivum* cv. Katepwa using primers specific to *TtPDI4A* (A), *TtPDIB* (B) and *AetPDI4D* (D). The numbers above the photo correspond to the AS-primer pair (eg. A1=PA1F/PA1R, D3=PD3F/PD3R, etc.). +: *T. aestivum* cv. Katepwa gDNA template, -: water only negative control, M: 100bp marker.

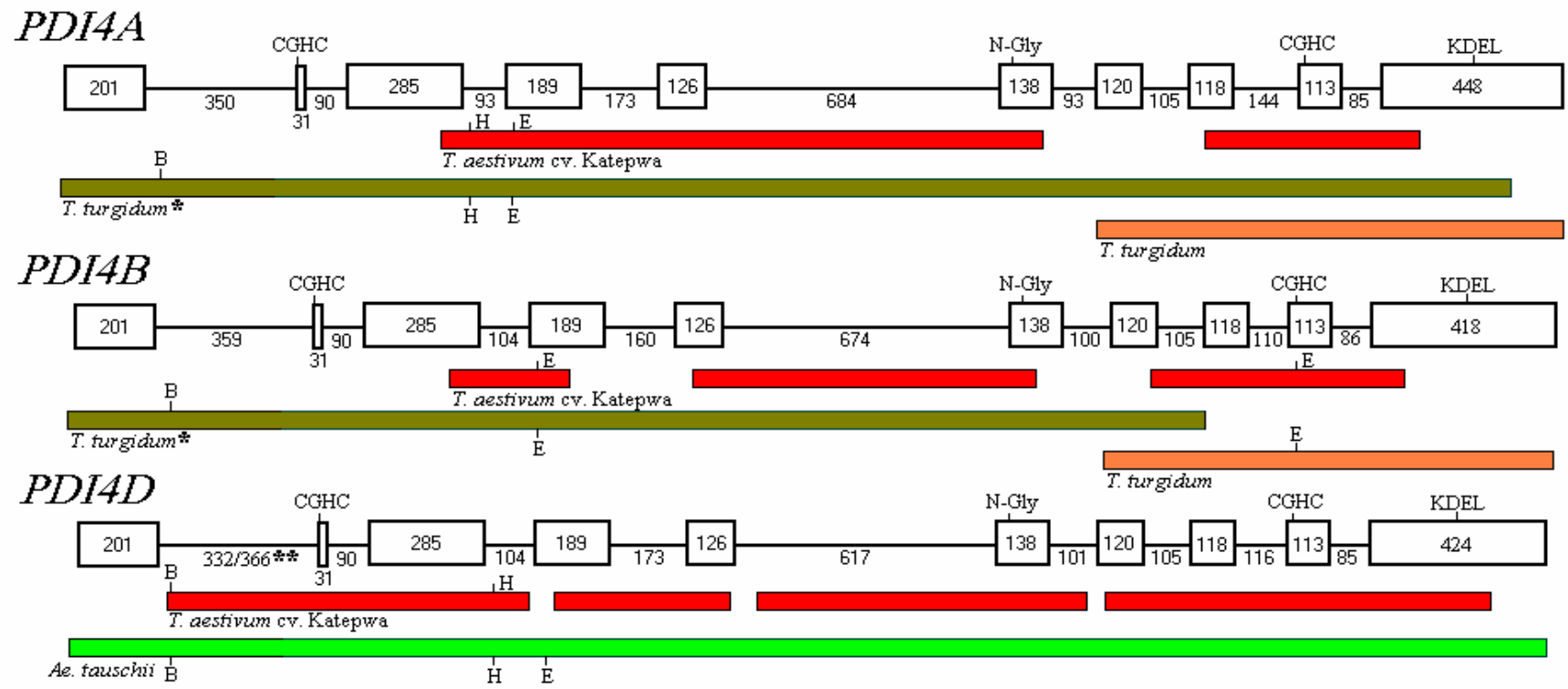


Figure 3.6 The characterised sections of the *PDI* genes of *T. aestivum*, *T. turgidum* and *Ae. tauschii*

Diagrammatic representation of the *PDI* genes of wheat and the sections characterized. Exons are depicted as boxes and introns as lines, with numbers indicating the size of each (in bp). The sections of PDI genes in *T. aestivum* cv. Katepwa characterized in this chapter are shown in relation to those of *T. turgidum* and *Ae. tauschii* characterized in Chapter 3 and an earlier study (Ciaffi *et al.* 2001). The positions of *Bam*HI (B), *Eco*RI (E) and *Hind*III (H) restriction sites, as identified from the sequence data and supported by Southern blot data (Section 3.2), are indicated. The positions of the sequences encoding the putative thioredoxin-like active site (CGHC), N-glycosylation site (N-Gly) and ER-retention signal (KDEL) are indicated. *Results of Ciaffi *et al.*, 2001. **The two different intron lengths represent the only sequenced area with a significant polymorphism (i.e. > a single base deletion) between the hexaploid (332bp) and diploid/tetraploid (366bp) sequences.

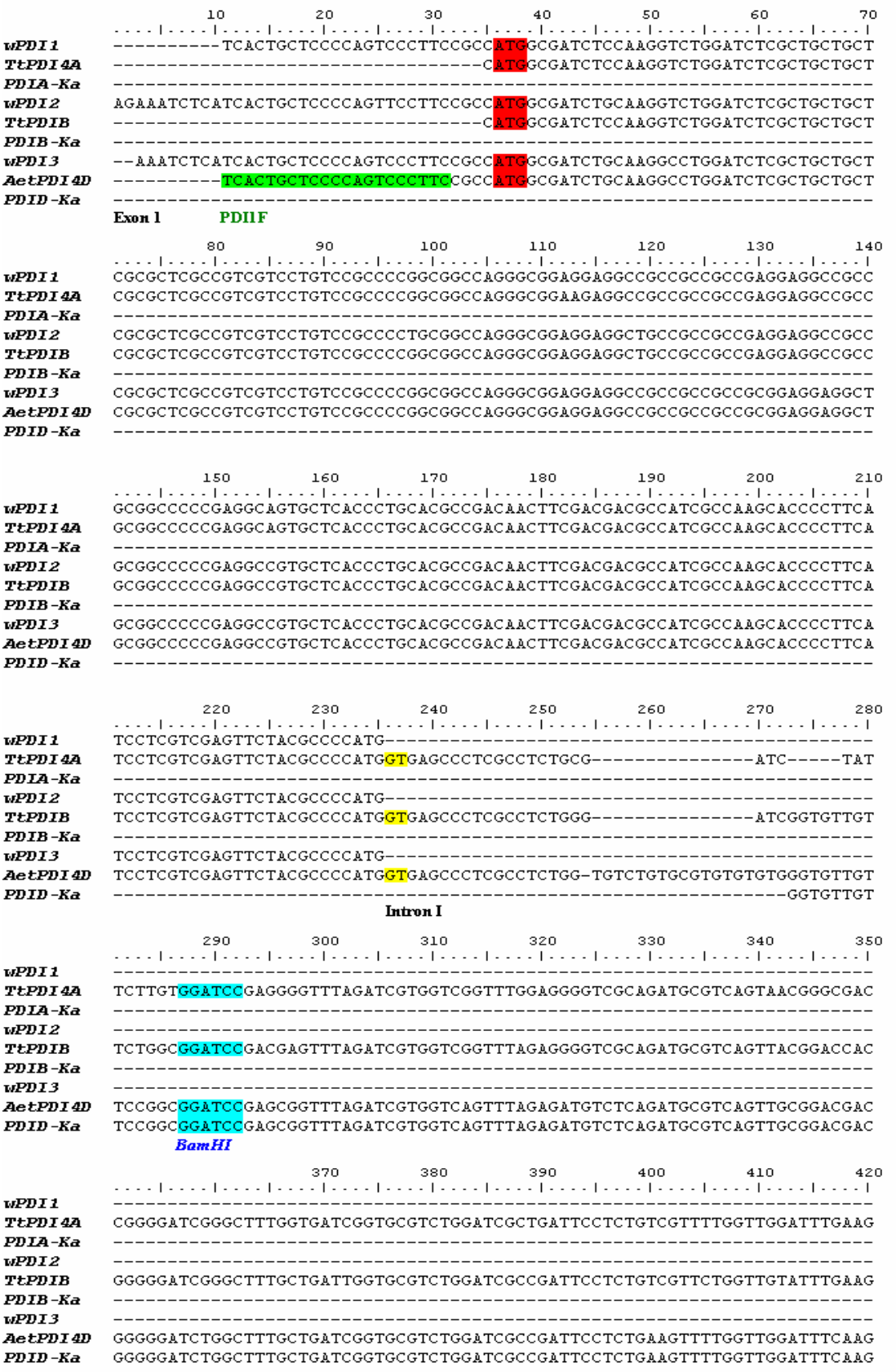


Figure 3.7 ***to be continued***

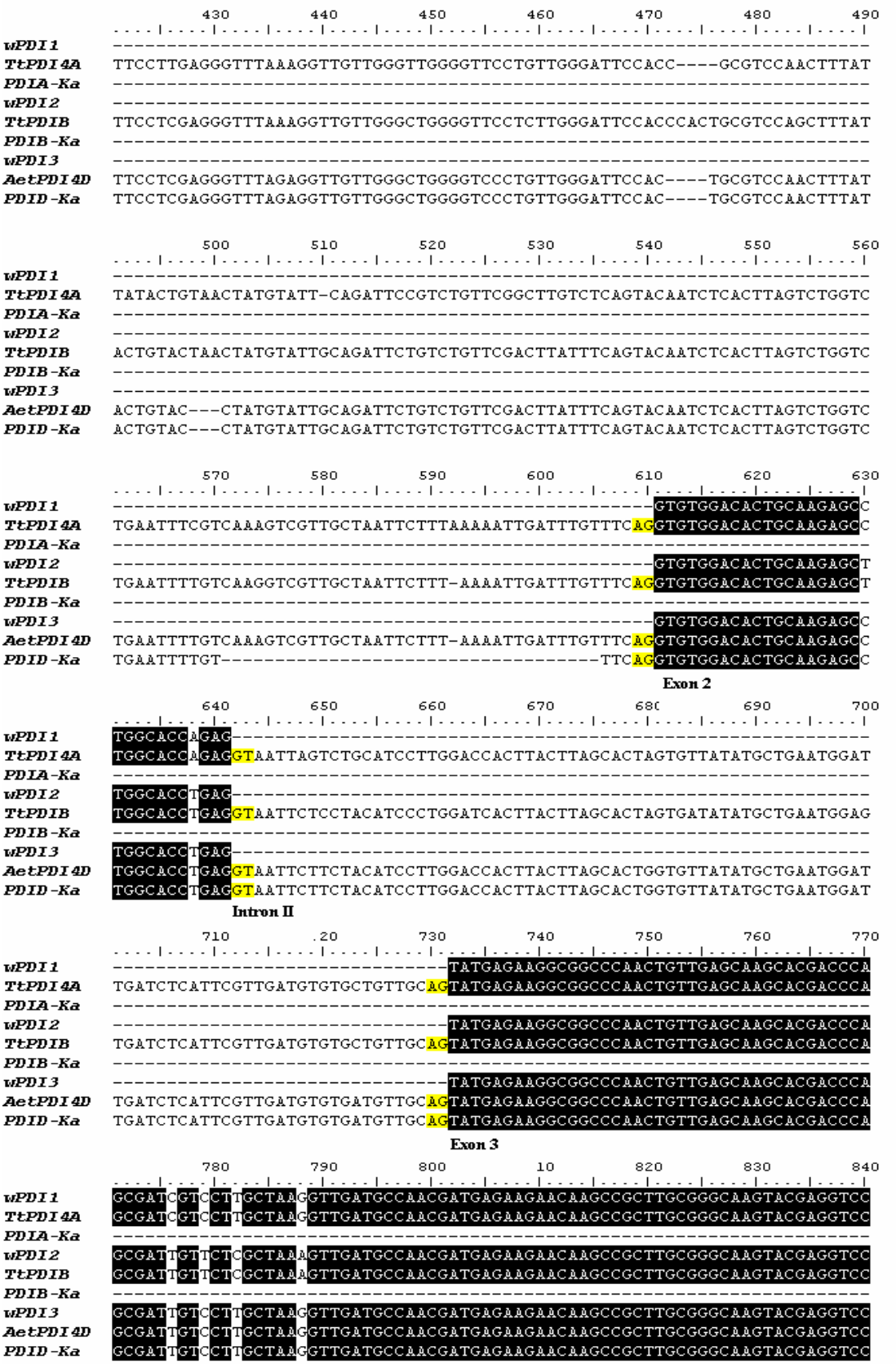


Figure 3.7 ***to be continued***

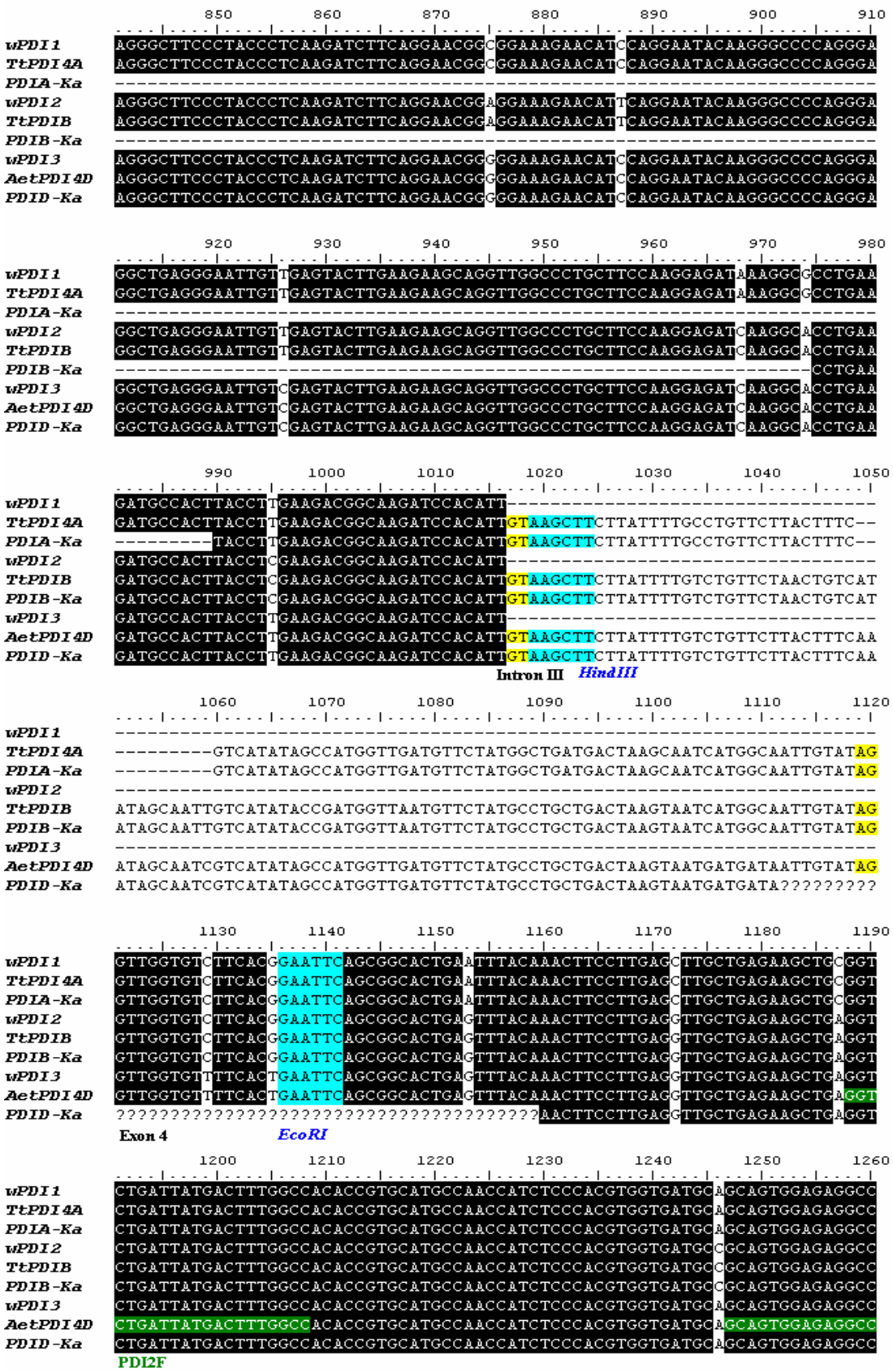


Figure 3.7 ***to be continued***

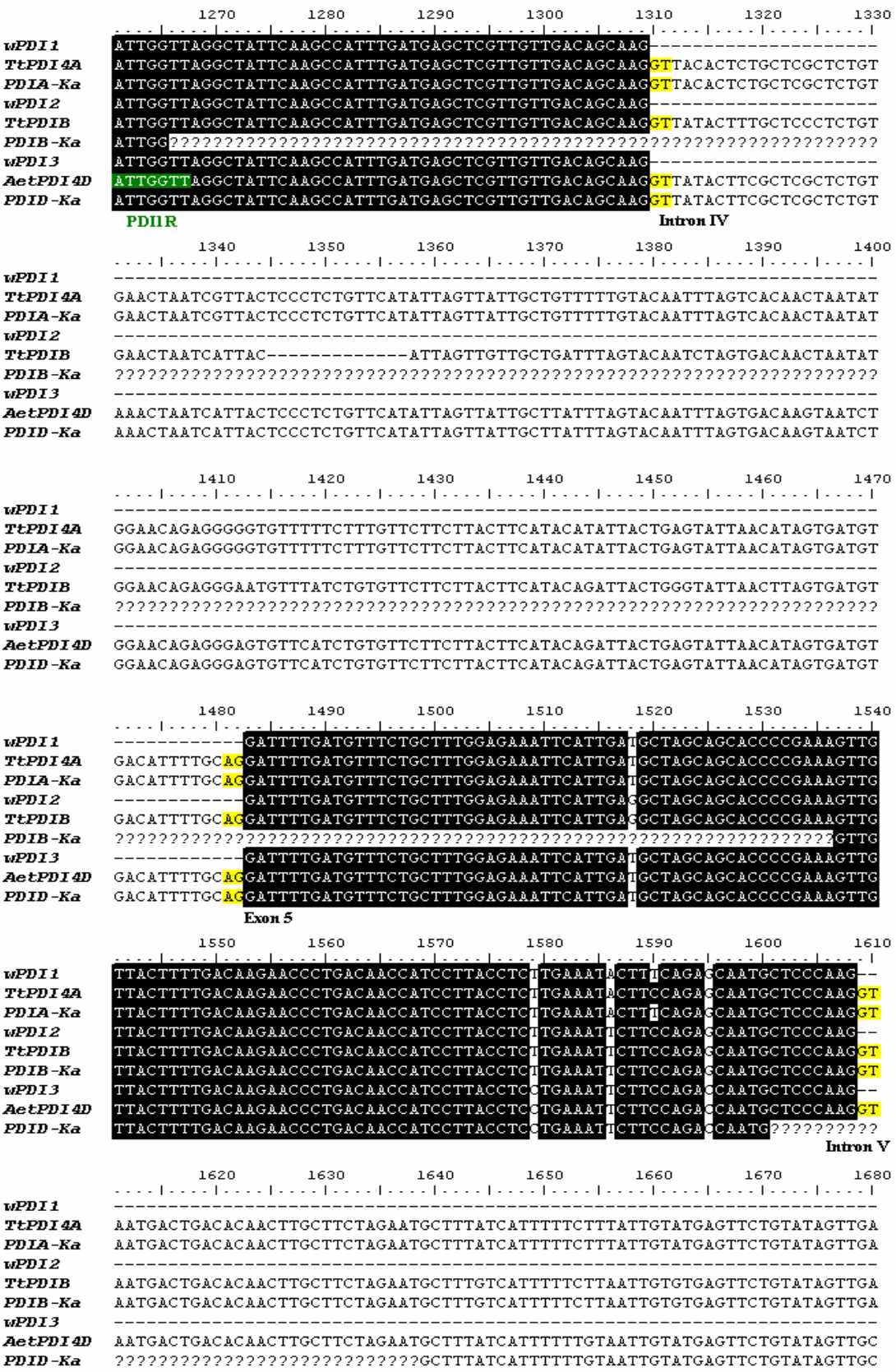


Figure 3.7 ***to be continued***

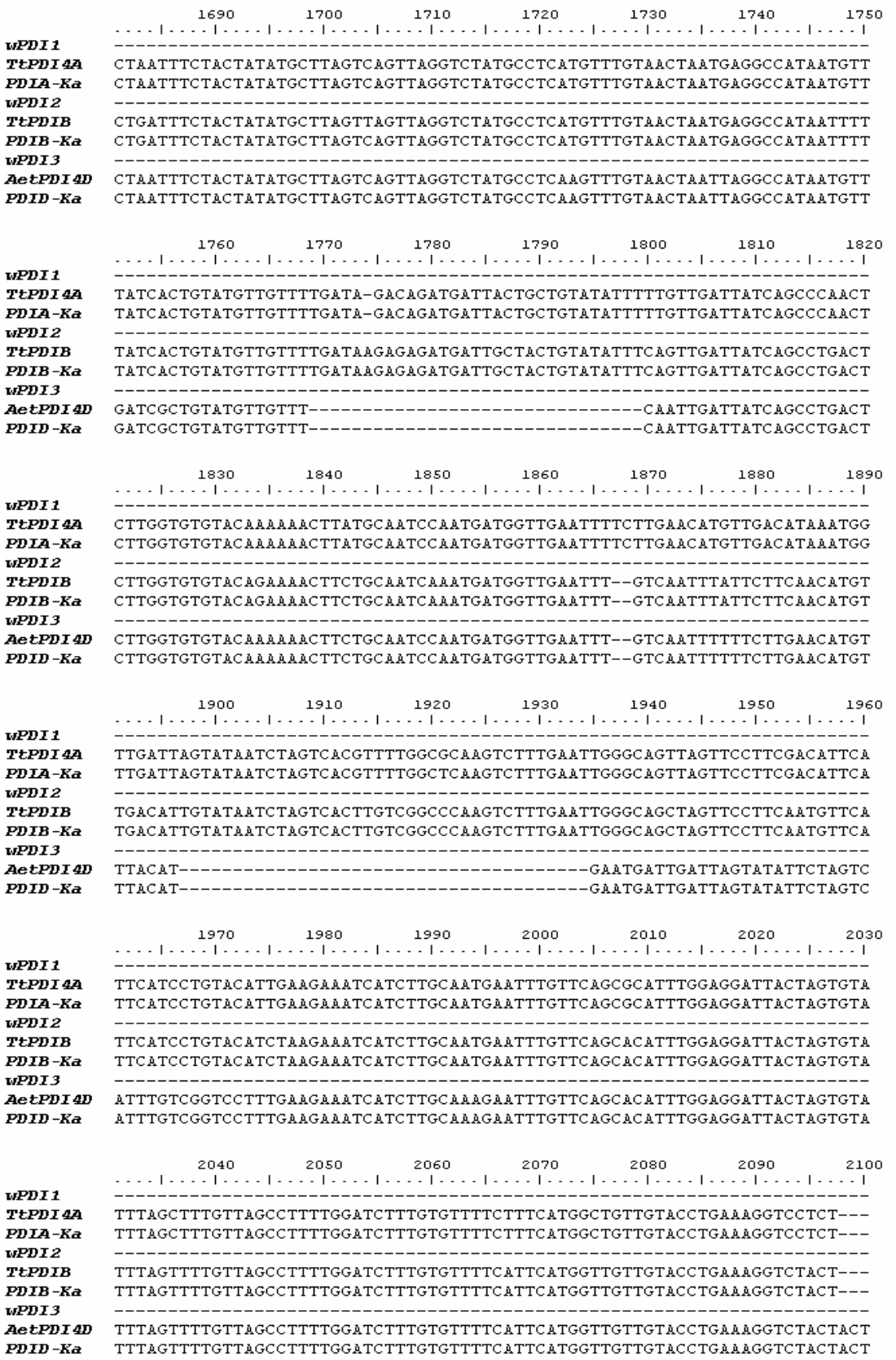


Figure 3.7 ***to be continued***

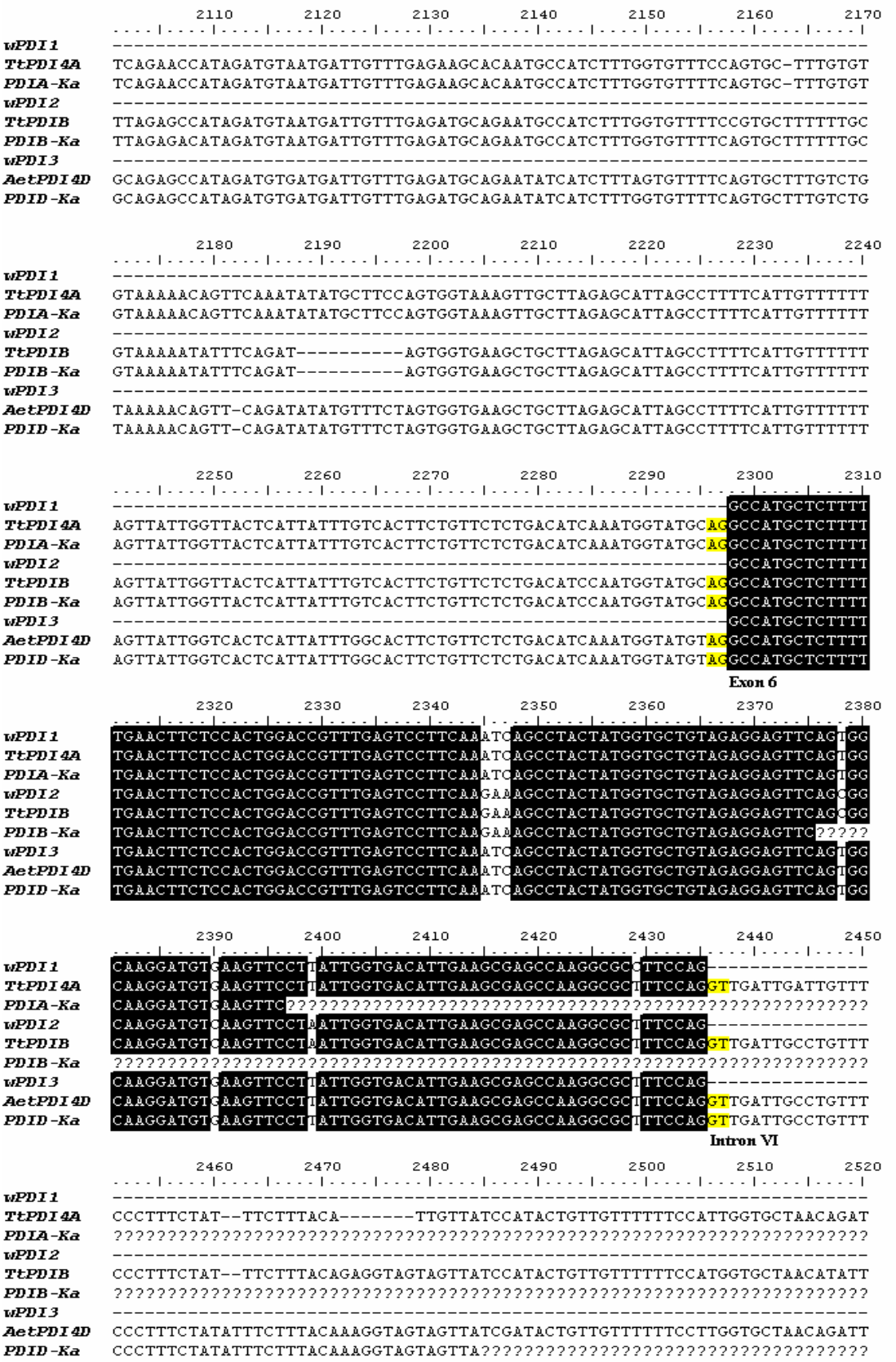


Figure 3.7 ***to be continued***

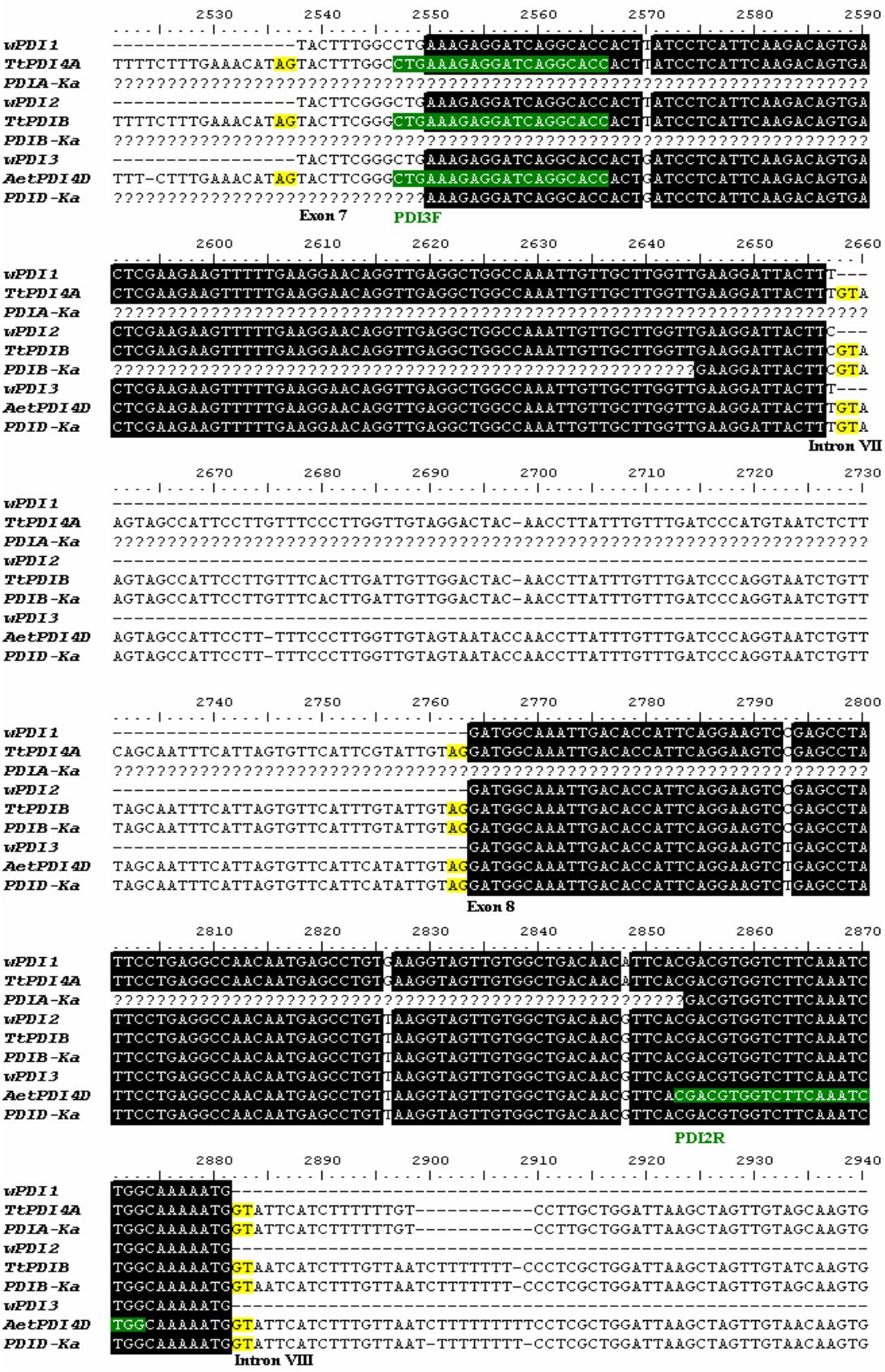


Figure 3.7 ***to be continued***

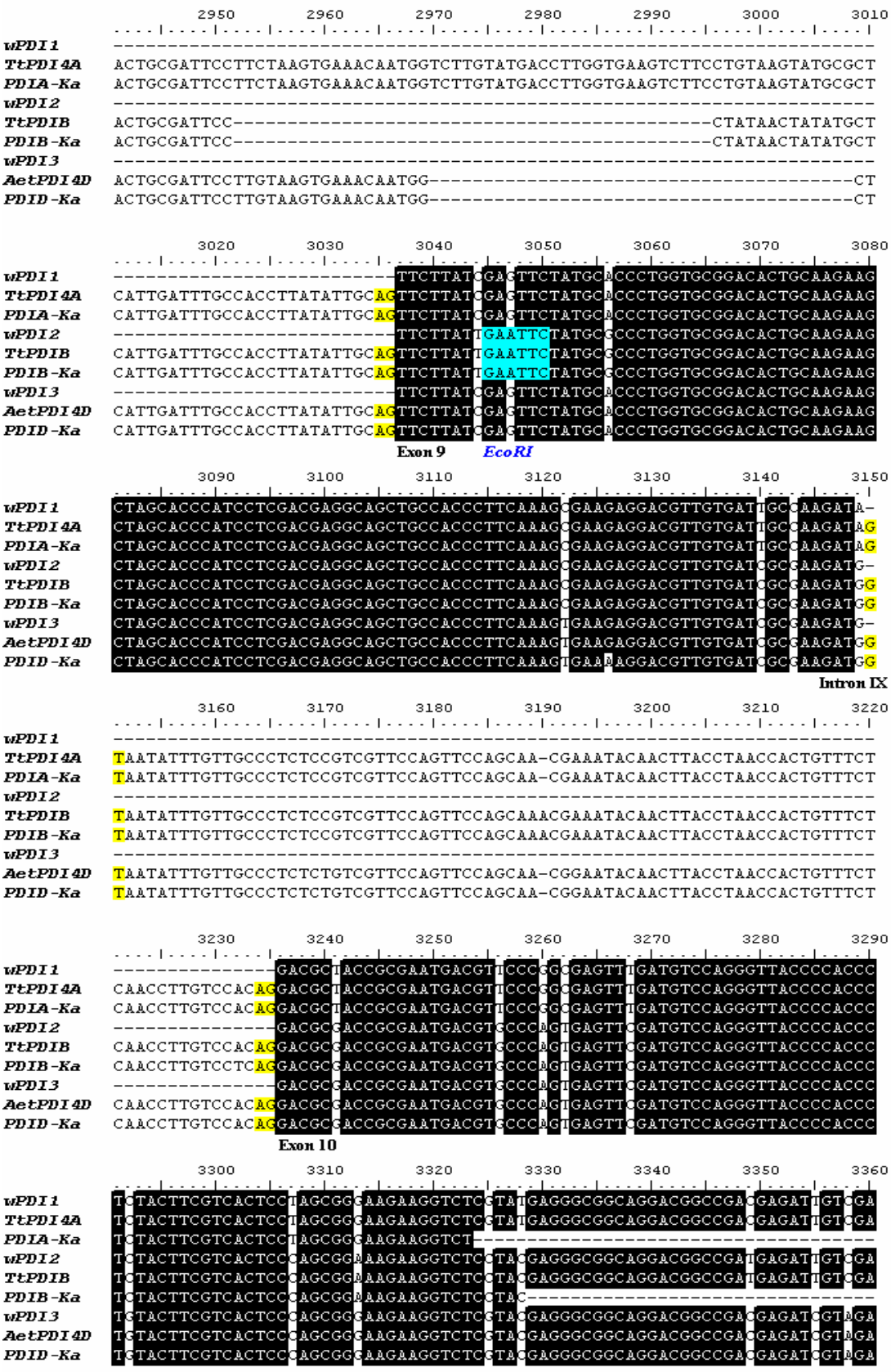


Figure 3.7 ***to be continued***

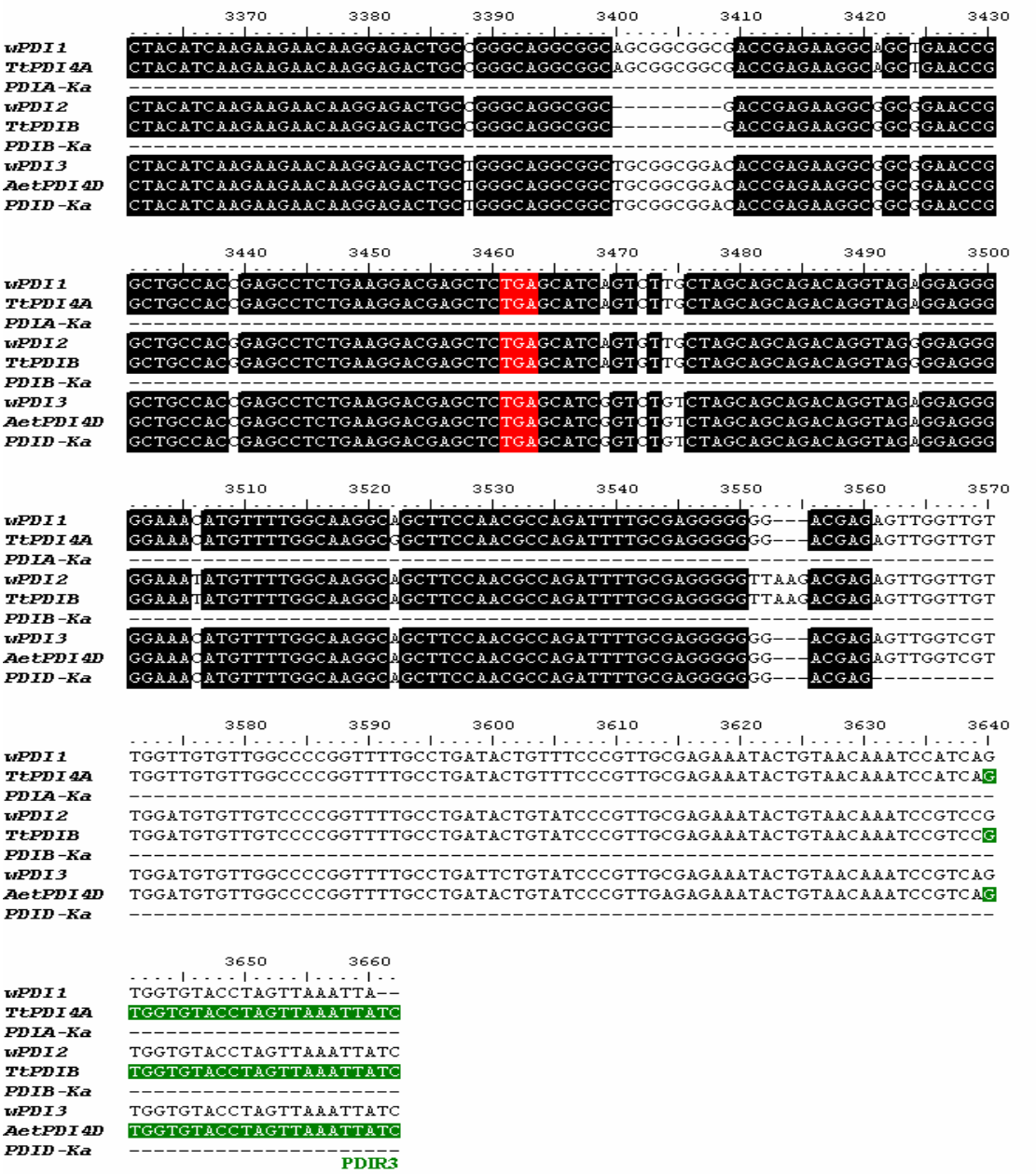


Figure 3.7 Alignment of the *PDI* genes and transcripts of wheat

The primers used to amplify the genes from the gDNA of *T. turgidum* and *Ae. tauschii* are highlighted in green. The sequences of *TtPDI4A* and *TtPDIB* represent a contig between the sequence of PCR products amplified with the PDIF3/PDIR3 primers in this study and the reported sequence data (Ciaffi *et al.* 2001). Uncharacterised gaps between contigs are indicated by "?". The exons and introns are indicated, the conserved GT/AG at the intron boundaries are highlighted in yellow, the *Eco*RI, *Bam*HI and *Hind*III restriction sites are highlighted in blue and the start (ATG) and stop (TGA) codons are highlighted in red.

3.5 ATTEMPTS TO IDENTIFY ADDITIONAL EXPRESSED *PDI* GENE(S)

As the amplification and sequencing of the three *PDI* genes in the hexaploid revealed single, unambiguous sequence types, it was undertaken to try to identify a fourth *PDI* gene corresponding to the hybridising fragment identified in Southern blots by Ciaffi *et al.* (1999) to ensure that the genetic and physical mapping experiments were conducted for this gene too, if indeed it was present.

3.5.1 Screening of cDNA clones for novel *PDI* genes

A cDNA library had previously been prepared from *T. aestivum* cv. Wyuna developing endosperm tissue from which the three *PDI* cDNA clones, *wPDI1*, *wPDI2* and *wPDI3* were isolated (Johnson *et al.* 2001). As 14 further cDNA clones co-isolated with these cDNAs had remained uncharacterised, it was undertaken to determine if any of these contained a unique sequence possibly representing the fourth *PDI* gene. The secondary screen of these with a ~1750bp section of the *wPDI1* cDNA clone resulted in the identification of positive plaques in 12 of these (Figure 3.8A), which were subsequently picked and excised as pBluescript SK+ plasmids (Section 2.11.4) with apparent RFLPs for *Eco*RI and *Xho*I (Figure 3.8B). Sequencing of the putative *PDI* clones and subsequent multiple sequence alignment with the *wPDI1*, *wPDI2* and *wPDI3* sequences revealed that all 12 clones contained *PDI*-like sequences, four aligned with *wPDI1*, three with *wPDI2* and five with *wPDI3* (raw data not shown). No new sequence differences were identified between each of these new cDNA sequences and one of either *wPDI1*, *wPDI2* or *wPDI3*; however, differences in the lengths of the 3' untranslated regions (UTRs) were observed, with sizes ranging from 193bp to 314bp in the *wPDI1*-type clones, 214bp to 270bp in the *wPDI2*-type clones and 206bp to 270bp in the *wPDI3*-like clones (Figure 3.9).

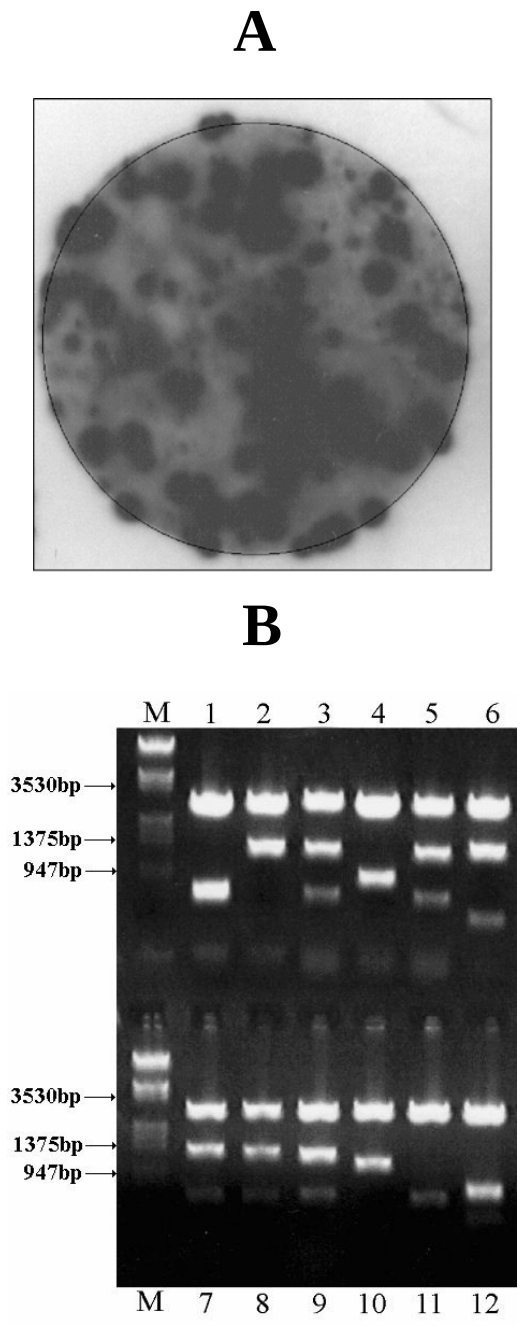
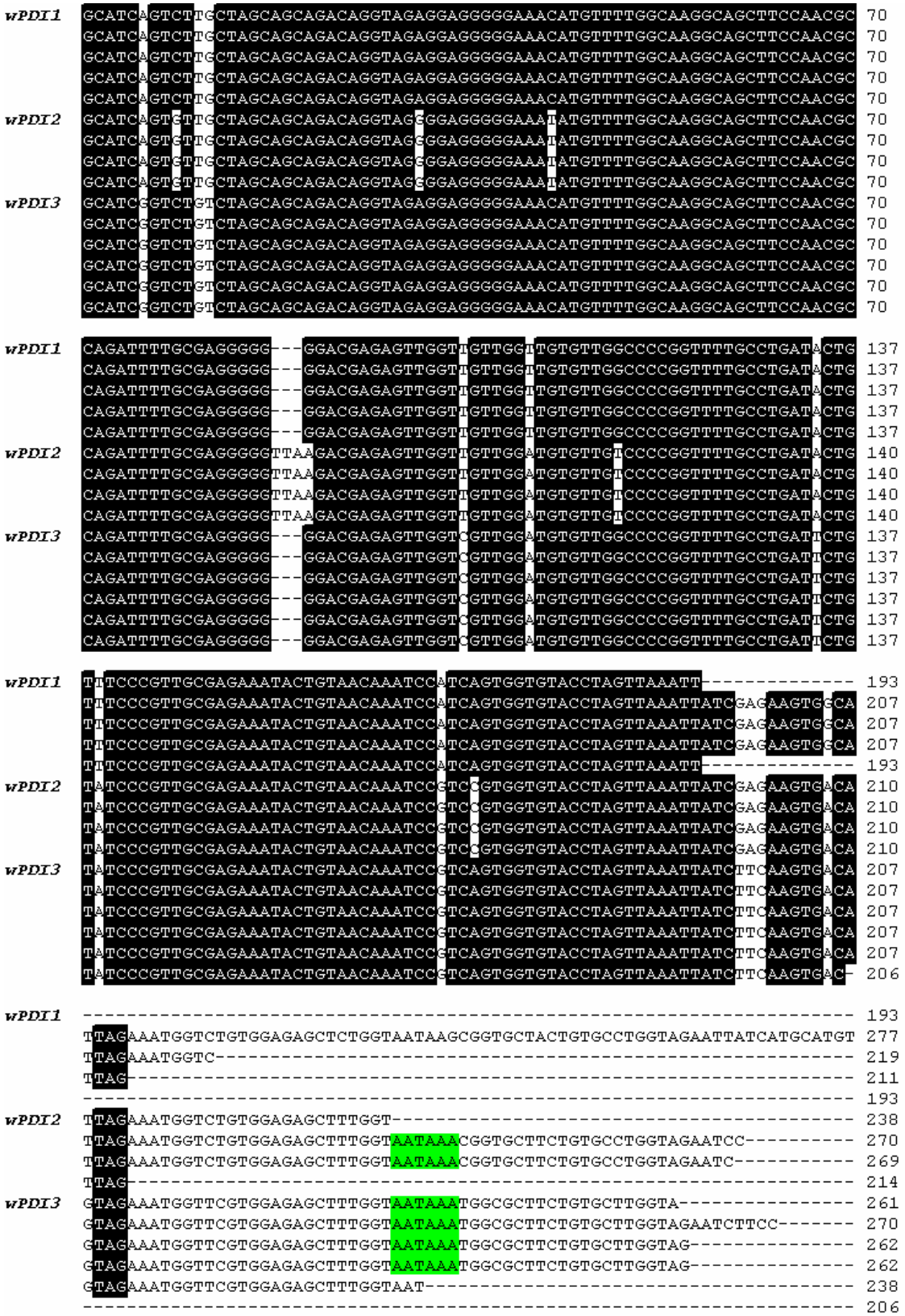


Figure 3.8 The secondary screen of putative *PDI* cDNA clones

(A) One of the twelve secondary screens of *PDI* cDNA clones probed with a 1750bp cDNA probe (Section 2.11). An individual hybridising plaque was selected from each of the twelve such plates containing a positive result and characterised further.

(B) *Eco*RI / *Xho*I restriction digestion of the 12 putative *PDI* cDNA plasmids. 1-12: The 12 digested putative *PDI* cDNA clones M: λ *Eco*RI/*Hind*III molecular weight marker.


wPDI1
wPDI2
wPDI3

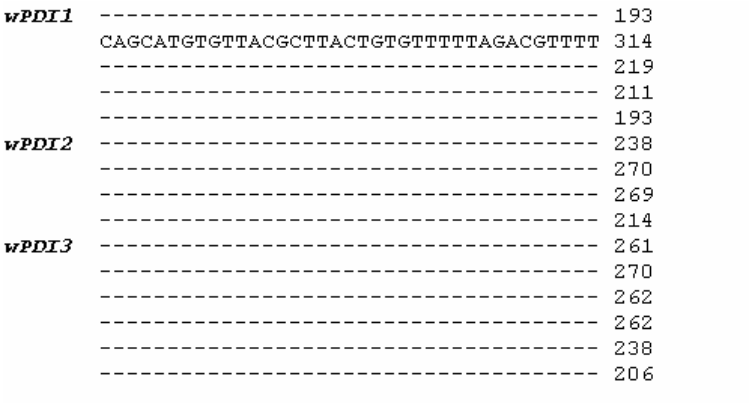


Figure 3.9 Variations in the length of the 3'UTRs of *PDI* cDNAs

A multiple sequence alignment of the 3'UTRs of the *PDI* cDNAs. The typical poly-adenylation sequence AATAAA is highlighted in green when present.

3.5.2 Screening of the TIGR TaGI database for novel *PDI* genes

The efforts to find any variants of *PDI* genes were also extended to mining the substantial number of publicly available wheat ESTs, through screening the TIGR *Triticum aestivum* gene index (TaGI version 7.0) using a query sequence consisting of the active site peptides encoded by the three *PDI* cDNAs (Section 2.12). This revealed three tentative consensus (TC) sequences, *TC105432*, *TC105037* and *TC104854*, which corresponded to *wPDI1*, *wPDI2* and *wPDI3*, respectively. Four other TC sequences constructed from more than 10 ESTs were identified, however, BLASTx analysis of these suggested *TC104975*, *TC127004* and *TC109004* most likely encode other members of the PDI superfamily containing the CXXC active site motif, due to their close sequence similarity on the amino acid level to alfalfa PDI-Dα (P38661) (86%), human P5 (Q15084) (68%) and *Arabidopsis thaliana* ERp72 (CAC81067) (78%), respectively, and *TC107055*, encoded an unknown member of the PDI superfamily. Analyses of the nucleotide sequences of these other members of the PDI superfamily suggest that they do not explain the fourth hybridising *PDI* gene reported by Ciaffi *et al.* (1999) due to their limited nucleotide sequence identity (41-50%) to the *PDI* cDNA used as a probe in this study. Thus, no additional *PDI*-like sequence, putatively encoded by the locus on chromosome 1B, was identified in this search also.

3.6 FURTHER INVESTIGATIONS INTO THE FOURTH *PDI* GENE

3.6.1 Copy number of *PDI* in *T. aestivum* cv. Chinese Spring

As bands corresponding to a fourth *PDI* gene could not be found on the Southern blot of *T. aestivum* cvs. Wyuna (Figure 3.1A), a Southern blot of *Eco*RI-digested gDNA of *T. aestivum* cv. Chinese Spring was also conducted to confirm the copy number of four hybridising fragments previously reported (Ciaffi *et al*. 1999) and identify any polymorphisms between Chinese Spring and Wyuna. Surprisingly, only three hybridising fragments were observed (Figure 3.1A and Table 3.1), corresponding to those on the *T. aestivum* cv. Wyuna blot.

3.6.2 The probing of Wyuna and Chinese Spring Southern blots with a 1466bp *PDI* cDNA

The absence of the reported 4.6kb fragment from the *Eco*RI digests on the Southern blots, corresponding to a gene on chromosome 1B in Chinese Spring (Ciaffi *et al.* 1999) was puzzling, however, this could have been due to the fact that the 838bp cDNA probe used for this work, although covering substantial sections of the gene, would not have hybridised to the 5' *Eco*RI fragments arising from the internal *Eco*RI site in (at least three of) the *PDI* genes (Figure 3.6). Therefore, a longer, 1466bp probe similar to that of Ciaffi *et al*. (1999) (Figure 2.2) was used for Southern hybridisations. This resulted in the hybridisation of at least 6 major bands in the *Eco*RI digests of both Wyuna and Chinese Spring (Figure 3.10), including the three bands identified previously (Figure 3.1), and additional bands at 4.4kb 5.7kb and 3.6kb, of which the 4.4 kb band could possibly represent the *PDI* gene on chromosome 1B reported by Ciaffi *et al.* (1999).

3.6.3 Characterization of the putative *PDI1B* gene

It was then undertaken to determine whether the above 4.4kb band, represented a variant *PDI* gene lacking the *Eco*RI site, or some form of partial gene, or whether it was the 5' or 3' fragment of some or all the *PDI* genes. Large-scale *Eco*RI digests of gDNA of *T. aestivum* cvs. Chinese Spring and Wyuna were electrophoresed and DNA in the size range of 4.2 to 4.9kb was gel-purified and used as a template to amplify sections of

the *PDI* gene (Section 2.13). Amplifications 5' to the *Eco*RI site (consensus primers PRB5F/PRB5R, Table 2.5, Figure 2.2) resulted in an ~800bp product in both of the cultivars, while amplifications 3' to this site (consensus primers PRB3F/PRB3R) amplified two faint products, between 1000 and 1100bp, in the gel-purified DNA and the undigested total gDNA of *T. aestivum* cv. Chinese Spring, but none in the Wyuna gel fraction (Figure 3.11). Amplifications using the primer pair PRB5F/PDI1R encompassing the conserved *Eco*RI site (Table 2.5, Figure 2.2) resulted in a ~1000bp product in all fractions and in the undigested total gDNA of both cultivars, which produced a ~920bp observable product when digested with *Eco*RI (results of Wyuna not shown) (Figure 3.11).

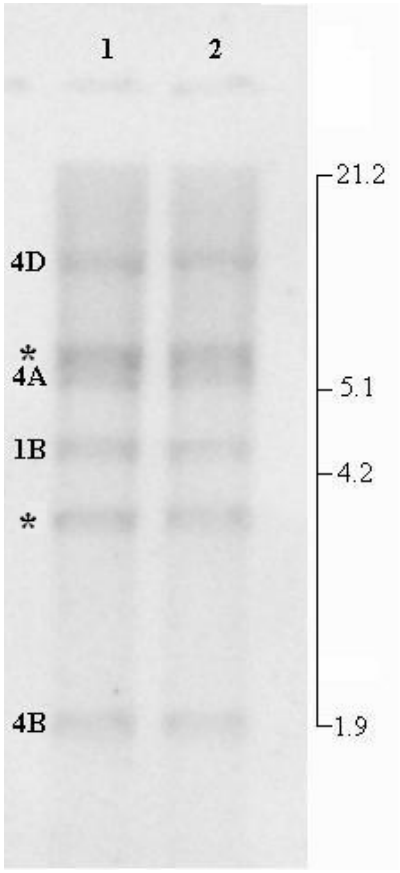


Figure 3.10 Southern blot of *PDI* in *T. aestivum* with a 1466bp probe

Southern analyses of *T. aestivum* cvs. Chinese Spring (1) and Wyuna (2) gDNA digested with *Eco*RI and probed with the 1466bp cDNA probe (Figure 2.2). The chromosomal location of the fragments, when known (Ciaffi *et al.* 1999), is indicated. *Fragments not reported previously. The molecular weight marker (kb) is given on the right side.

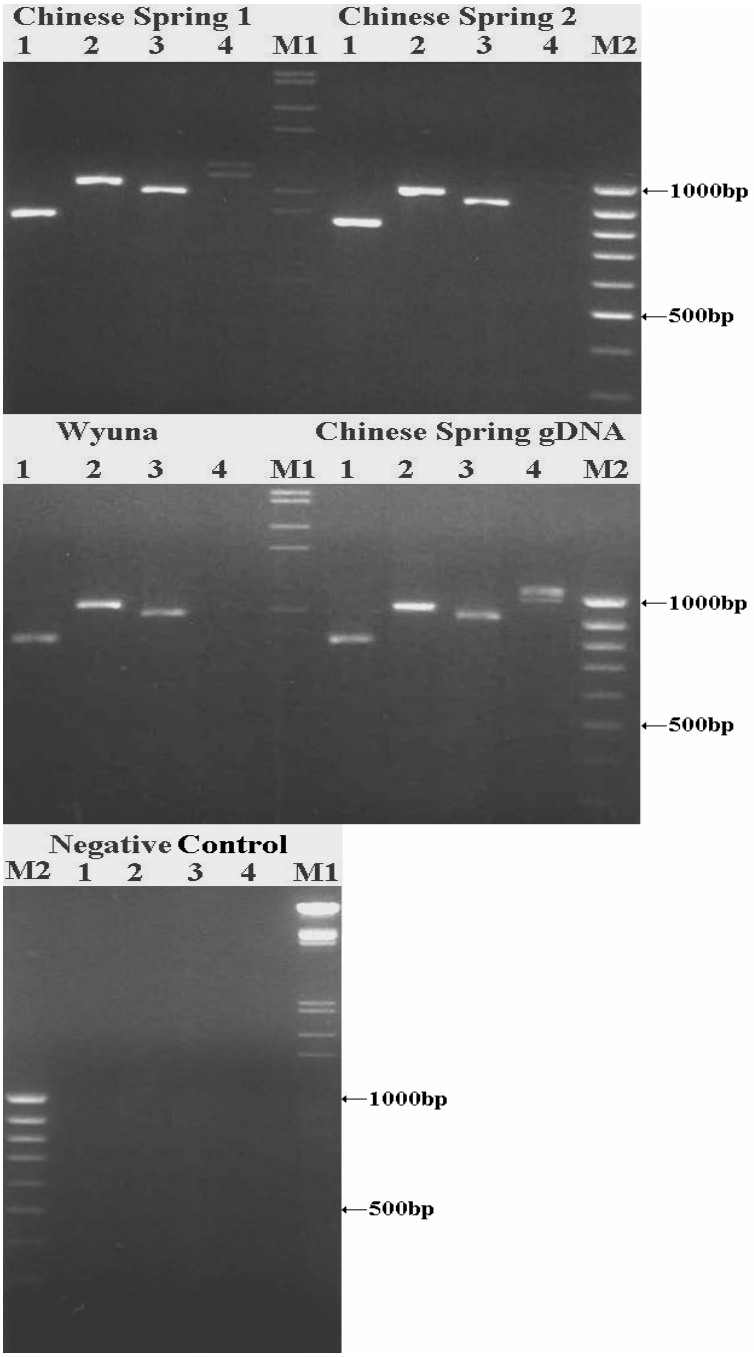


Figure 3.11 PCR investigation of the *PDI* gene reported on chromosome 1B

Amplification of sections of the *PDI* genes with primers located 5' to the conserved *Eco*RI site (Figure 3.6) (primers PRB5F/PRB5R) (Lane1), encompassing this conserved *Eco*RI site (primers PRB5F/PDI1R) (Lane 2) and 3' to this conserved *Eco*RI site (primers PRB3F/PRB3R) (Lane 4). Digestion of the PRB5F/PDI1R PCR product (Lane 2) with *Eco*RI, is shown (Lane 3). The templates for these PCRs were the ~4.4 kb *Eco*RI fraction of Chinese Spring (in duplicate) and Wyuna gDNA, total gDNA from Chinese Spring and a no-template negative control. M1: λ *Eco*RI/HindIII marker, M2: 100bp marker.

3.7 DISCUSSION

The main biochemical role of the PDI enzyme, i.e., mediation of disulfide bond isomerisation or "foldase" activity, has made it a factor of high potential in terms of regulation of the folding and deposition of the rapidly synthesised and precisely deposited storage proteins in cereals. As described in Section 1.5.6 circumstantial evidence such as its up-regulation in the developing endosperm of wheat and its physical presence within the storage protein bodies, as well as more direct evidence such as the improper segregation of the protein bodies in a rice mutant lacking PDI add weight to this hypothesis. However, in spite of potentially playing a role of great significance in terms of determining or influencing various parameters of cereal grain quality, PDI studies in cereals have been rather limited. Partial sequence information exists for two of the durum *PDI* genes, however, the structure and organisation of the *PDI* gene family and any variations/ mutations therein in common wheat, the cultivars of which vary greatly in protein content and dough properties, the trends that affect global wheat trade, are largely unstudied; some of these shortcomings are addressed in the present chapter.

Towards this purpose, firstly a better understanding of the structure and composition of this gene family was required. During the course of addressing this issue, our detection of three bands on the Southern blots of *T. aestivum* cv. Wyuna and Chinese Spring gDNA, in conjunction with our previous isolation of three cDNA classes from a Wyuna endosperm library (Section 1.5.5) suggested that the *PDI* gene family of *T. aestivum* consisted of three members. However, while our studies were underway, another study reported four *PDI* bands in aneuploid lines of Chinese Spring, localised to chromosome arms 4AL, 4BS, 4DS and 1BS, and subsequent studies showed that the gene on chromosome 1BS was not present in all cultivars and showed significant RFLP variability when present (Section 1.5.5). This raised questions as to whether this gene was functional and important in the deposition of storage proteins in wheat, and whether there were differences in PDI gene copy number between different cultivars of common wheat, a factor of significance for the genetic and physical mapping of these genes (Chapters 4 and 5). Thus, to elucidate the size and organization of the *PDI* gene family, we extended our studies of these genes to *T. turgidum,* the tetraploid progenitor of the A and B genomes, and *Ae. tauschii*, the diploid progenitor of the D genome as well.

3.7.1 The A and D genomes of wheat contain a single, highly conserved *PDI* gene each

The presence of a single intense band on the Southern blot of *Ae. tauschii* gDNA and the isolation of PCR products of a single sequence type strongly suggest that this species contains a single *PDI* gene, supporting the results of Ciaffi *et al*. (1999; 2000). The sizes of bands in *Hind*III and *Bam*HI digests for this gene were similar to those estimated for the PDI gene on chromosome 4D of *T. aestivum* (Ciaffi *et al.* 1999) and our results for *T aestivum* cvs. Wyuna and Chinese Spring, but a size difference existed for *Eco*RI, probably due to differences in the flanking sequences, as the sequence data for the *Ae. tauschii PDI* gene and sections of its homologue isolated from *T. aestivum* did not exhibit variations for this site. Further, the latter appears to be expressed in developing wheat endosperm, as a cDNA clone *wPDI3*, highly similar (99.8%) to the exons in the *Ae. tauschii* gene, has been isolated from a cDNA library from this tissue (Johnson *et al*. 2001). As has been found for other genes (Section 1.5.5), the *PDI* gene from *Ae. tauschii* and its homologue from *T. aestivum,* revealed an extremely high degree of conservation, particularly for exon sequences (99.9%) in comparison to introns (97.7%). One noteworthy difference was a 34bp deletion in intron 1 of the *T. aestivum* D-genome based gene, a polymorphism which may be useful for the development of a molecular marker for this gene.

Our Southern blots also confirmed a single strong hybridisation signal of a size corresponding to the *PDI* gene reported to be on chromosome 4A in *T. aestivum* (Ciaffi *et al*. 1999). Isolation and analysis of the tetraploid *T. turgidum PDI* genes was then utilised to address the *PDI* gene from the A genome of *T. aestivum*. Sequencing of most of the A-genome *PDI* gene in *T. turgidum* had been reported as our work was underway (Ciaffi *et al*. 2001), the current study thus confirmed these results and contributed an additional 86bp of the 3' UTR of the gene. The near-complete identity between the *wPDI1* cDNA clone isolated previously from a *T. aestivum* developing endosperm cDNA library (Johnson *et al.* 2001), exons of the *T. turgidum* A-genome *PDI* gene (Ciaffi *et al.* 2001; Section 3.3.3), and those on sections of the corresponding gene cloned from *T. aestivum* (Section 3.4.1) suggests that the *PDI* gene on the A genome is

highly conserved between the tetraploid and hexaploid wheat and actively expressed in the developing endosperm in the latter.

3.7.2 The sequenced *PDI* gene of *T. aestivum* and *T. turgidum* is from chromosome 4B

Due to the complexity associated with identifying individual genes in cases of multi-gene families, a number of factors were brought together, such as (i) the partial sequence information for a *T. turgidum* B genome *PDI* gene, of unknown chromosomal location (Ciaffi *et al.* 2001); (ii) our sequence data for *T. turgidum PDI* genes showing only two sequence classes, one of these being the *PDI* 4A gene (Section 3.3.3); (iii) our second sequence class, i.e., for a *T. turgidum* B-genome *PDI* gene, overlapping the Ciaffi *et al.* (2001) sequence with 100% identity and extending it further, leading to identification of a second *Eco*RI in this gene (Section 3.3.3); (iv) the three sections of the equivalent gene from *T. aestivum* showing very high identity (>99.5%) to the corresponding sections of the *T. turgidum* contig and the extra *Eco*RI site (Section 3.4.2); (v) the two internal *Eco*RI sites, observed only in this gene, being able to clearly explain the 1.8 kb bands on the Southern blots (Ciaffi *et al.* 1999; Figures 3.1A and 3.10), which had been localised to the chromosome 4B (Ciaffi *et al.* 1999); and (vi) the cDNA clone *wPDI2* isolated previously (Johnson *et al.* 2001), which showed two internal *Eco*RI restriction sites and most homology to the exon sequences in *TtPDIB* and the three sections of its homologue in *T. aestivum*. Considering these factors together, it appears that the gene of *T. turgidum* reported by Ciaffi *et al*. (2001) and extended by us, and its three homologous sections from *T. aestivum*, represent the *PDI* gene on chromosome 4B, it being highly conserved between *T. turgidum* and *T. aestivum*, exhibiting an extra internal *Eco*RI site, and expressed in the developing endosperm of common wheat, as attested by an equivalent cDNA being isolated from this tissue. Further, based on the reported sizes in the *Eco*RI digests for the second putative B-genome *PDI* gene (from chromosome 1B), the intensity observed for the 1.8kb band in our blots (Figures 3.1A and 3.10), and our sequence data (Section 3.3.3 and 3.4), it appears that this band and the gene discussed above do not represent the 1B gene, therefore the contig *TtPDIB* will henceforth be referred to as *TtPDI4B*.

3.7.3 The *PDI* gene on chromosome 1B may be a partial, non-functional gene

The lack of a third sequence type from *T. turgidum*, a fourth sequence type for *T. aestivum* cv. Katepwa or a fourth type of cDNA, and the Southern blots on *Eco*RI-digested DNAs using the 838bp cDNA probe clearly lacking the appropriate fragment strongly indicated that the gene suggested to be present on chromosome 1B was absent, at least in the accessions we were studying. However, confirming its presence or absence and addressing its sequence (if present) were critical for the genetic and physical mapping work (Chapters 4 and 5). The restriction maps of the three *PDI* genes were then constructed utilising the data discussed above, i.e., results of Southern blots of DNAs of various genotypes (Figure 3.1), the sequence data for various genes or gene sections discussed above (Sections 3.3.3 and 3.4), the conserved *Eco*RI site in the *PDI* genes from *T. turgidum*, *Ae. tauschii* and their corresponding sections from *T. aestivum*, and all three cDNA types from *T. aestivum* and a second *Eco*RI site unique to the *PDI* gene on chromosome 4B and the cDNA *wPDI2*. The results of Southern blots of *Eco*RI digests utilising the 838bp probe (mainly covering 659bp, from exons 4 to 8), ie., one strongly hybridising band for *Ae. tauschii*, two for *T. turgidum* and three for both cultivars of *T. aestivum*, explained the sequence data and the maps put together for these genes, but did not exhibit the ~4.6kb fragment from the chromosome 1B gene in *T. aestivum*; this was expected at least in Chinese Spring (Ciaffi *et al.* 1999). All of these bands would comprise the sections 3' to the conserved *Eco*RI site in exon 4 in the respective genes, as this probe was complementary to only a small section of the genes 5' to this site (179bp of exons 3 and 4) possibly producing bands too faint to observe, and also did not cover the section 3' to the second *Eco*RI site in the 4B gene. The longer, 1466bp cDNA probe, stretching through nearly all exons, was then used to address this shortcoming, and as expected, produced (at least) 6 observable bands, covering all EcoRI fragments. It is not clear why Ciaffi *et al*. (1999; 2000) were unable to visualise the ~3.6kb and ~5.7kb EcoRI bands on their Southerns, conducted with a similar probe; differences in hybridisation conditions and/or detection methods may be able to address this issue.

It was then necessary to investigate whether the ~4.4kb fragment observed with this probe represented the *PDI* gene on chromosome 1B as suggested by (Ciaffi *et al.* 1999), or the 5' *Eco*RI fragment(s) of one (or more) of the three genes which could not be seen

with the shorter probe, or both. The results of a unique experiment conducted for this purpose, i.e., isolation of the corresponding gel fraction, followed by a successful amplification from it of a section of the gene preceding the conserved *Eco*RI site, but the low levels or complete absence of a PCR product with primers for the section 3' to it, suggest that the ~4.4kb band does represent a 5' *Eco*RI fragment of *PDI* gene(s). It might represent one or more of the genes discussed above (4A, 4B or 4D), or additionally, include the 1B gene, which might be a partial gene, missing some downstream sequences. These possibilities would need to be explored further through the analysis of other cultivars of *T. aestivum*. However, the absence of a fourth class of cDNAs in our search or ESTs or TC sequences amongst the 492,694 ESTs used to construct the TIGR TaGI database suggests that, if at all present, this gene is not expressed.

This Chapter has provided a comprehensive characterisation of the *PDI* gene family in *T. aestivum*, *Ae. tauschii* and *T. turgidum*. The data generated in this Chapter has permitted the characterisation of the parental lines used in the construction of mapping populations (Section 1.8.4.1) to identify inter-cultivar polymorphism and design molecular markers for the genetic mapping of these genes discussed in Chapter 4. Further, the design of AS-primers for each of the *PDI* genes in this Chapter has allowed the physical mapping of the *PDI* genes of wheat to chromosomal 'bins' and preliminary investigations into orthology between rice and wheat discussed in Chapter 5.

CHAPTER 4

DEVELOPMENT OF MOLECULAR MARKERS AND GENETIC MAPPING OF THE *PDI* GENES

List of Figures and Tables in Chapter 4

Chapter 4 Development of Molecular Markers and Genetic Mapping of the *PDI* Genes

ABSTRACT

The characterisation of the *PDI* gene family in the previous chapter has permitted the analysis of these genes in seven other cultivars used in the development of genetic maps. Partial sequencing of the three *PDI* genes identified a single SNP in each of the genes on chromosomes 4AL and 4BS, however no polymorphism was identified in the *PDI* gene on chromosome 4D after sequencing over 80% of this gene in the eight cultivars. The SNP identified in the *PDI* gene on chromosome 4A was found to be polymorphic between the Cr x Hb, Sc x Ta and CD x Ka crosses, while the SNP identified in the *PDI* gene on 4BS was found to be polymorphic in the CD x Ka, only. These SNPs were used to develop molecular markers for both the 4AL and 4BS genes, exploiting RFLPs for *Sml* and *Bsu*36I, respectively, to identify each allele through the generation of characteristic-sized restriction fragments due to the presence of the SNPs. The molecular markers were subsequently used to screen the doubled haploid progeny of each of the crosses containing the polymorphic SNPs revealing co-segregation of the PDI4A marker (*XvutPDI*) with a number of AFLP and SSR markers localised to chromosome 4A as well as the *GERMIN* marker, linked to an oxalate oxidase gene (*germin*) involved in a number of important traits including, germination, stress response and pathogen defence. Similarly, the PDI4B marker (*XvutPDIb*) was found to be linked to a number of anonymous, random SSR, RFLP and AFLP markers localised to chromosome 4B and the *XGERMINB* marker, linked to a homeologue of the germin gene described above. A number of studies have revealed QTLs mapped to chromosomes 4A and 4B and future studies will allow the association between these *PDI* markers and these QTLs.

4.1 INTRODUCTION

As stated in Section 1.10 one of the main aims of this project was to develop molecular markers for the *PDI* genes of wheat to permit their integration into wheat genetic maps. This process will provide a number of benefits in elucidating the role of PDI in wheat, particularly, (1) allowing the involvement of PDI in dough processing traits to be dissected through assessment of any association of the genes encoding PDI with QTLs for storage protein deposition and dough mixing properties, (2) identifying other *in vivo* roles for PDI in physiological processes in wheat through the association of *PDI*-specific markers with QTLs for such traits, (3) identifying beneficial alleles of *PDI* influencing traits of agronomic importance, and allowing their rapid selection in breeding programs using marker assisted selection and, (4) as for all new markers, integrating the *PDI* molecular markers into genetic maps will improve the genome coverage of those maps and thus, their effectiveness in the genetic analysis of wheat.

The partial characterisation of three *PDI* genes putatively localised to chromosomes 4A, 4B and 4D in *T. aestivum* cv. Katepwa in Chapter 3 has provided the basis for comparing the homologous regions in the *PDI* genes other cultivars previously used to establish genetic mapping populations (Section 1.8.4.1). The amplification and sequencing of the equivalent sections of these three genes in the cultivars Cranbrook, Halberd, Egret, Sunstar, Sunco and CD87 will allow the identification of any inter-cultivar polymorphism in these genes that can be used to develop *PDI*-specific molecular markers. The inheritance of these markers and thus, the *PDI* genes, can then be determined in the doubled haploid progeny of these crosses to trace their segregation with the numerous RFLP, SSR, AFLP and gene-specific markers present in the genetic maps (Section 1.8.4.1) developed from these crosses to identify any linkage and allow the integration of these genes into the genetic maps. The integration of these genes into the genetic maps would prove useful in analysing the contribution of the *PDI* genes in dough processing traits as all three segregate for such traits: dough extensibility in Sunco x Tasman and CD87 x Katepwa, dough strength in Cranbrook x Halberd and dough strength and extensibility in a genetic background with similar storage protein alleles in Egret x Sunstar (Table 1.2). Thus, this Chapter describes the use of the AS-primers used in the amplification of the *PDI* genes from *T. aestivum* cv. Katepwa to characterise the homologous *PDI* genes in these seven other parental cultivars, the

development of *PDI*-specific molecular markers and genetic mapping of the *PDI* genes of wheat.

4.2 THE *PDI* GENES IN THE PARENTAL CULTIVARS

4.2.1 Isolation and characterization of the *PDI* gene on chromosome 4A

To assess the level of conservation of the *PDI4A* gene in *T. aestivum* and identify any inter-cultivar polymorphism, the homologous regions in the cultivars Cranbrook, Halberd, Egret, Sunstar, Sunco, Tasman and CD87 were amplified with the four pairs of allele-specific primers used in the characterisation of the *PDI* gene in *T. aestivum* cv. Katepwa (Table 2.6, Figure 2.4) producing products similar in size to those generated from Katepwa (Figure 4.1). Direct sequencing of these products allowed the assembly of two contiguous regions in all cultivars, one ranging in size from 1316bp (Sunco) to 1391bp (Katepwa) and the other 458bp (Sunco, Tasman, CD87 and Katepwa) to 506bp (Cranbrook, Halberd, Egret and Sunstar) (Appendix IIIA). The differences in the lengths of these sequences in these cultivars were due solely to trimming of the ends of the sequences due to differences in the quality of the sequence data generated from each and were not due to size polymorphisms between the cultivars. Multiple sequence alignments of the characterised sections in the 8 mapping cultivars revealed only a single inter-cultivar polymorphism in the characterised sections of the *PDI* gene from the A genome of *T. aestivum*, a G-T substitution at a position equivalent to base pair 1850 in *TtPDI4A* that was polymorphic in the Cranbrook x Halberd, Sunco x Tasman and CD87 x Katepwa crosses (Appendix IIIA; Figure 4.3A).

4.2.2 Isolation and characterization of the *PDI* gene on chromosome 4B

To assess the level of conservation of the *PDI4B* gene in *T. aestivum* and identify any inter-cultivar polymorphism, the three allele-specific primer pairs used in the characterisation of this gene in *T. aestivum* cv. Katepwa (Table 2.4, Figure 2.4) were used to amplify the corresponding gene sections from the seven cultivars of *T. aestivum* listed in the previous section (Section 4.2.1) producing PCR products equivalent in size to those amplified from *T. aestivium* cv. Katepwa (Figure 4.2). Direct sequencing of these PCR products allowed the assembly of three contiguous regions of 247bp (Egret) to 291bp (Halberd, Sunco, Tasman, CD87 and Katepwa), 696bp (Sunstar) to 824bp

(Katepwa) and 560bp (Sunco) to 638bp (CD87 and Katepwa) (Appendix IIIB). Again, these size differences were due to quality trimming of the sequences and not true polymorphisms. Multiple sequence alignment of these gene sections from the eight mapping cultivars revealed a single SNP, an A-T substitution at the end of intron 9 at a position equivalent to base pair 3106 in *TtPDI4B* that was polymorphic in the CD87 x Katepwa cross (Appendix IIIB; Figure 4.4B).

4.2.3 Isolation and characterization of the *PDI* gene on chromosome 4D

Six allele-specific primer pairs used to characterise the *PDI4D* gene from *T. aestivum* cv. Katepwa (Table 2.4, Figure 2.4) were used to specifically amplify the homologous sections from the seven cultivars of *T. aestivum* described above (Section 4.2.1) revealing sizes similar to those from *T. aestivum* cv. Katepwa (Figure 4.2). Direct sequencing of these products allowed the construction of four contiguous sections ranging in size from (i) 769bp (Egret) to 839bp (Cranbrook); (ii) 436bp (Tasman) to 488bp (Sunstar); (iii) 765bp (Cranbrook) to 889bp (Sunstar); and (iv) 950bp (Cranbrook) to 1032bp (CD87) (Appendix IIID). Multiple sequence alignment of these gene sections revealed 100% sequence conservation of gene sequences in all cultivars over the characterised regions (Appendix IIID).

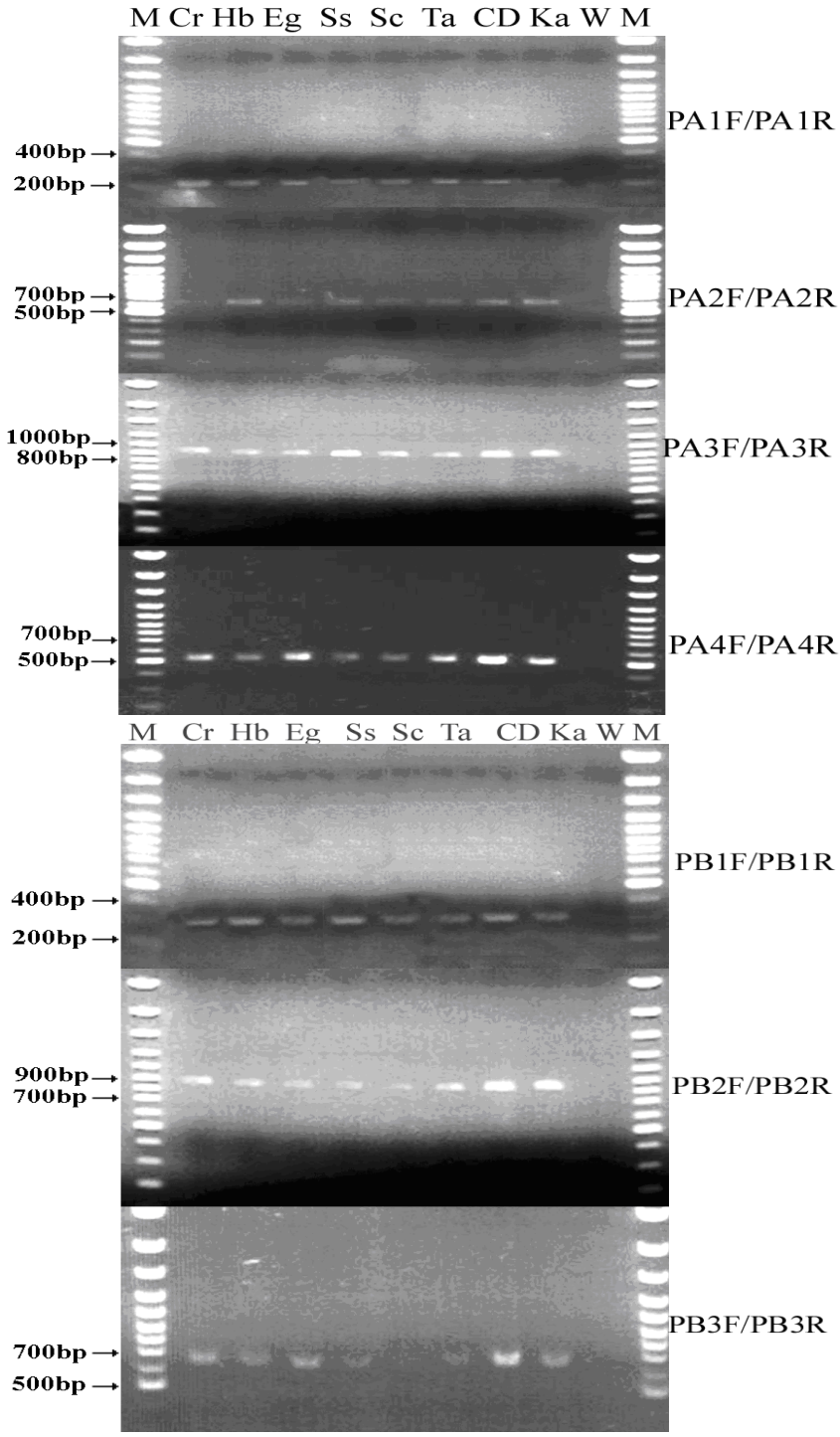


Figure 4.1 Amplified sections of the A and B genome *PDI* genes of *T. aestivum*

The PCR products generated from the A and B genomes of eight cultivars of *T. aestivum*. Primer pairs are listed on the right (Table 2.4 and Figure 2.4). Legend: M: 100bp marker; Cr: Cranbrook; Hb: Halberd; Eg: Egret; Ss: Sunstar; Sc: Sunco; Ta: Tasman; CD: CD87; Ka: Katepwa.

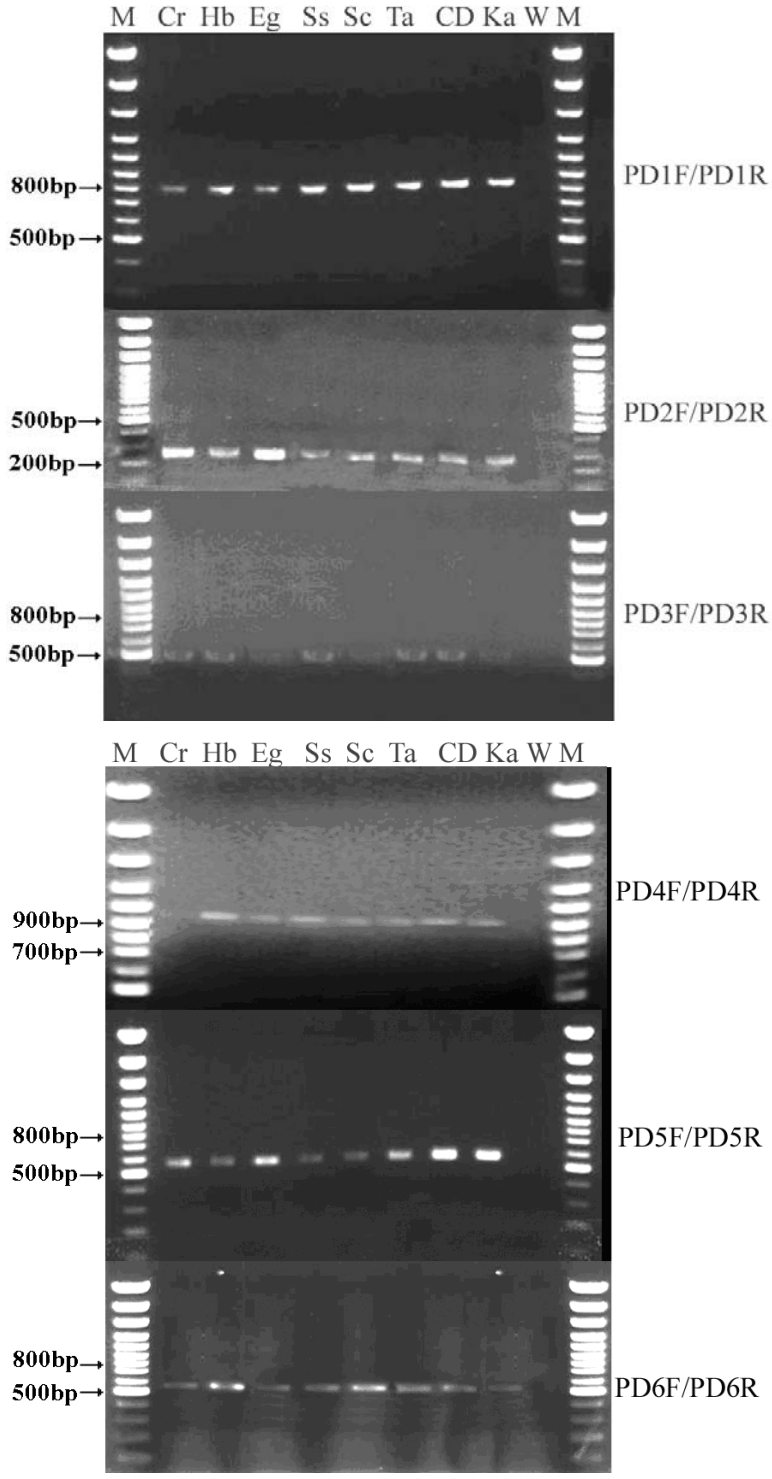


Figure 4.2 Amplified sections of the D genome *PDI* gene of *T. aestivum*

The PCR products generated from the D genome of eight cultivars of *T. aestivum*. Primer pairs are listed on the right (Table 2.4 and Figure 2.4). Legend: M: 100bp marker; Cr: Cranbrook; Hb: Halberd; Eg: Egret; Ss: Sunstar; Sc: Sunco; Ta: Tasman; CD: CD87; Ka: Katepwa.

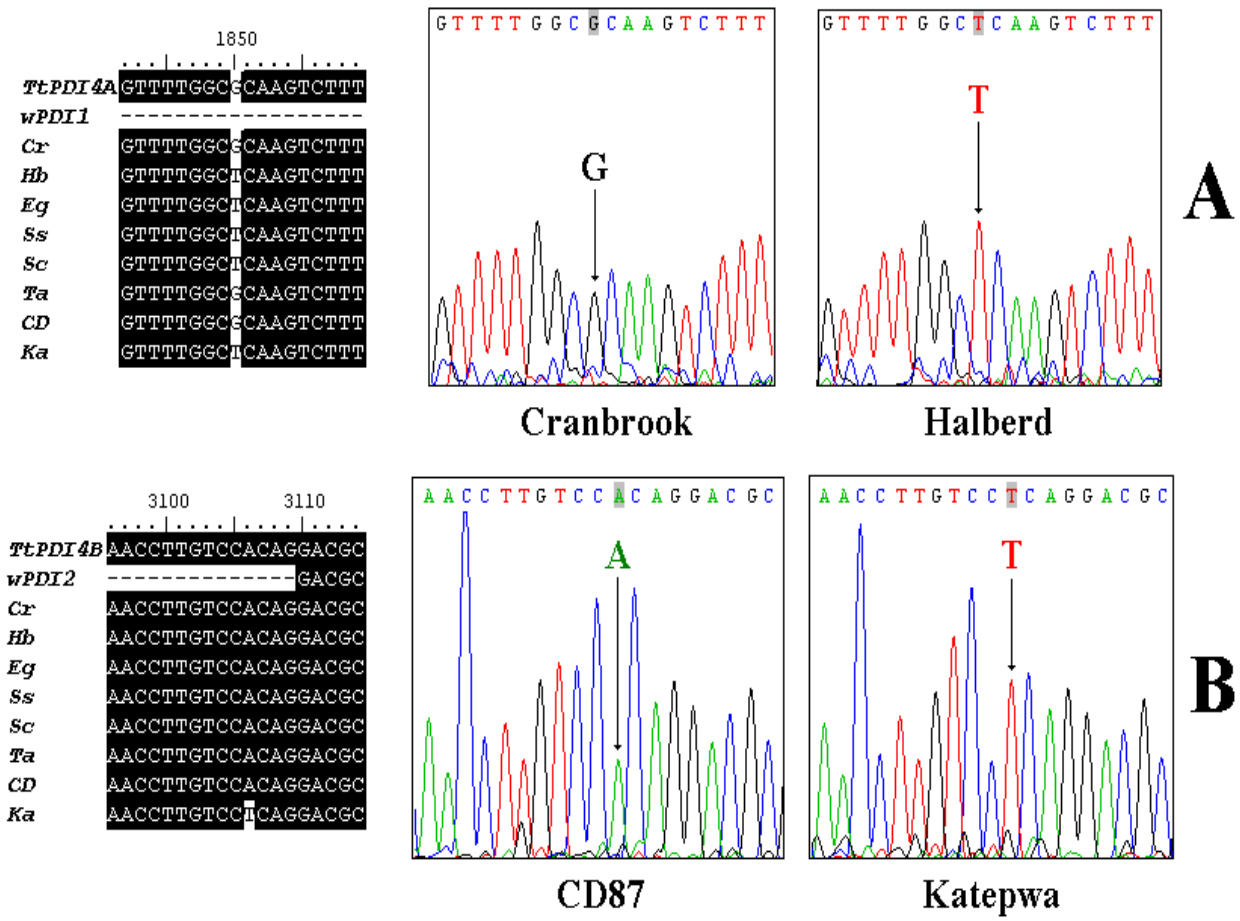


Figure 4.3 Intercultivar polymorphism in the *PDI* genes of wheat

The alignments highlighting the SNPs identified in the *PDI* genes in the A (A) and B (B) genomes of *T. aestivum.* The sequencing trace files show the unambiguous peaks at the polymorphic nucleotides in both cases. The Cranbrook and Halberd sequencing trace files displayed are representative of the similar result obtained for the Sunco x Tasman and CD87 x Katepwa cultivars. The alignments showing the sequenced sections of all cultivars are shown in full in Appendix III.

4.3 DESIGN AND TESTING OF A *TaPDI4A*-SPECIFIC MARKER

The G/T SNP in the 900bp section of the *TaPDI4A* genes amplified with the PA3F/PA3R primer pair from Cranbrook x Halberd, Sunco x Tasman and CD87 x Katepwa (Figure 4.1; Section 4.2.1) was analysed with the restriction mapping program of the BioEdit software package (Section 2.6.3.4). This was undertaken to determine whether the polymorphism resulted in the loss or gain of any restriction sites that could be exploited to distinguish the two alternative genotypes in the mapping populations generated from these crosses. This analysis revealed that the single *Sml*I restriction site at position 357 of the above 900bp PCR product generated from Halberd, Sunco and Katepwa was absent in that amplified from Cranbrook, Tasman and CD87. An alternative reverse primer PA3RB located 87bp downstream of the PA3R used in the amplification of the *TaPDI4A* PCR-RFLP marker was predicted to produce a 987bp product in all of the parental lines which after *Sml*I digestion would be expected to produce an undigested 987bp product in Cranbrook, Tasman and CD87 and 630bp and 357bp products in Halberd, Sunco and CD87 (Figure 4.4A).

The effectiveness of the *Sml*I PCR-RFLP marker was tested on the eight parental cultivars used in the generation of the genetic mapping populations. Amplification of gDNA with the PA3F/PA3RB primer pair resulted in a 987bp product in all lines (data not shown), which, after digestion with *Sml*I, resulted in two bands in Halberd, Egret, Sunstar, Sunco and Katepwa, with a faint band at 987bp, presumably as a result of incomplete digestion of the original PCR product and an intact 987bp product in Cranbrook, Tasman and CD87 (Figure 4.4B).

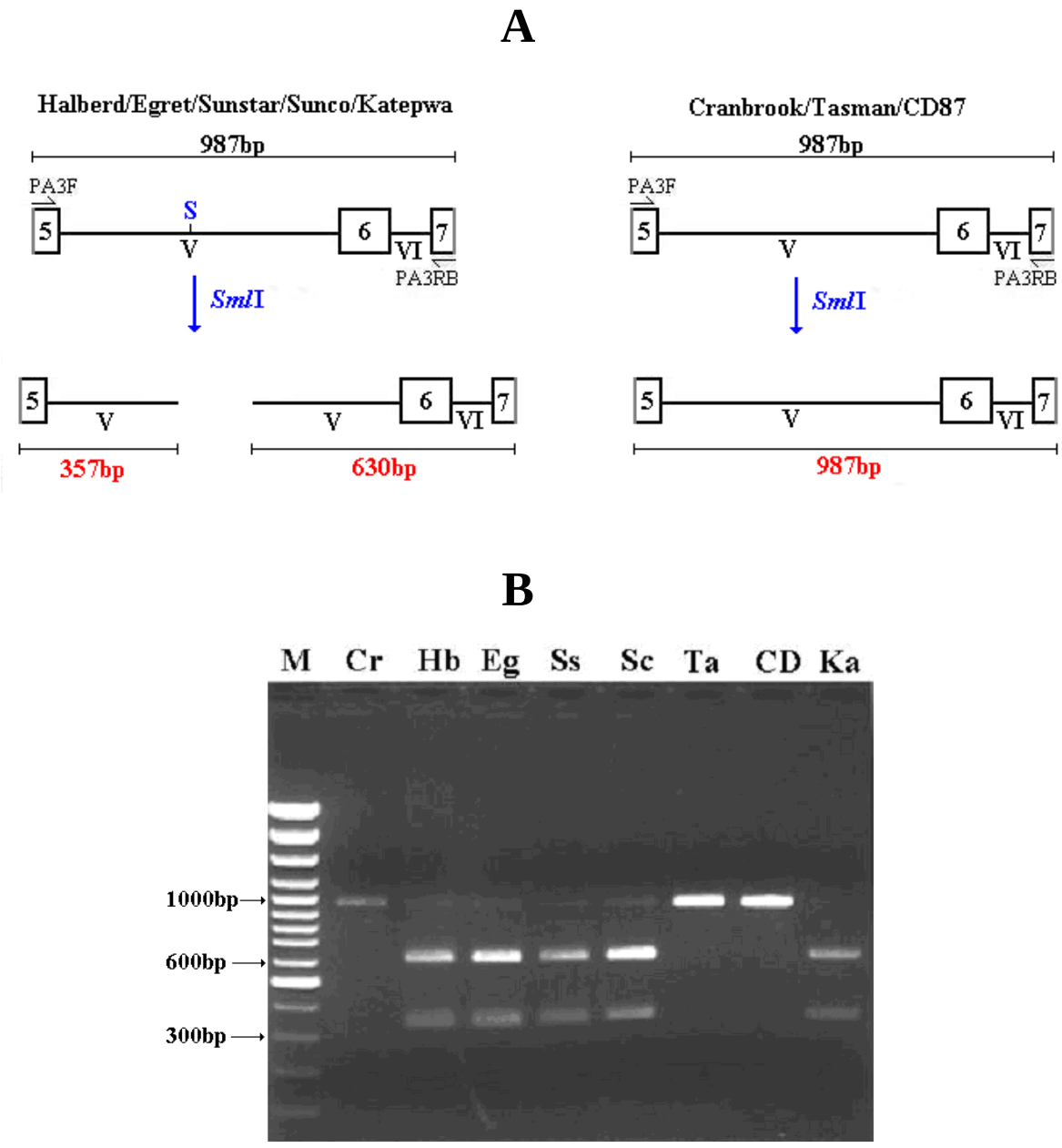


Figure 4.4 Development and testing of the *TaPDI4A* PCR-RFLP marker

A: The principle of the PCR-RFLP markers for the *TaPDI4A* gene. The size of the predicted amplicon (987bp) and *Sml*I restriction site (S) and the predicted sizes of the products of restriction digestion with *Sml*I are shown. The positions of the primers, the exons (5-7) and introns (V and VI) are shown. B: The products of digestion of the PA3F/PA3RB amplicons with *Sml*I from the eight parental cultivars. The different genotypes of the *TaPDI4A* gene in the Cr x Hb, Sc x Ta and CD x Ka crosses are clearly identifiable with this marker. M: 100bp marker.

4.4 DESIGN AND TESTING OF A *TaPDI4B*-SPECIFIC MARKER

In a similar method to that undertaken to identify the presence of any unique restriction sites in the PA3F/PA3R PCR product described above (Section 4.3), the 653bp PB3F/PB3R PCR product, which was found to contain an A/T SNP polymorphic between the CD87 and Katepwa cultivars (Section 4.2.2), was analysed to identify the loss or gain of any restriction sites to distinguish the two alternative genotypes in the mapping population derived from this cross. The restriction maps revealed a conserved *Bsu*36I restriction site at position 169 of both 653bp PCR products, however the A/T SNP in the PCR product from Katepwa introduced a second *Bsu*36I site in this PCR product at position 551 (Figure 4.5A). Thus, a PCR-RFLP assay was devised that would identify the CD87 genotype as the presence of 484bp and 169bp restriction fragments after digestion of the PB3F/PB3R amplicon with *Bsu*36I and the Katepwa genotype as the presence of 382bp, 169bp and 102bp restriction fragments after *Bsu*36I digestion (Figure 4.5A).

The effectiveness of the *Bsu*36I PCR-RFLP marker was first tested on all eight parental cultivars used in the generation of the genetic mapping populations. Amplification of gDNA with the PB3F/PB3R primer pair resulted in the predicted 653bp product from all cultivars (data not shown), which, after digestion with *Bsu*36I resulted in the generation of the 382bp fragment from Katepwa only, however the predicted 169bp and 102bp fragments could not be visualised on the agarose gel (Figure 4.5B). In contrast, the PCR products from the 7 other cultivars were digested to produce the predicted 484bp and faint 169bp fragments, in addition, the 653bp, undigested product was still readily visible, again presumably due to partial digestion of the PCR product (Figure 4.5B).

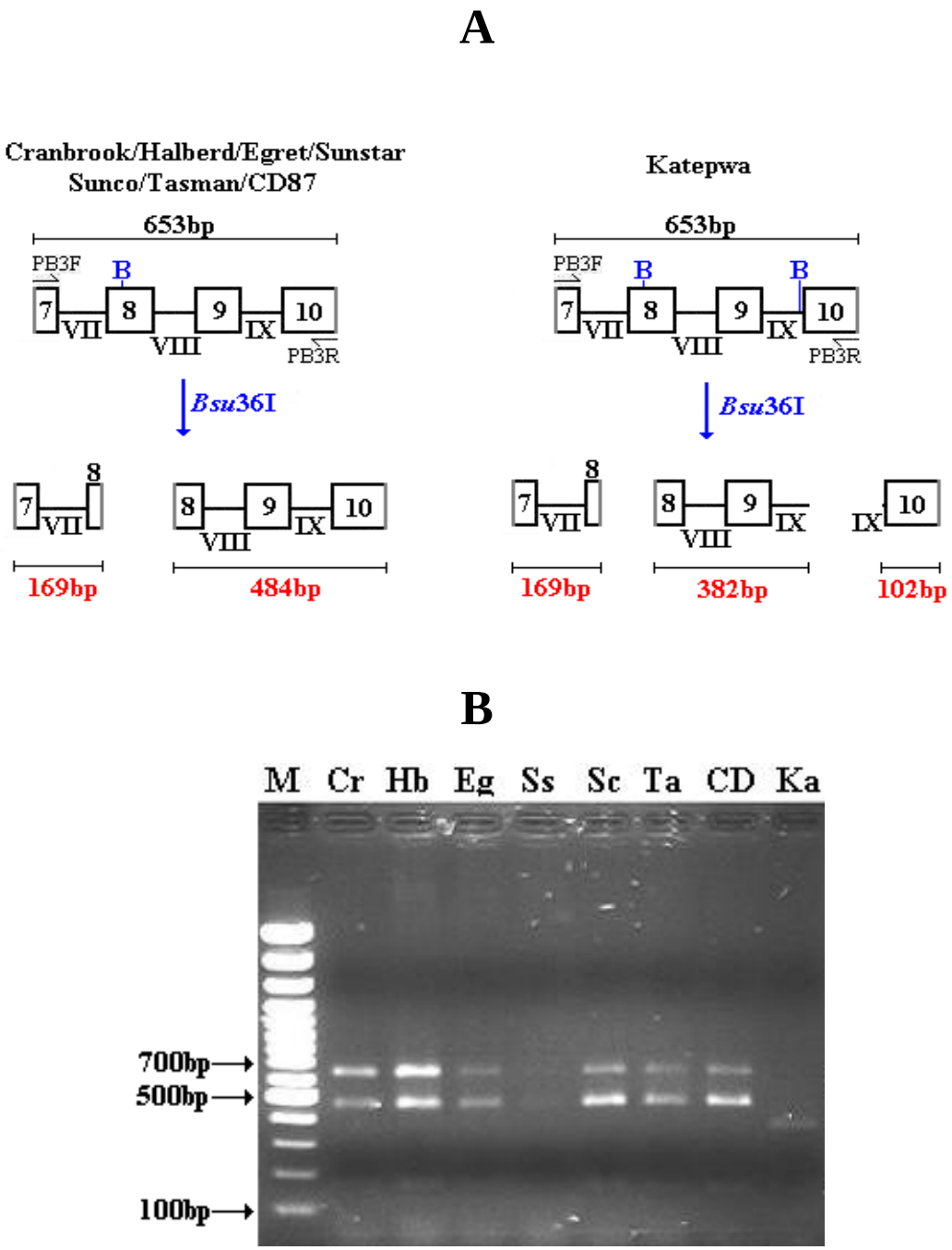


Figure 4.5 Development and testing of the *TaPDI4B* PCR-RFLP marker

A: The principle of the PCR-RFLP markers for the *TaPDI4B* gene. The size of the predicted amplicon (653bp) and internal *Bsu*36I restriction sites (B) and the predicted sizes of the products of restriction digestion with *Bsu*36I are shown. The positions of the primers, the exons (7-10) and introns (VII-IX) are shown. B: The products of digestion of the PB3F/PB3R amplicons with *Bsu*36I from the eight parental cultivars. The different genotypes of the *TaPDI4B* gene in the CD x Ka cross are clearly identifiable with this marker. M: 100bp marker.

4.5 GENOTYPING OF *PDI* MARKERS IN MAPPING POPULATIONS

As the *PDI* molecular markers were found to effectively identify the genotypes of the *TaPDI4A* and *TaPDI4B* genes in the parental cultivars, they were subsequently tested on the doubled haploid genetic mapping populations generated from these crosses (Section 2.14.2), to identify any linkage between these genes and other markers genotyped in the same populations (Section 1.8.4.1).

4.5.1 Genotyping the *TaPDI4A* marker

The *TaPDI4A* marker was amplified from gDNA isolated from the 160 doubled haploid progeny of the Cranbrook x Halberd cross with the PA3F/PA3R primers and digested with *Sml*I resulting in the genotyping of this PCR-RFLP in 154 of the 160 available progeny and the absence of a PCR product in the remaining 6 progeny (Figure 4.6).

Likewise, the assay of the 180 doubled haploid progeny from the Sunco x Tasman cross revealed very clear differentiation of the *TaPDI4A* genotypes in 174 of the 180 progeny, with the absence of a PCR product in the remaining 6 progeny (Figure 4.7).

The *TaPDI4A* PCR-RFLP marker was also used to genotype the 180 available doubled haploid progeny of the CD87 x Katepwa cross, revealing a much lower success rate than above two crosses (Figure 4.8) with only 60 of the 180 lines positively genotyped and the failure of the PCR in 120 of the progeny. As such, the assay was repeated to discount experimental errors in the original assay, however, a lower success rate than that obtained for the other crosses was again obtained, improving the total genotyped progeny in the two assays to 84 of the 180 available with apparent failure of the PCR in the 96 remaining progeny (Figure 4.8).

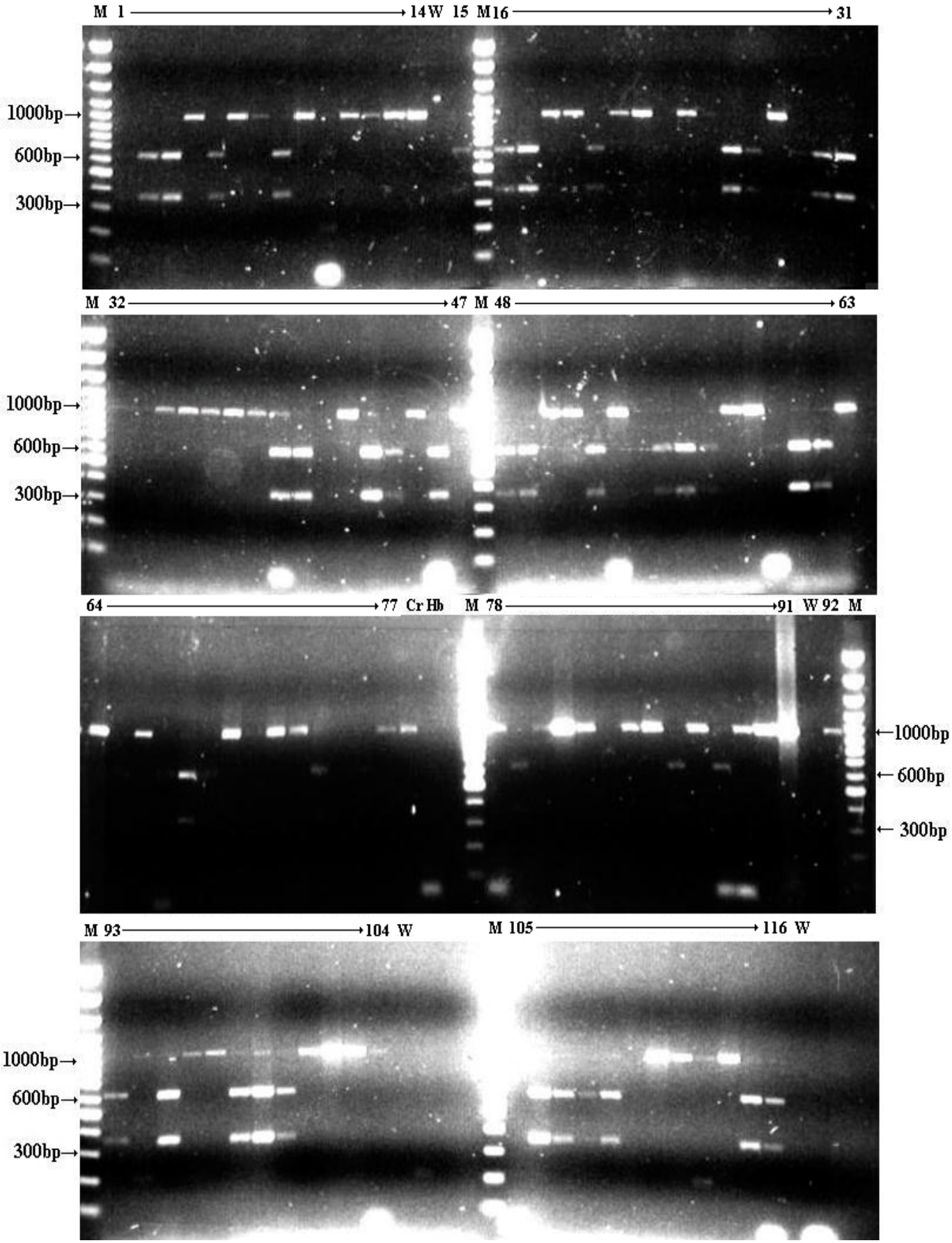


Figure 4.6 To be continued.

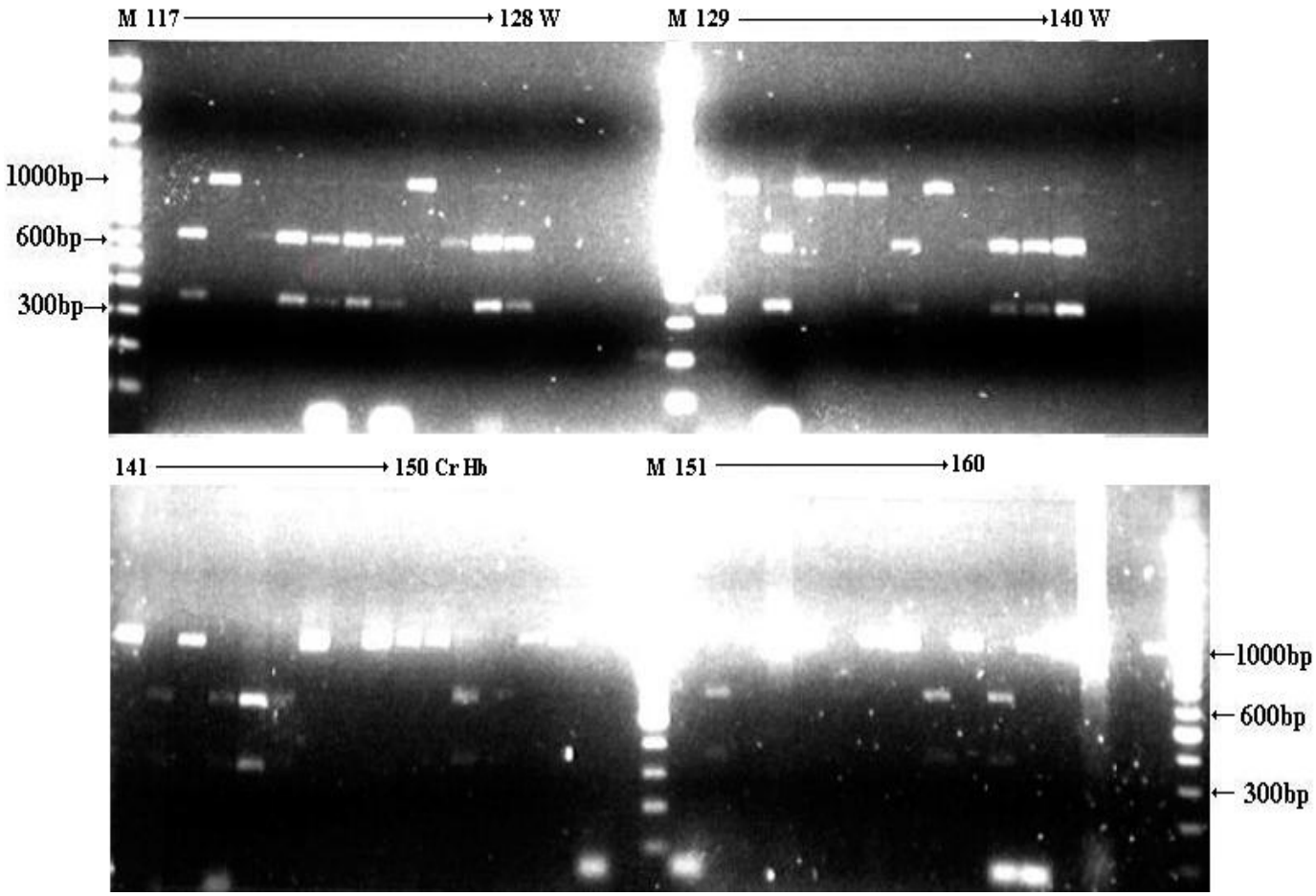


Figure 4.6 The *TaPDI4A* marker assayed in the Cr x Hb DH progeny

The data generated from the 160 doubled haploid progeny of the Cr x Hb cross assayed for the *TaPDI4A* PCR-RFLP marker. The figure shows the results of restriction digestion of the PA3F/PA3RB PCR product from 160 DH progeny of Cr x Hb with *Sml*I. 1-160: various DH progeny of the Cr x Hb cross, M: 100bp marker, W: No template negative control, Cr: Cranbrook gDNA, Hb: Halberd gDNA.

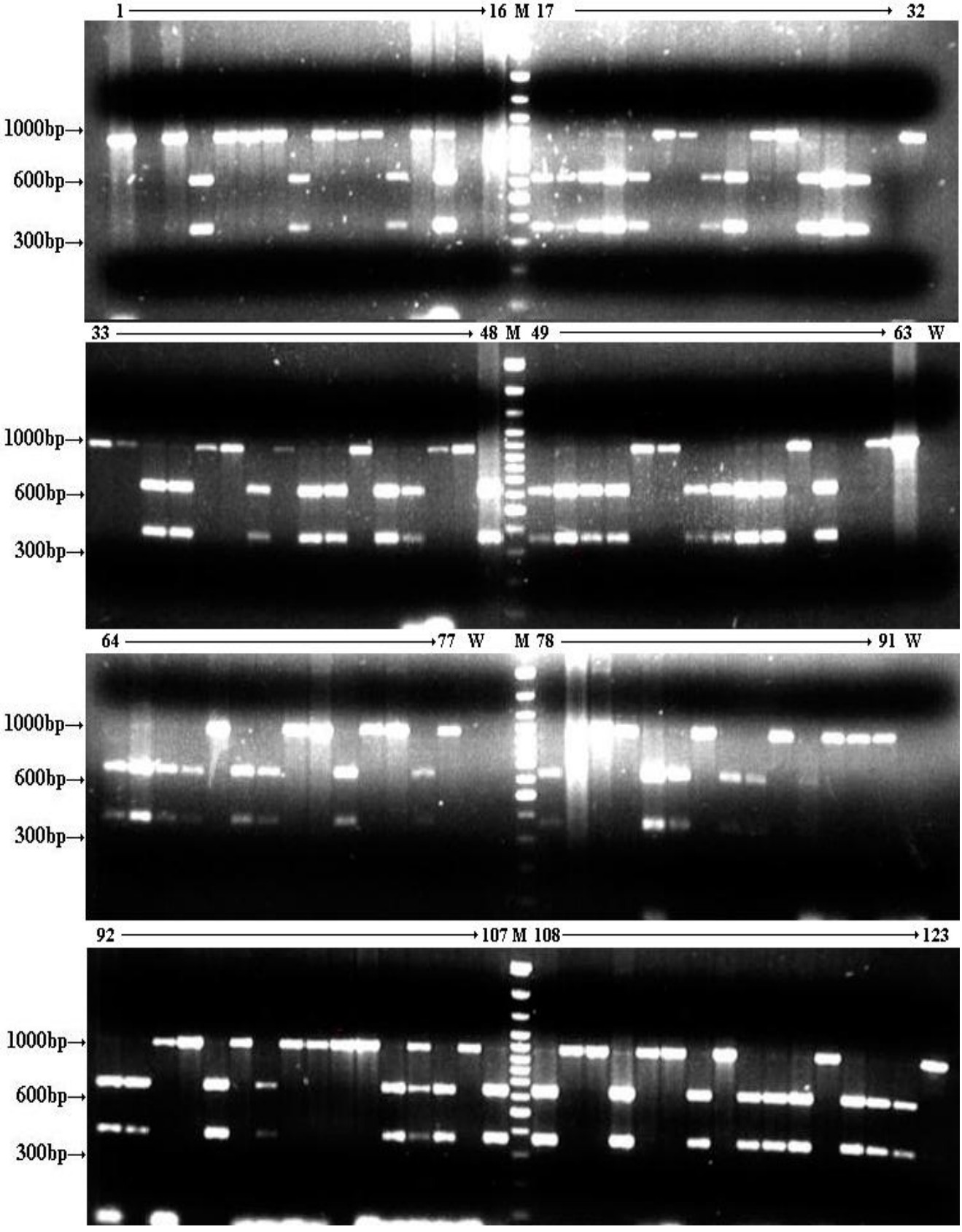


Figure 4.7 To be continued.

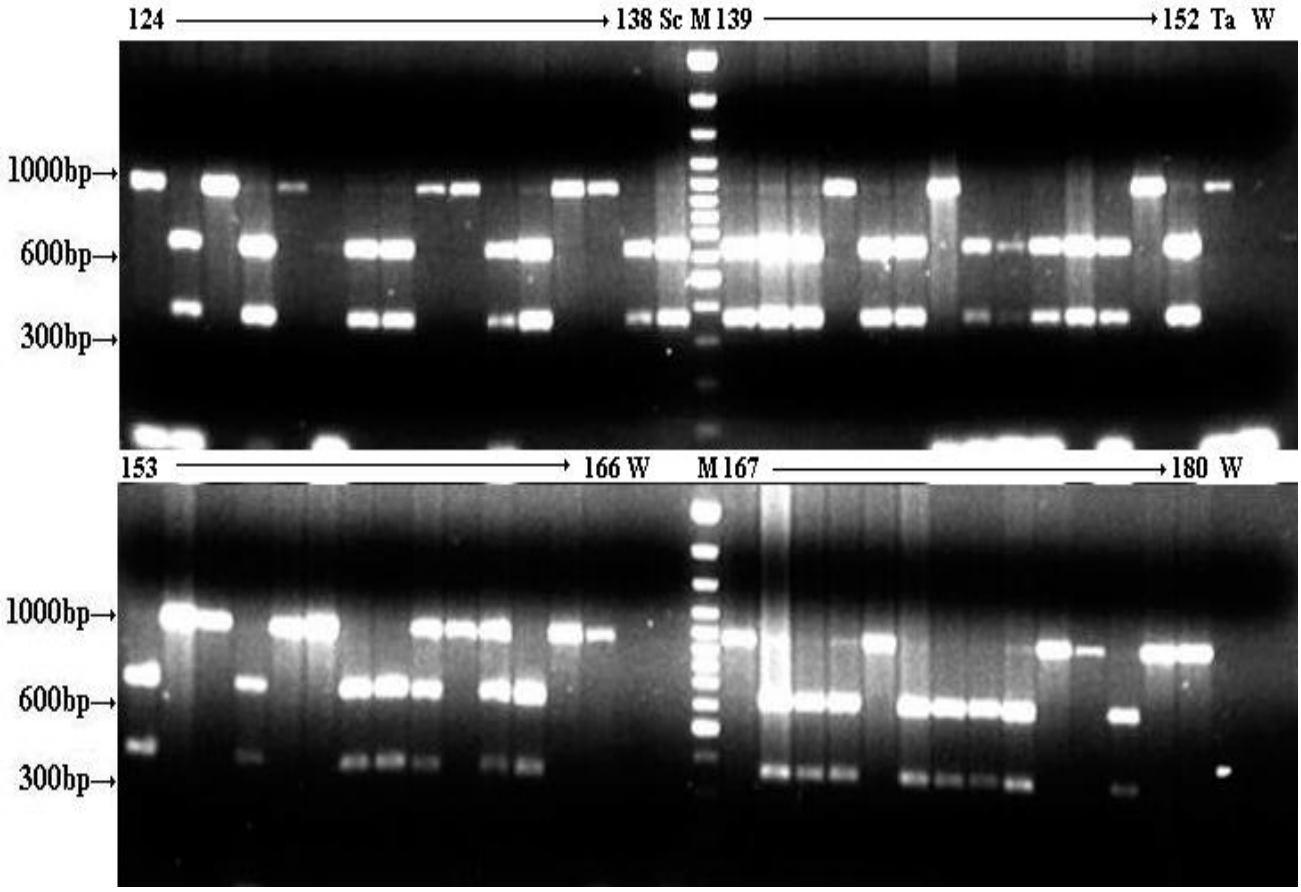


Figure 4.7 The *TaPDI4A* marker assayed in the Sc x Ta progeny

The data generated from the 180 doubled haploid progeny of the Sc x Ta cross assayed for the *TaPDI4A* PCR-RFLP marker. The figure shows the results of Sc x Ta with *Sml*I. 1-180: various DH progeny of the Sc x Ta cross, M: 100bp marker, W: No template negative control, Sc: Sunco gDNA, Ta: Tasman gDNA.

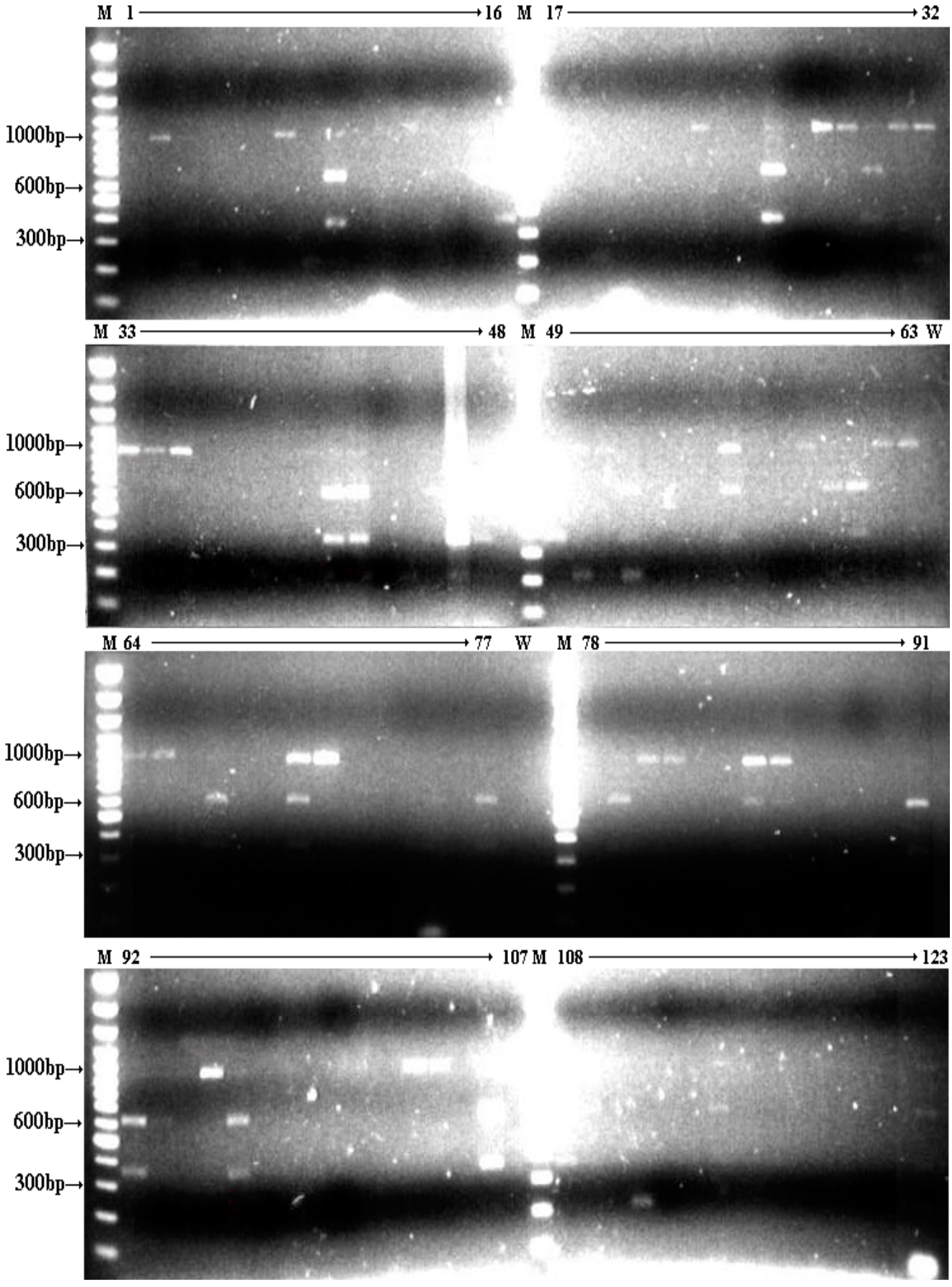


Figure 4.8 To be continued.

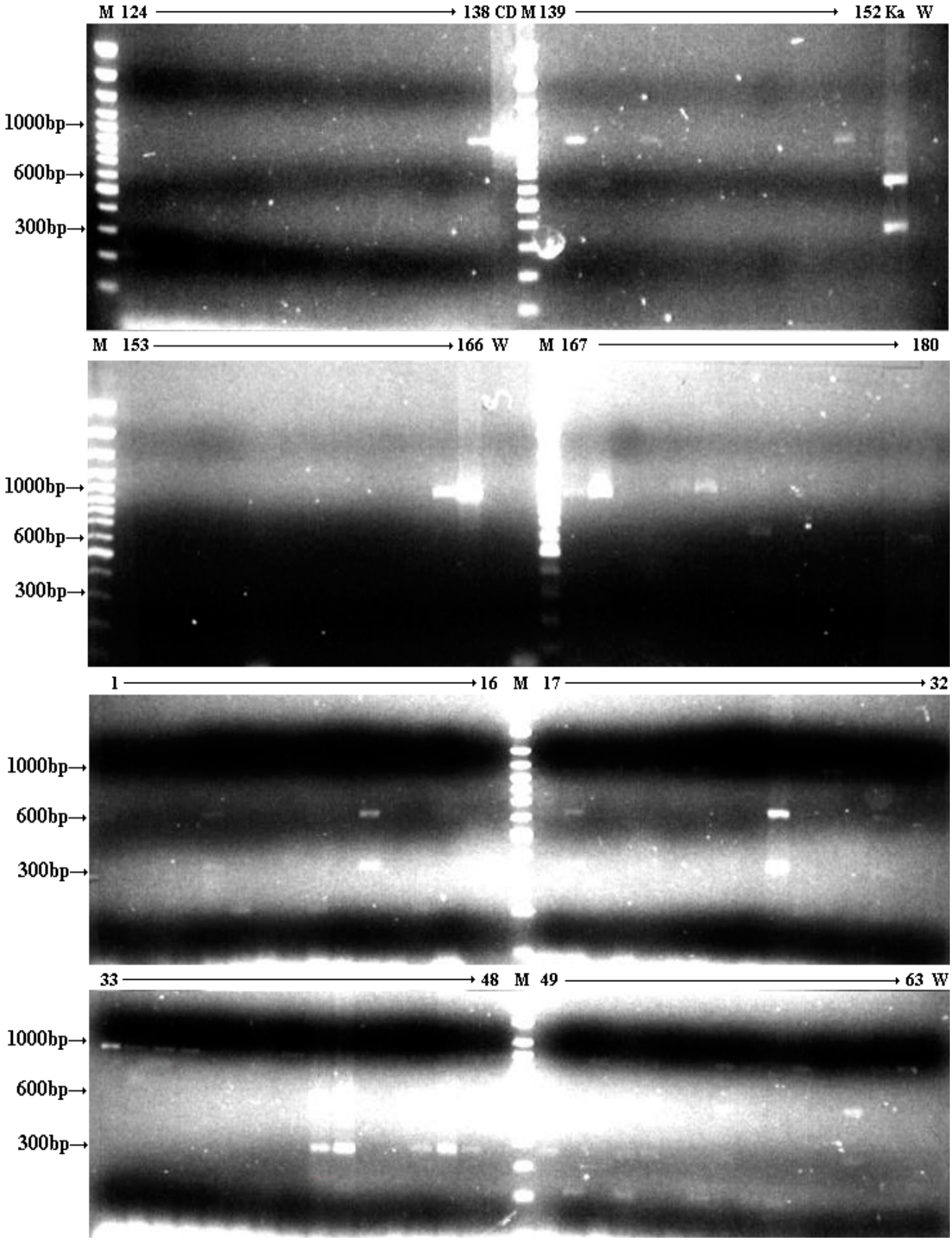


Figure 4.8 To be continued.

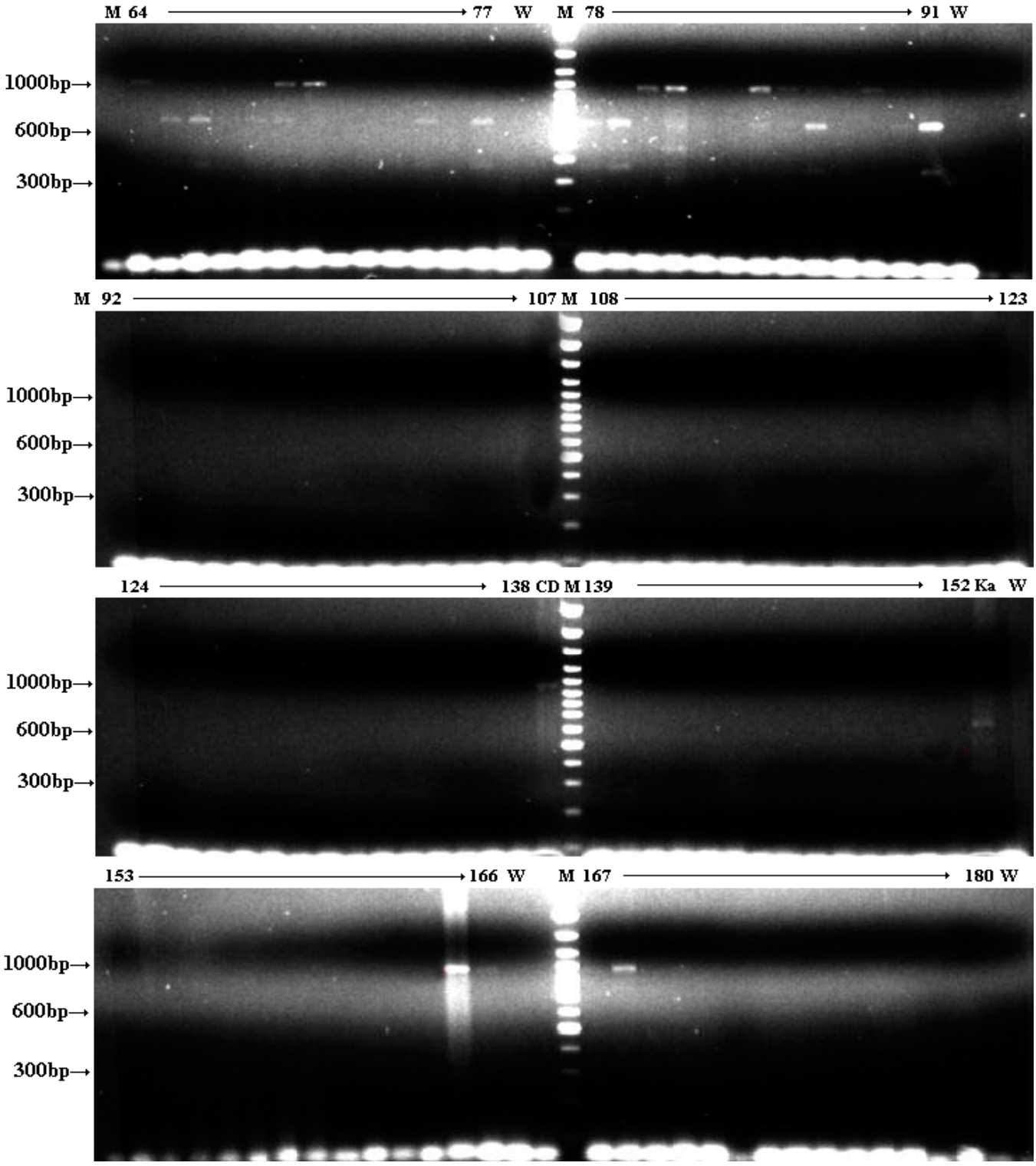


Figure 4.8 The *TaPDI4A* marker assayed in the CD x Ka progeny

The data generated in duplicate from the 180 DH progeny of the CD x Ka cross assayed for the *TaPDI4A* PCR-RFLP marker. The figure shows the results of restriction digestion of the PA3F/PA3RB PCR product from 180 DH progeny of CD x Ka with *Sml*I. 1-180: various DH progeny of the CD x Ka cross M: 100bp marker, W: No template negative control, CD: CD87 gDNA, Ka: Katepwa gDNA.

4.5.2 Genotyping the *TaPDI4B* marker

Amplification of the gDNA of the doubled haploid progeny of the CD87 x Katepwa cross with the PBF3/PBR3 primer pair and subsequent restriction digestion with *Bsu*36I resulted in the genotyping of 118 of 180 progeny and the absence of PCR products in the remaining 62 progeny lines (Figure 4.9).

4.6 LINKAGE MAPPING OF *TaPDI4A* AND *TaPDI4B*

Analysis of the above marker data for the *TaPDI4A* gene (*XvutPDI*) with the Map Manager QT software package (Section 2.14.2) allowed the integration of this gene-specific marker into the genetic maps generated from the Cr x Hb, Sc x Ta and CD x Ka crosses (Figure 4.10). In agreement with our results in Chapter 3 in conjunction with the findings of Ciaffi *et al*. (1999; 2001; Section 1.5.5), the gene mapped to chromosome 4A in all crosses and further, was found to be closely linked to the *GERMIN* marker (Lane *et al.* 1993) and a number of anonymous markers, especially the *Xwmc48a* SSR marker (Harker, *et al.* 2001), in all of the crosses.

Similarly, the analysis of the marker data generated for the *TaPDI4B* gene (XvutPDIb) allowed its integration into the genetic map generated from CD x Ka cross (Figure 4.11). Although fewer markers had been scored in these progeny at the time of analysis (12/03), the *XvutPDIB* marker was found to be linked to the *XGERMINB* marker (Lane *et al.* 1993) and a number of anonymous markers on chromosome 4B (Figure 4.11).

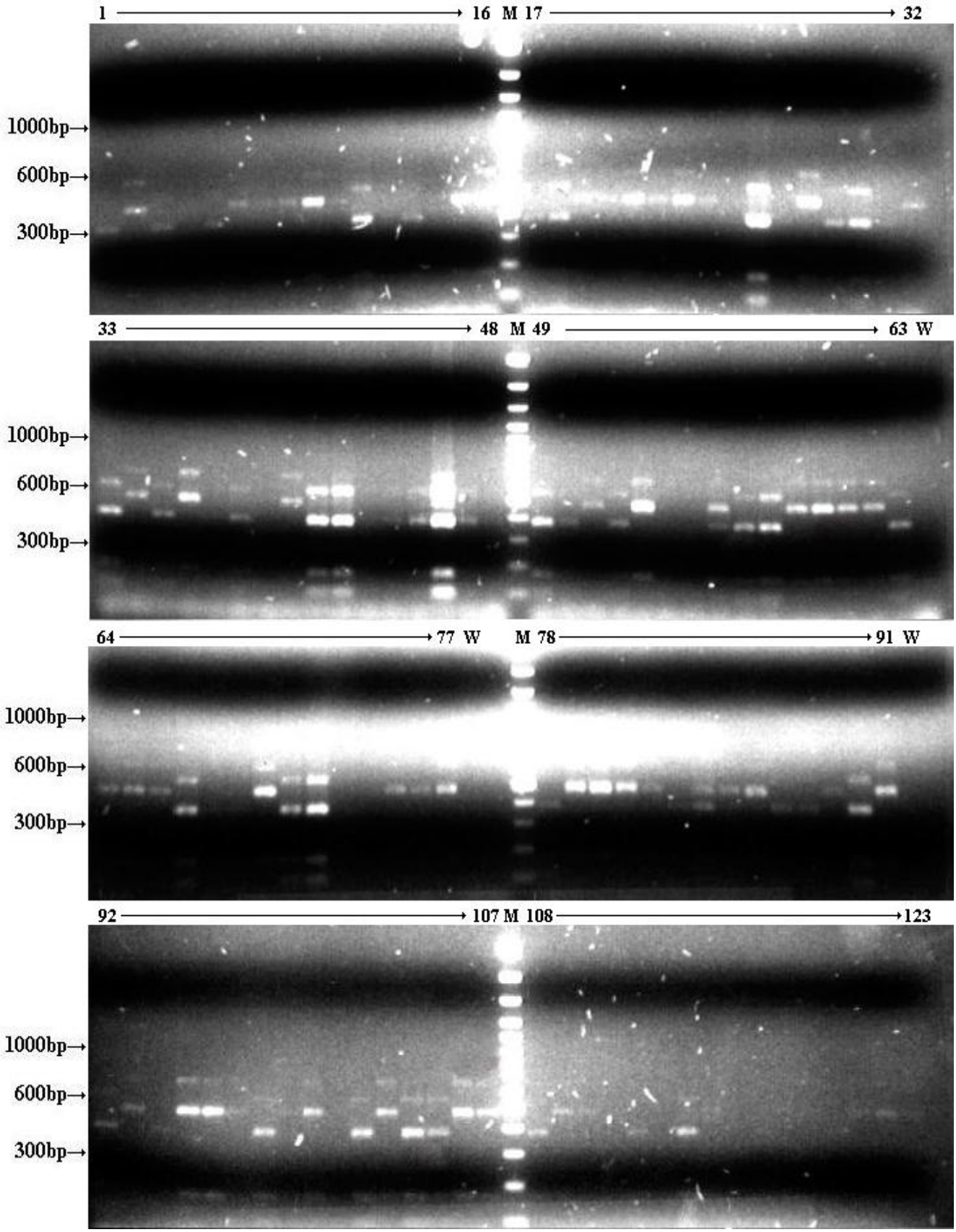


Figure 4.9 To be continued.

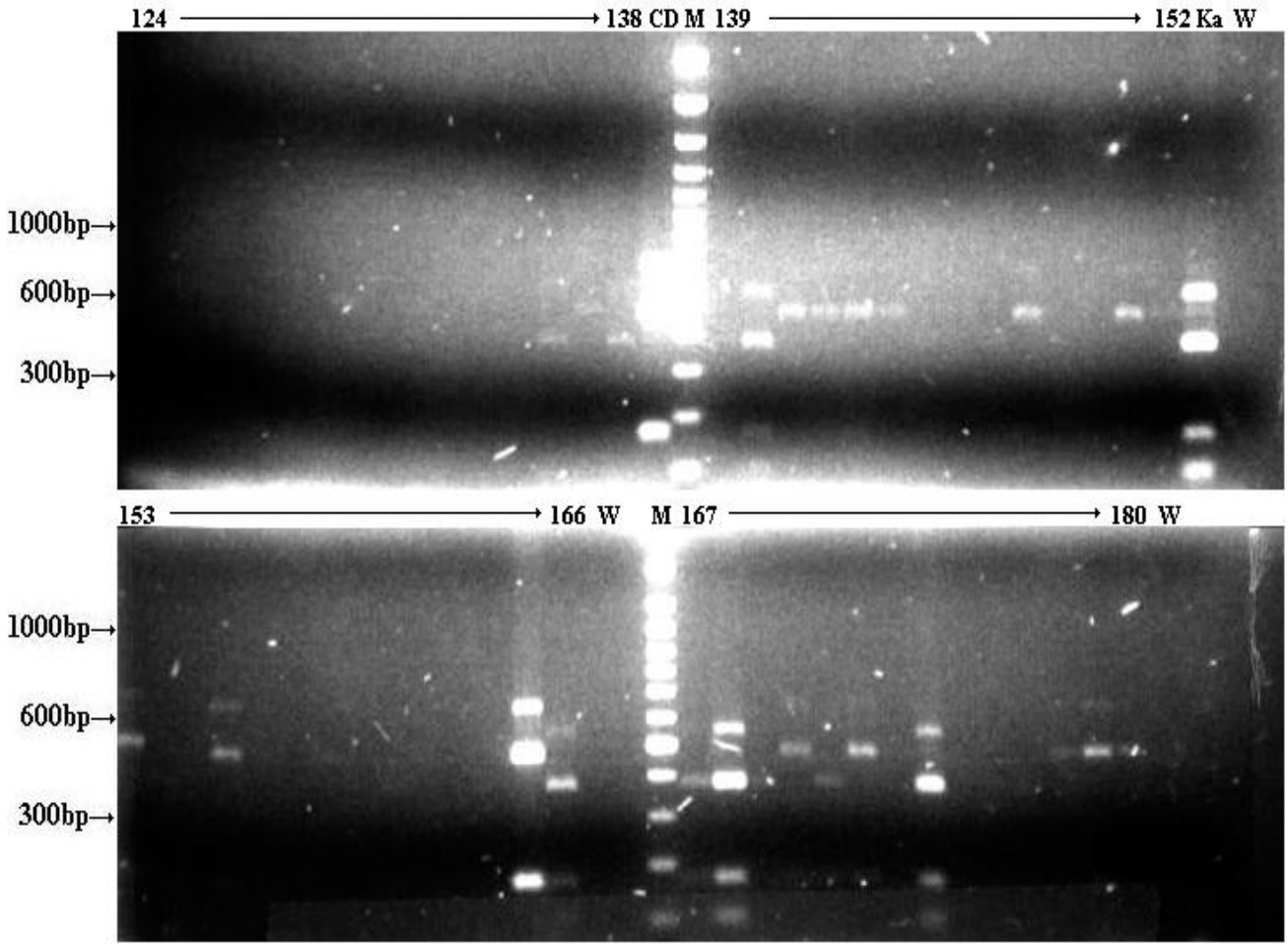


Figure 4.9 The *TaPDI4B* marker assayed in the CD x Ka DH progeny

The data generated from the 180 doubled haploid progeny of the CD x Ka cross assayed for the *TaPDI4B* PCR-RFLP marker. The figure shows the results of restriction digestion of the PB3F/PB3R PCR product from 180 DH progeny of CD x Ka with *Bsu*36I. 1-180: various DH progeny of the CD x Ka cross M: 100bp marker, W: No template negative control, CD: CD87 gDNA, Ka: Katepwa gDNA.

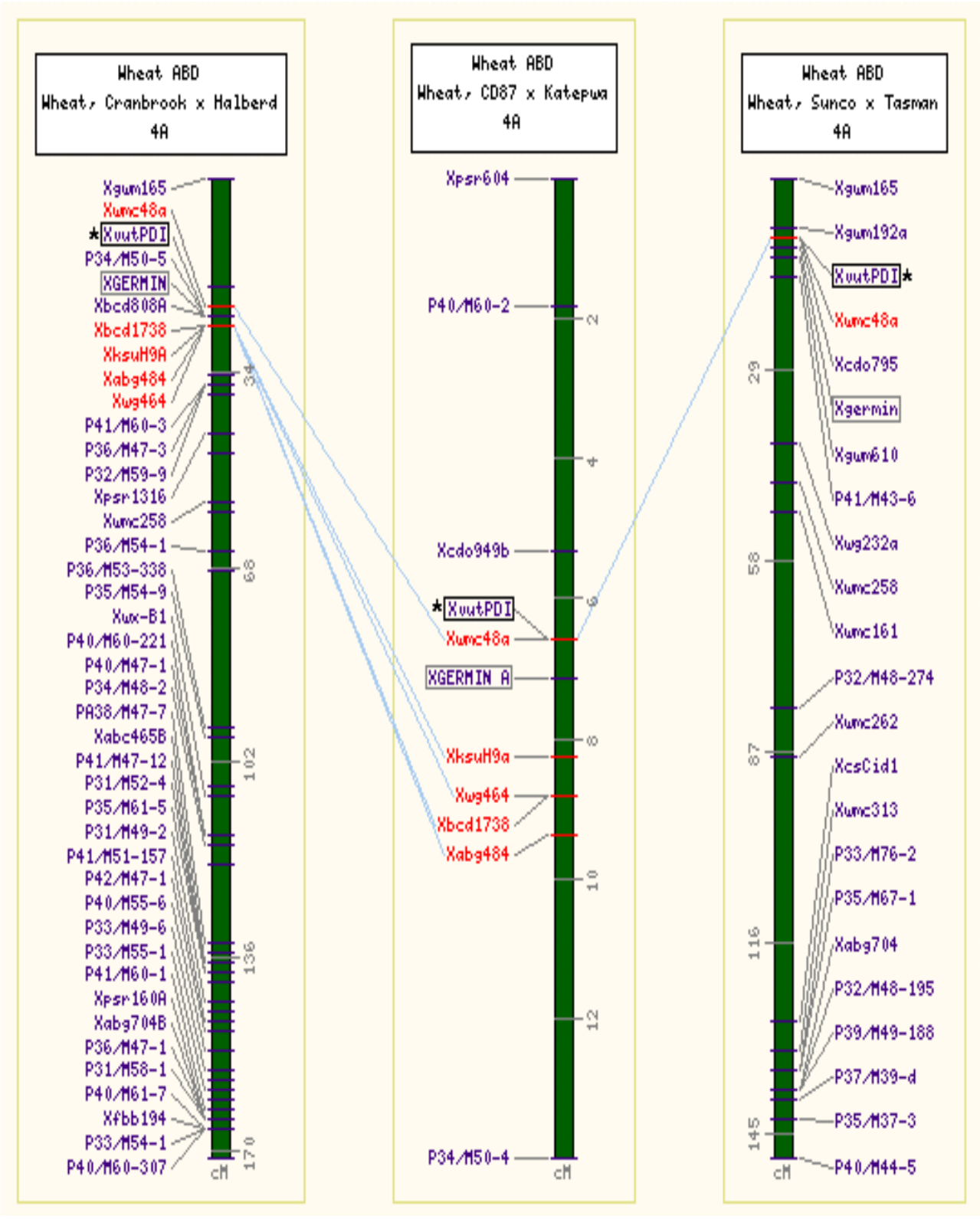

Wheat ABD
Wheat, Cranbrook x Halberd
4A
Xgum165
Xumc48a
*XvutPDI
P34/M50-5
XGERMIN
Xbcd808A
Xbcd1738
XksuH9A
Xabg484
Xwg464
P41/M60-3
P36/M47-3
P32/M59-9
Xpsr1316
Xumc258
P36/M54-1
P36/M53-338
P35/M54-9
Xwx-B1
P40/M60-221
P40/M47-1
P34/M48-2
PA38/M47-7
Xabc465B
P41/M47-12
P31/M52-4
P35/M61-5
P31/M49-2
P41/M51-157
P42/M47-1
P40/M55-6
P33/M49-6
P33/M55-1
P41/M60-1
Xpsr160A
Xabg704B
P36/M47-1
P31/M58-1
P40/M61-7
Xfbb194
P33/M54-1
P40/M60-307
cM
34
68
102
136
170
Wheat ABD
Wheat, CD87 x Katepwa
4A
Xpsr604
P40/M60-2
Xcdo949b
*XvutPDI
Xumc48a
XGERMIN A
XksuH9a
Xwg464
Xbcd1738
Xabg484
P34/M50-4
cM
2
4
6
8
10
12
Wheat ABD
Wheat, Sunco x Tasman
4A
Xgum165
Xgum192a
XvutPDI*
Xumc48a
Xcdo795
Xgermin
Xgum610
P41/M43-6
Xwg232a
Xumc258
Xumc161
P32/M48-274
Xumc262
XcsCid1
Xumc313
P33/M76-2
P35/M67-1
Xabg704
P32/M48-195
P39/M49-188
P37/M39-d
P35/M37-3
P40/M44-5
cM
29
58
87
116
145

Figure 4.10(A)

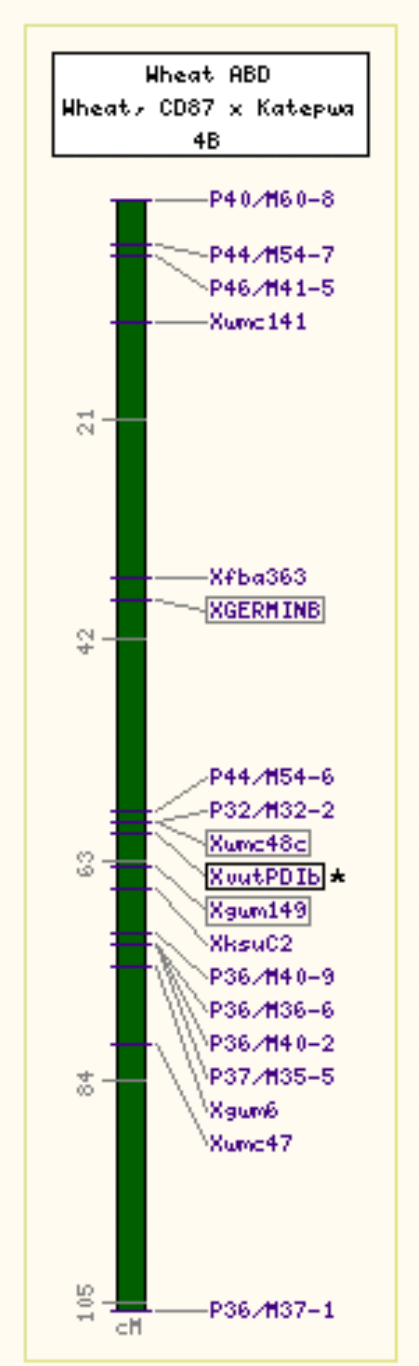


Figure 4.10(B)

Figure 4.10 Comparative linkage maps from (A) chromosome 4A and (B) chromosome 4B of *T. aestivum*

The PCR-RFLP markers for the *TaPDI4A* and *TaPDI4B* genes are shown in boxes with an asterisk and denoted as *XvutPDI* and *XvutPDIb*, respectively. Markers common to the different crosses are shown connected by lines. The *Germin* markers, linked to the *PDI* genes, are boxed, as are the flanking markers defining part of a wheat QTL on chromosome 4B for reaction to head scab. Distances (in cM) and marker names are shown on the right and left of the map, respectively.

4.7 DISCUSSION

4.7.1 The design of markers for the *TaPDI4A* and *TaPDI4B* genes

The identification of an inter-cultivar polymorphism in the *PDI* genes putatively located on chromosomes 4A and 4B through our analysis of the *PDI* gene family (Sections 3.7.1 and 3.7.2) and the work of Ciaffi *et al.* (1999; 2001) has permitted the development of allele-specific PCR-RFLP markers to differentiate the alleles from different cultivars. The identification of a unique *Sml*I site in a gene section amplified from the *TaPDI4A* gene allowed a restriction digestion profile to be generated easily from each of the two alleles. Likewise, the unique *Bsu*36I restriction site in the gene section amplified from the *TaPDI4B* gene also permitted a straight-forward differentiation of the different alleles. Although both markers were based on a similar premise, eg. assaying for the presence of a SNP in a PCR product with a restriction enzyme, the *TaPDI4A* marker had a limitation that it could only distinguish between digested and undigested products. Thus, there was a small possibility of any band of the original size (987bp) being the result of unsuccessful digestion (due to experimental errors) rather than a true allele lacking the restriction site. Therefore, every possible care had to be taken to ensure that the restriction enzyme was functional and digested the expected products. The chances of such an occurrence were minimised through extended restriction digestion incubation times (4 hours) and the use of total genomic DNA of the parental lines to act as positive controls. In contrast, the *TaPDI4B* marker had the benefit of digesting both alleles, producing characteristic patterns of digestion. Thus, failure of the digestion was easily recognised as a 653bp product and the two alleles could be easily identified through the presence of the expected sized bands, especially the clearly visible 484bp or 382bp bands in the two alternate alleles (Figures 4.5 and 4.9).

The analysis of over 80% of the *PDI* gene putatively localised to the D genome (Chapter 3) in the eight cultivars of *T. aestivum* used in the development of genetic maps (Kammholz *et al*. 2001) failed to identify any polymorphism in this gene between these cultivars (Section 4.2.3). This compares to the identification of a single SNP in the *PDI* genes localised to the A and B genome when only ~50% of these genes were characterised (Sections 4.2.1 and 4.2.2). This finding is in agreement with reports of lower levels of polymorphism within the D genome of *T. aestivum*, as reflected in the

lower numbers of markers mapped to this genome when random, anonymous markers are used in the assembly of genetic maps (Section 1.8.4.1) and has led to concerted efforts to develop D-genome specific markers (Pestova *et al.* 2000). Thus, future efforts to develop a molecular marker for the *PDI* gene in the D genome of *T. aestivum* may require looking for polymorphisms in the genomic sequences flanking this gene.

4.7.2 Genetic mapping of two *PDI* genes

The genotyping of the *XvutPDI* (*TaPDI4A*) marker in the doubled haploid progeny of the Cr x Hb, Sc x Ta and CD x Ka crosses permitted the identification of segregation of this marker with a number of anonymous (RFLP, SSR and AFLP) and gene-specific (ie. GERMIN) markers previously analysed in these progeny (Chalmers *et al*. 2001). Thus, the *XvutPDI* marker was successfully incorporated into linkage groups localised to chromosome 4A in all three of these crosses (Chalmers *et al.* 2001), providing genetic evidence supporting the chromosomal location of this gene (Chapter 3; Ciaffi *et al.* 1999; 2001). Likewise, scoring of the *XvutPDIb* (*TaPDI4B*) marker in the doubled haploid progeny of the CD x Ka cross produced sufficient data to identify segregation of this *PDI* marker with a number of anonymous AFLP (eg. *P32/M32-2*), SSR (eg. *Xgwm149*) and gene-specific (*XGERMINB*) markers (Chalmers *et al.* 2001). Thus, the *XvutPDIb* marker could be incorporated into the linkage group localised to chromosome 4B, supporting the putative chromosomal location of this gene (Chapter 3; Ciaffi *et al.* 2001). It must be noted that the results of scoring of both the *XvutPDI* and XvutPDIb markers in the CD x Ka cross produced relatively lower success rates compared to the other crosses (Figures 4.8 and 4.9) and, in the case of the *XvutPDI* marker, required a repeat of the assay to increase the number of scored markers (Figure 4.8). The problem was most likely due to the quality of the template DNA isolated from the doubled haploid progeny of the CD x Ka cross, as the *XvutPDI* marker assay was successful in the Sc x Ta and Cr x Hb screens and lower success rates were observed in *both* of the assays using this gDNA.

Further analysis of the markers linked to the *TaPDI4A* and *TaPDI4B* genes revealed that on the CD x Ka and Sc x Ta chromosome 4A maps, the *xgermin* (*XGERMIN A*) marker was found to be linked to the *TaPDI4A* gene and on the chromosome 4B map, the *XGERMINB* marker appears linked to the *TaPDI4B* gene. Thus, the *PDI* genes

appear to be closely associated with another gene of agronomic importance, as the *Germin* genes encode a family of extracellular matrix oxalate oxidases that are involved in the production of H_2O_2 through the oxidative breakdown of oxalate. Germins were first utilised as a marker of early plant germination, as they are synthesised at the onset of growth of germinating wheat embryos, where they are believed to be involved in embryo hydration and restriction of cell growth through cross-linking of cell walls (Lane *et al*., 1993; Caliskan and Cuming, 1998). They have subsequently been found to be important in a number of other processes, including pathogen defence and stress tolerance in wheat (Berna and Bernier, 1999; Patnaik and Khurana, 2001; Lane, 2002).

4.7.3 QTLs mapping to the group 4 chromosomes of wheat

The genetic mapping of these two *PDI* genes has opened the possibility of determining their association with QTLs of agronomic importance. There have been a number of reports of QTLs for important traits mapped to the chromosome arms carrying the mapped *PDI* genes of wheat, 4AL and 4BS, in the populations developed by Kammholz *et al.* (2001) (Section 1.8.5.1). Specifically, chromosome 4AL has been shown to have QTLs involved in starch characteristics including starch viscosity and flour swelling (Batey *et al*. 2001). This chromosome arm also carries a QTL involved in pre-harvest sprouting (Mares and Mrva, 2001; Mares *et al.* 2002). Recent research on this pre-harvest sprouting QTL has found that the *XvutPDI* marker flanks this QTL and has proven useful as a flanking marker in following the inheritance of this locus in wheat crosses (Çakir *et al.* 2003). Chromosome 4BS has been found to have QTLs involved in many important traits including plant height (Batey *et al*. 2001; Mares and Campbell, 2001; Rebetzke *et al.* 2001), leaf 1 length (+GA), internode length, coleoptile length (Rebetzke *et al.* 2001), flour colour, grain size (Mares and Campbell, 2001), hardness (Osborne *et al*. 2001) and % of B-type starch granules (Batey *et al*. 2001). It is unclear at this stage as to whether the *XvutPDIb* marker co-segregates with these QTLs. Relatively few QTL studies have focussed on dough processing traits in wheat, however QTLs for dough strength have been identified on chromosomes 1A, 5D and 3B (Perretant *et al*. 2000) and QTLs for a number of dough processing traits were identified including chromosomes 4A and 4D (Zanetti *et al.* 2001). Interestingly, recent work by Ma *et al.* (in press, 2005) has revealed a significant epistatic interaction between the *GluB1* locus on chromosome 1B and the 4AL region near the centromere where the *PDI*

has been localised. This observation is consistent with the predicted significance of the PDI locus and further work is needed to develop this further. Future analysis of these QTLs will identify any association between the *PDI* markers and as more QTL studies are conducted in the future, especially those focussed on dough processing traits, any association between any of these new QTL and the *PDI* genes of wheat can be identified and any functional association between these enzymes and such traits can be further elucidated.

CHAPTER 5

INVESTIGATIONS INTO ORTHOLOGY BETWEEN THE *PDI* LOCI OF WHEAT AND THE *ESP2* LOCUS OF RICE

List of Figures and Tables in Chapter 5

Chapter 5 Investigations into Orthology Between the *PDI* Loci of Wheat and the *esp2* Locus of Rice

ABSTRACT

Reports on the identification and characterisation of a rice mutant, esp2, revealed the presence of aberrant storage protein deposition in conjunction with the absence of PDI expression. This report, thus, provides evidence of a direct role for PDI in storage protein depositions and thus the presence of a hypothetical QTL for such a trait at the *esp2* locus of rice. As orthology has been established between many of the crop plants, it was possible that this hypothetical QTL for storage protein deposition may be present in wheat also. As the esp2 mutant exhibited a lack of PDI expression it was conceivable that the mutation was directly to the *PDI* gene or flanking regulatory sequences. A comparison of the *esp2* locus and the *PDI* gene of rice revealed they were present at similar positions in the rice genome on the short arm of chromosome 11 distal to the S20163S RFLP marker. In order to determine whether flanking genes present at the *PDI* locus of rice were also present at similar loci in wheat an analysis of a 1Mb region of the TIGR rice genome assembly 2004 on chromosome 11 containing the *PDI* gene to identify putative expressed wheat orthologs present in the TIGR TaGI database revealed 34 tentative consensus (TC) sequences with high sequence similarity to sequences at the *PDI* locus of rice. Subsequent analysis revealed that 9 of the ESTs used in the assembly of these TC sequences had been sequenced from 5 probes used in physical mapping experiments as part of the US Wheat EST project. Two of these probes were specific for the *PDI* gene of wheat and the three others for genes putatively orthologous to those flanking the rice *PDI* gene. Analysis of this data revealed that all five probes identified loci in the group 4 chromosomes of wheat; however the reported physical loci for the two *PDI* probes were contradictory, with different loci on chromosome 4AL. To clarify this contradiction the *PDI* genes of wheat were localised to chromosomal ‘bins’ using AS-PCR, confirming their presence on, and refining their locations to bins on the proximal section of chromosome 4AL and the distal sections of chromosomes 4BS and 4DS. Thus, the presented data provides preliminary evidence of orthology between the *PDI* loci of rice and wheat and therefore the presence of a QTL involved in storage protein deposition at the *PDI* loci of wheat.

5.1 INTRODUCTION

One of the potentially important roles of the *PDI* gene family in wheat is their involvement in the folding and deposition of the storage proteins in the developing endosperm. As explained in Section 1.5.6, this role could be critical in determining the end-use quality characteristics of different wheat cultivars, as allelic variations or differences in expression patterns of the *PDI* genes could potentially have downstream effects on the size and/or organization of the protein bodies in the endosperm and thus, on the rheological characteristics of the dough. The characterization of the *PDI* gene family in wheat (Chapter 3) and subsequent development of molecular markers and genetic mapping of two of the three *PDI* genes (Chapter 4) has provided the means for the further quantitative genetic analyses of these genes in wheat to elucidate a genetic link between these genes and variations in storage protein deposition and/or the rheological characteristics of dough.

As described in Section 1.5.6, work by Takemoto *et al.* (2002) has provided important evidence supporting a direct role for PDI in the deposition of storage proteins in a grass species. This work found that the chemically-induced rice mutant, esp2, which exhibits aberrant storage protein deposition in the rice grain, was lacking *PDI* expression, providing the first direct link between PDI and the deposition and organisation of the storage proteins, specifically the prolamins, in a grass crop. Such a relationship provides the basis for the presence of a hypothetical QTL involved in the deposition of storage proteins at the *esp2* locus of rice. Whether an orthologous QTL for storage protein deposition is present wheat is currently unclear, however, support for such a locus is provided by the conservation of other QTLs at orthologous loci in barley, wheat, rice and maize for a number of important traits including, vernalisation, flowering time, plant height, dwarfism and shattering (Section 1.9.2). Further, the absence of *PDI* expression in the esp2 mutant suggests that such an orthologous QTL could be at the *PDI* loci in wheat. However, it is yet to be established that the *esp2* locus contains the *PDI* gene in rice, as the absence of *PDI* expression in this mutant could conceivably be due to mutations in *trans*-acting regulators of *PDI* expression. Thus, it first must be established whether the *PDI* gene in rice and the *esp2* locus are present at the same genetic map positions and the physical chromosomal loci of the *PDI*

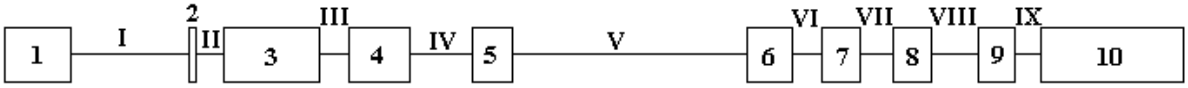
genes in wheat must be firmly established. Our current knowledge of the *PDI* loci in wheat was provided by nullisomic/tetrasomic and ditelosomic analysis by Ciaffi *et al.* (1999; Section 1.5.5) wherein, they identified three loci on chromosomes 4AL, 4BS and 4DS and in some cultivars a fourth locus on chromosome 1B. Our investigations of this fourth locus in Chapter 3 suggested that, if present at all, it is a partial, non-expressed gene. This Chapter describes the investigation of the *esp2* locus of rice, providing evidence that it is at the same position as the rice *PDI* gene and the finer physical mapping of the *PDI* genes of wheat using Chinese Spring chromosomal deletion lines (Section 1.8.3). The use of publicly-available rice and wheat genomic data to allow a comparison of putative genes at the *PDI* locus of rice with the physical map positions of expressed wheat genes is also described, providing preliminary evidence for orthology between these loci and, thus, a QTL for storage protein deposition at the *PDI* loci of wheat.

5.2 COMPARISON OF THE OF RICE *PDI* AND *esp2* LOCI

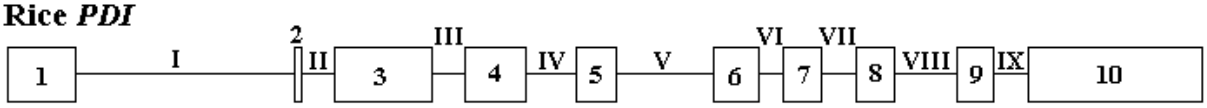
5.2.1 Identification of a rice BAC containing the putative *PDI* orthologue

As a preliminary investigation into the orthology between the *esp2* locus of rice and the *PDI* loci of wheat, a BLASTn search of the Gramene TIGR pseudomolecule assembly release 2 (accessed 1/05) available in the "Rice_genome_japonica_TIGR" database on the Gramene website (www.gramene.org) was conducted using the wheat *PDI* cDNA *wPDI1* as a query sequence. This search identified a ~173kb BAC, OSJNBa0058p12 (AC139170), containing a sequence highly similar to wheat *wPDI1* located between base-pairs 4,960,745 and 4,964,540 on rice chromosome 11 of the rice genome assembly. Alignment of the *wPDI1* and *TtPDI4A* sequences with the orthologous sequences in the rice BAC revealed a conserved 10 exon structure with 56.8% sequence identity, specifically ~77% sequence identity over the *PDI* exon sequences but far less conservation over intron sequences (~40% identity) (Figure 5.1). The ORFs encoded putative proteins of 512 (TtPDI4A) or 515 (rice) amino acids, which exhibited 84.9% identity and 94.2% similarity. The high degree of sequence conservation suggests that this BAC most likely contains the rice orthologue of the wheat *PDI* gene.

Wheat *PDI*



Rice *PDI*



		Length (bp)	
	% Identity	**Rice**	**Wheat**
Exon 1	81.9	204	201
Exon 2	87.1	31	31
Exon 3	83.0	288	285
Exon 4	88.4	189	189
Exon 5	90.5	126	126
Exon 6	84.8	138	138
Exon 7	84.2	120	120
Exon 8	82.2	118	118
Exon 9	80.5	113	113
Exon 10	58.3	512	422
Total Exon	77.1	1839	1743
Intron I	37.1	639	350
Intron II	57.5	105	90
Intron III	56.1	103	93
Intron IV	44.8	141	173
Intron V	30.8	290	684
Intron VI	55.2	82	93
Intron VII	47.7	110	105
Intron VIII	45.3	190	144
Intron IX	42.5	112	85
Total Intron	39.9	1772	1817
Total	56.8	3611	3560
CDS	82.5	1539	1548
Protein	84.9 (94.2)[a]	512[b]	515[b]

Figure 5.1 Sequence comparison of the wheat and putative rice *PDI* genes

Schematic diagrams of the wheat *TtPDI4A* gene and putative rice *PDI* gene on chromosome 11 are shown with exons as boxes connected by introns as lines. Various sequence comparisons between the rice and wheat *PDI* genes are shown below the figure. [a]The number in parentheses is the sequence similarity between the putative proteins. [b]The lengths of the proteins in amino acids.

5.2.2 Comparing the genetic positions of the rice BAC and the *esp2* marker

An investigation to determine whether the *PDI* gene of rice is present at the *esp2* locus was conducted by a search of the marker database on the Gramene website (www.gramene.org) to identify the genetic position of the *esp2* marker in rice. This revealed the *esp2* marker is present on the Hokkaido Morphological 2000 (Morph 2000) map on chromosome 11, 17 cM from the proximal phenotype marker (D53) (Figure 5.2). As the rice BAC containing the putative *PDI* gene (discussed above) also appeared to be present on chromosome 11, the cMap interface of Gramene was used to construct comparative genetic/sequence maps to determine whether this BAC and the *esp2* were located at similar loci. The chromosome 11 Rice Gramene TIGR Assembly sequence map (containing the *PDI* gene sequence on the BAC) and the chromosome 11 Morph 2000 map (containing the *esp2* marker) however did not share correspondences, preventing a direct comparison. As a result, a third map, the JRGP RFLP 2000 genetic map (www.gramene.org) was used, as it was found to contain markers with correspondence to markers on both the above maps (Figure 5.2).

The “sp” phenotype marker located at 38cM on the Morph 2000 chromosome 11 map was found to be present as an interpolated phenotype marker on The IRGP RFLP 2000 chromosome 11 map between 27.1cM and 42.5cM on the short arm (Figures 5.2 and 5.3). This region of the IRGP RFLP map was found to contain at least 10 RFLP markers with correspondences to the Rice Gramene TIGR Assembly sequence map (Figure 5.3). These 10 markers were co-linear on both maps and ranged from the S20163S marker at 27.8cM on the IRGP RFLP 2000 map to the C827S marker at 35.6cM (Figure 5.3). Comparing these genetic positions with their base-pair positions in the rice genome revealed the *S20163S* marker was located between 5,266,027 and 5,266,669bp and the *C827S* was located between 7,071,940 and 7,072,464bp on chromosome 11 (Figure 5.3). Thus, the *PDI* gene on the Rice Gramene TIGR sequence map and the *esp2* marker were found to be located at similar map positions, distal to the sp2, *S20163S* and *C827S* markers. Attempts to identify markers distal to the *esp2* marker on the Morph 2000 map failed to identify any with correspondence to any currently available maps, thus preventing the determination of whether the *esp2* marker and *PDI* gene are flanked by similar markers. However, the presence of the *PDI* gene and *esp2* marker on chromosome 11 at similar map positions, in conjunction with the

characterisation of the *esp2* mutant as lacking *PDI* expression (Takemoto *et al*., 2002), suggests that the *esp2* marker is closely linked to the *PDI* structural gene or *cis*-acting regulatory sequences in rice.

5.3 ORTHOLOGY AT THE *PDI* LOCI OF RICE AND WHEAT

5.3.1 Identification of putative wheat orthologs in the rice BAC sequence

Analysis of the rice genome sequence encompassing its *PDI* gene was conducted to identify any potential wheat orthologues at this locus. The Genome Browser interface of Gramene was used to identify a 1Mb region encompassing bp 4,462,642 - 5,462,642 of rice chromosome 11, flanking the rice *PDI* gene (at bp 4,960,745 - 4,964,540 of chromosome 11; see above). Analysis of wheat sequences present in the TIGR TaGI database (v9.0; accessed 1/05) including tentative consensus sequences (TCs) and EST singletons that are believed to be orthologous with this region of the rice genome, due to high sequence similarity, were identified by selecting the "Wheat_ESTCluster_TGI" feature track on the Gramene Genome Browser. This analysis revealed 42 EST singletons and 34 TCs, including 17 EST singletons and 7 TCs that aligned with the rice *PDI* loci, that were putatively orthologous (Table 5.1). These 42 EST singletons and 34 TCs were constructed from 745 individual EST and cDNA sequences present in the GenBank database, many aligning at the same or overlapping sections (Figure 5.4), The BLASTn alignments of these singletons and TCs (76 in total) with the entire available rice genome sequence revealed e-values (probability of the alignments occurring due to chance) ranging from 7.4e-23 for the 1018bp EST singleton CA707119 to 8.0e-233 for the 2438bp TC207667, with the potentially *PDI*-encoding TCs (including TC220284: discussed below) showing e-values of 7.4e-23 to 2.6e-192 (Table 5.1).

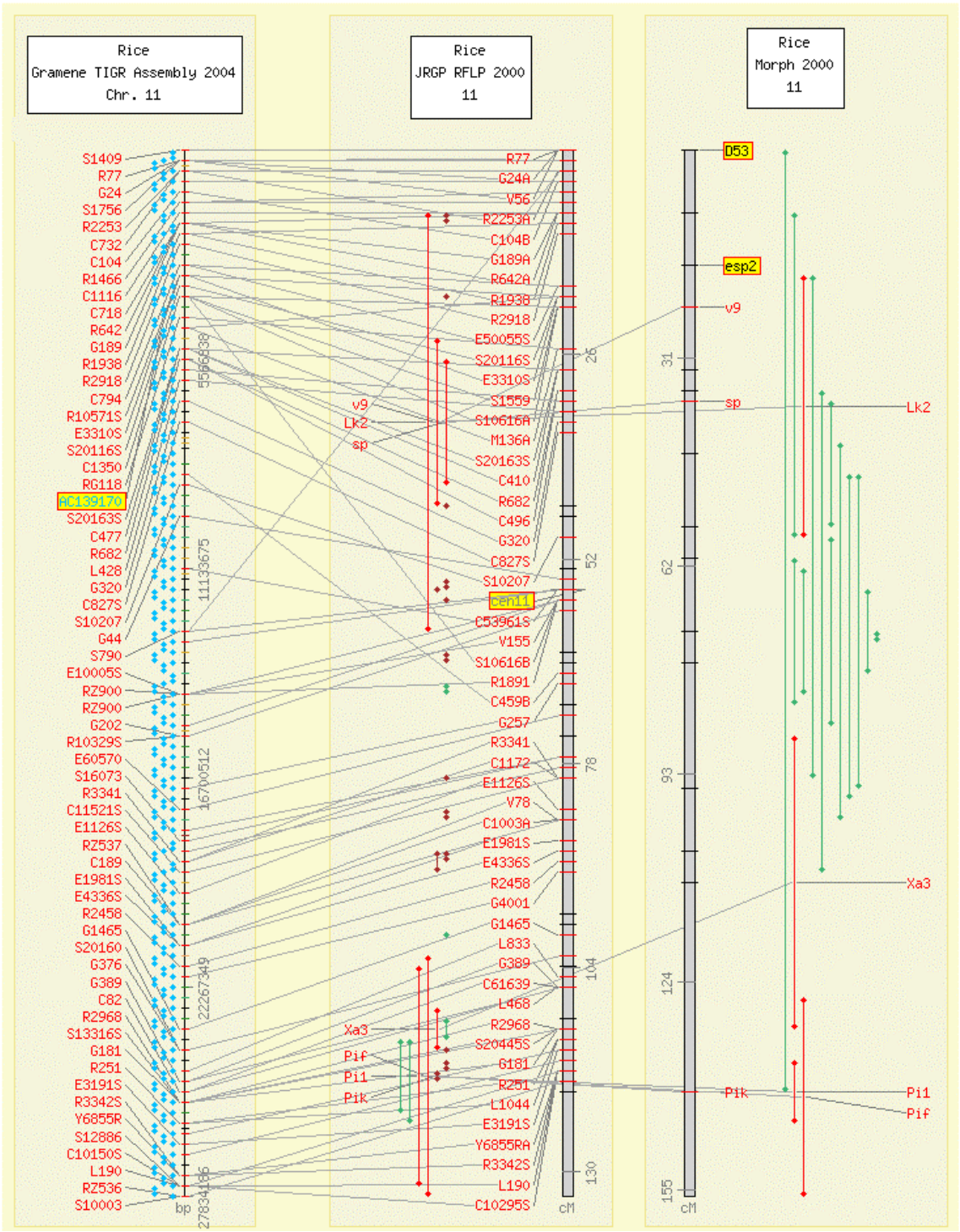


Figure 5.2 Comparing the map positions of *esp2* and *PDI* in rice

A comparison of the *esp2* locus (highlighted) on the Morph 2000 chromosome 11 map with the position of the rice *PDI* gene on the BAC AC139170 (highlighted) in the rice chromosome 11 genome assembly. The chromosome 11 JRGP RFLP containing markers common to the Morph 2000 and genome sequence map is shown. Markers with correspondences are shown in red, interpolated phenotype markers are shown as lines next to the maps, the BACs making up the sequence map are shown in blue. Map units are shown next to the map in cM for the genetic maps or base-pairs for the sequence map. The centromere-specific marker (cen11) is shown highlighted on the JRGP RFLP map. A more detailed comparison of the *esp2* and rice *PDI* gene is shown in Figure 5.3.

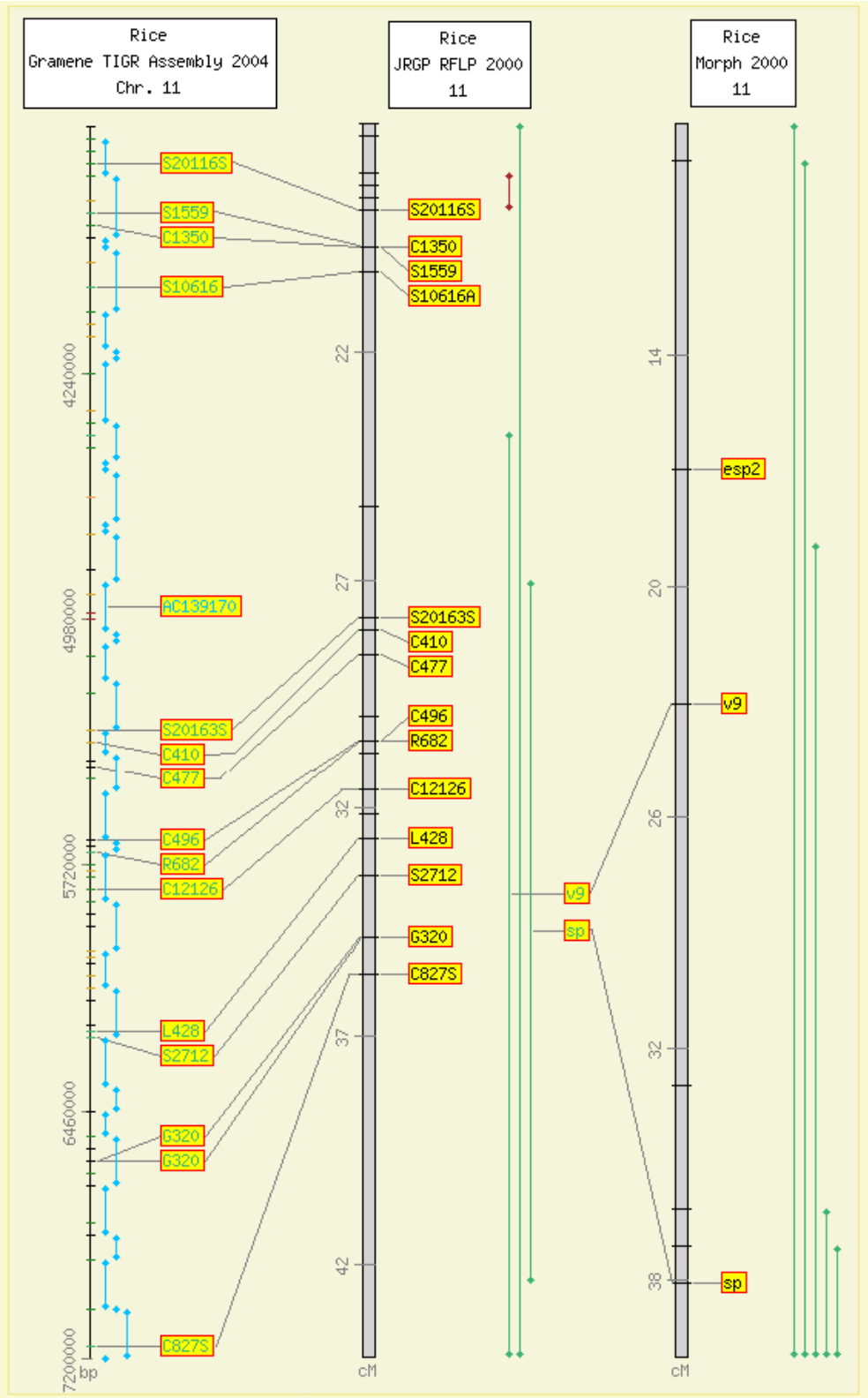


Figure 5.3 A detailed comparison of the *PDI* and *esp2* loci of rice

Sections of the genetic maps in Figure 5.2 enlarged to highlight the position of markers proximal to the *esp2* marker on the Morph 2000 map (sp) with those proximal to the *PDI*-containing AC139170 BAC on the Gramene TIGR Assembly 2004 map. The absence of distal markers common to these genetic maps is also highlighted. Markers with correspondences are shown in red, interpolated phenotype markers are shown as lines next to the maps, the BACs making up the sequence map are shown in blue. Map units are shown next to the map in cM for the genetic maps or base-pairs for the sequence map.

Table 5.1 TIGR TaGI sequences orthologous to the *PDI* locus of rice

TIGR TaGI v9.0[a]	Rice Genome (bp)[b]	BLASTn e-value[c]	# of ESTs[d]	Wheat Map Position[e]	TIGR Tentative Annotation[f]
BJ258989	4543558 - 4546374	2.7e-43	1	None	*similar to* putative cycloartenol synthase {Oryza sativa}, *partial (21%)*
TC196384	4616426 - 4617811	1.5e-104	6	None	*similar to* UP|Q9XI15 (Q9XI15) F8K7.5 protein, *partial (10%)*
TC196709	4618606 - 4621267	2.0e-119	5	None	*homologue to* UP|Q9XI15 (Q9XI15) F8K7.5 protein, *partial (26%)*
CA618761	4724292 - 4726488	1.4e-29	1	None	*homologue to* putative RNA polymerase II subunit hsRPB10 {Arabidopsis thaliana}, *partial (93%)*
TC195799	4727428 - 4729462	5.6e-138	6	None	*weakly similar to* UP|TGR2_HUMAN (P37173) TGF-beta receptor type II precursor (TGFR-2) (TGF-beta type II receptor), *partial (5%)*
TC195800	4727551 - 4728364	6.7e-35	2	None	*similar to* UP|Q8PX92 (Q8PX92) Chemotaxis protein CheW, *partial (10%)*
CA728958	4729600 - 4731045	1.8e-36	1	None	*weakly similar to* protein kinase homolog T14N5.13 - Arabidopsis thaliana, *partial (8%)*
TC214820	4753280 - 4754819	5.3e-33	3	None	*similar to* UP|Q9XI14 (Q9XI14) F8K7.6 protein, *partial (11%)*
TC215350	4760772 - 4762509	3.4e-76	2	None	*similar to* UP|Q9XI14 (Q9XI14) F8K7.6 protein, *partial (8%)*
BQ240481	4778396 - 4779053	7.9e-57	1	None	*weakly similar to* GP|6457331|gb|A phytoalexin-deficient 4 protein {Arabidopsis thaliana}, *partial (7%)*
BM138475	4778453 - 4779098	7.5e-50	1	None	*weakly similar to* phytoalexin-deficient 4 protein {Arabidopsis thaliana}, *partial (8%)*
CN012385	4779314 - 4780335	1.7e-82	1	None	none
TC224813	4779323 - 4780386	4.1e-81	8	None	*weakly similar to* GB|AAF09479.1|6457331|AF188329 phytoalexin-deficient 4 protein {Arabidopsis thaliana;} , *partial (5%)*
TC206416	4790395 - 4791565	2.1e-70	14	None	*similar to* UP|O82044 (O82044) Amino acid carrier, *partial (46%)*
TC206408	4790395 - 4792383	6.2e-183	28	None	*similar to* UP|O82044 (O82044) Amino acid carrier, *partial (87%)*
TC206414	4791678 - 4792311	4.5e-44	5	None	*similar to* UP|O82044 (O82044) Amino acid carrier, *partial (16%)*
CK204373	4799000 - 4799926	3.5e-34	1	None	*similar to* hypothetical protein F24M12.260 - Arabidopsis thaliana, *partial (15%)*
TC192774	4857466 - 4858637	8.7e-60	7	C-4BS4-0.37 4AL12-0.43-0.59 4AL12-0.43-0.59 4DS3-0.67-0.82	*similar to* (Q8RZP9) B1065E10.15 protein, *partial (5%)*
TC192772	4857662 - 4862668	2.6e-111	13	C-4BS4-0.37 4AL12-0.43-0.59 4AL12-0.43-0.59 4DS3-0.67-0.82	*similar to* GB|AAF79266.1|8778257|AC023279 F12K21.21 {Arabidopsis thaliana;} , *partial (81%)*
TC192773	4858576 - 4860254	4.1e-108	8	None	*similar to* GB|AAF79266.1|8778257|AC023279 F12K21.21 {Arabidopsis thaliana;} , *partial (61%)*

CA643561	4859113 - 4860280	3.3e-56	1	None	*similar to* F12K21.21 {Arabidopsis thaliana}, *partial (36%)*
CA593533	4859600 - 4860280	1.7e-35	1	None	*similar to* F12K21.21 {Arabidopsis thaliana}, *partial (23%)*
CA605437	4862033 - 4862668	1.2e-32	1	None	*similar to* F12K21.21 {Arabidopsis thaliana}, *partial (10%)*
CK207473	4882016 - 4882760	1.7e-25	1	None	*similar to* PIR|G86239|G862 protein F20B24.6 [imported] - Arabidopsis thaliana, *partial (2%)*
CK209806	4908332 - 4909705	2.3e-79	1	None	*similar to* hypothetical protein AT4g09630 [imported] - Arabidopsis thaliana, *partial (5%)*
TC196242	4909206 - 4913219	1.5e-182	7	None	*similar to* UP|Q9LNL6 (Q9LNL6) F12K21.12, *partial (49%)*
CK210115	4960689 - 4961813	3.6e-37	1	None	protein disulfide isomerase {Triticum turgidum subsp. durum}, *partial (44%)*
TC225348	4960691 - 4962585	4.7e-82	2	None	UP|Q7FYS2 (Q7FYS2) Protein disulfide isomerase 1 precursor , *partial (46%)*
TC206472	4960691 - 4964698	3.1e-184	112	None	UP|Q7FYS2 (Q7FYS2) Protein disulfide isomerase 1 precursor , *complete*
TC220284	4960691 - 4964698	2.6e-192	151	4DS1-0.53-0.67 C-4BL1-0.71 4AL5-0.66-1.00* C-4BL1-0.71 C-4BL1-0.71 OR 4AL12-0.43-0.59	UP|Q93XQ8 (Q93XQ8) Protein disulfide isomerase 2 precursor , *complete*
CK209051	4960716 - 4961813	9.3e-29	1	None	None
TC219115	4960716 - 4962132	7.1e-65	4	None	*homologue to* UP|Q93XQ8 (Q93XQ8) Protein disulfide isomerase 2 precursor , *partial (43%)*
TC190587	4960807 - 4964698	4.7e-147	181	None	protein disulfide isomerase [Triticum aestivum]
TC225347	4960988 - 4962579	8.6e-66	6	None	*homologue to* UP|Q7FYS2 (Q7FYS2) Protein disulfide isomerase 1 precursor , *partial (48%)*
CK206490	4961143 - 4963514	5.7e-124	1	None	None
CK204484	4961378 - 4963217	7.1e-77	1	None	None
CA618500	4961642 - 4962935	1.3e-57	1	None	*homologue to* SP|P52588|PDI_ Protein disulfide isomerase precursor (PDI) (EC 5.3.4.1). [Maize] {Zea mays}, *partial (28%)*
CA717271	4961659 - 4962576	5.5e-40	1	None	None
CK168537	4961659 - 4962923	3.7e-37	1	None	*homologue to* protein disulfide isomerase {Triticum turgidum subsp. durum}, *partial (34%)*
CA707119	4961930 - 4962947	7.4e-23	1	None	None
CA711864	4962144 - 4963176	1.1e-40	1	None	None
BE414158	4962486 - 4963521	1.2e-55	1	None	None
CK161783	4962496 - 4963856	3.4e-96	1	None	None
CA696449	4962503 - 4963264	2.7e-44	1	None	None
BQ606805	4962519 - 4963521	1.7e-55	1	None	None
CA665402	4962817 - 4963678	9.7e-39	1	None	None
TC201070	4963068 - 4964698	2.3e-77	2	None	*homologue to* UP|Q93XQ8 (Q93XQ8) Protein disulfide isomerase 2 precursor , *partial (34%)*
CA726808	4963086 - 4963690	5.0e-37	1	None	*similar to* protein disulfide isomerase {Triticum turgidum subsp. durum}, *partial (18%)*

CA626889	4963134 - 4964552	3.8e-67	1	None	*homologue to* Protein disulfide isomerase precursor (PDI) (EC 5.3.4.1). [Maize], *partial (26%)*
CK195525	4964091 - 4964698	1.1e-32	1	None	None
TC192879	4980685 - 4981821	2.1e-47	9	None	*similar to* UP|Q6NMK2 (Q6NMK2) At5g49400, *partial (55%)*
TC192878	4980685 - 4981904	5.1e-63	15	None	*similar to* UP|Q6NMK2 (Q6NMK2) At5g49400, *partial (45%)*
TC207667	4983246 - 4995545	8.0e-233	41	4DS3-0.67-0.82 4BS8-0.57-0.81	*homologue to* UP|Q43486 (Q43486) B32E protein, *complete*
CA657932	4986713 - 4987583	2.3e-39	1	None	*similar to* Unknown protein {Arabidopsis thaliana}, *partial (9%)*
BF484909	5299709 - 5306011	1.5e-40	1	None	*similar to* unnamed protein product {Mus musculus}, *partial (10%)*
TC212686	5313720 - 5314576	1.0e-55	5	None	*weakly similar to* UP|Q6Z8L0 (Q6Z8L0) Mitochondrial transcription termination factor-like, *partial (29%)*
BQ237761	5319873 - 5321800	7.9e-79	1	None	*similar to* ESTs C74776(E51022) C26123(C116681) correspond to a region of the predicted gene.~Similar to, *partial (25%)*
BQ240822	5321450 - 5323000	3.0e-67	1	None	*similar to* ESTs C74776(E51022) C26123(C116681) correspond to a region of the predicted gene.~Similar to, *partial (22%)*
TC217165	5323512 - 5324483	1.8e-73	2	None	*similar to* UP|Q9LI02 (Q9LI02) ESTs C74776(E51022), *partial (16%)*
CA593343	5342566 - 5343494	5.9e-45	1	None	*homologue to* unknown protein {Oryza sativa}, *partial (10%)*
TC196715	5347261 - 5368277	9.7e-135	5	None	*similar to* UP|Q9W539 (Q9W539) CG16902-PA, *partial (3%)*
CA645055	5348033 - 5348742	1.4e-29	1	None	*similar to* unknown protein {Oryza sativa}, *partial (9%)*
CA601498	5348049 - 5368700	2.9e-65	1	None	None
TC212622	5348141 - 5349135	4.3e-41	5	None	UP|Q6IIJ1 (Q6IIJ1) HDC17998, *partial (11%)*
TC231682	5361355 - 5362563	5.5e-79	2	None	*homologue to* UP|Q8MKX3 (Q8MKX3) CG2368-PD (Cg2368-pe), *partial (5%)*
BE516134	5367079 - 5368052	1.1e-98	1	None	*similar to* unknown protein {Oryza sativa}, *partial (21%)*
CA708335	5367101 - 5367879	5.2e-61	1	None	*similar to* unknown protein {Oryza sativa}, *partial (11%)*
TC210736	5367788 - 5368956	9.0e-58	10	None	*similar to* GB|AAH45472.1|28279616|BC045472 zgc:55839 {Danio rerio;} , *partial (3%)*
TC207984	5401824 - 5406551	4.3e-207	28	4DS3-0.67-0.82 4AL12-0.43-0.59	*similar to* UP|Q6RFY4 (Q6RFY4) MAP3Ka, *partial (50%)*
TC207986	5403647 - 5405577	2.8e-59	3	None	*weakly similar to* UP|Q6RFY3 (Q6RFY3) MAP3Ka, *partial (38%)*
CK208019	5404431 - 5405431	2.2e-43	1	None	*similar to* similar to mitogen-activated protein kinases {Oryza sativa}, *partial (21%)*
CD906245	5404786 - 5405661	1.5e-36	1	None	*similar to* similar to mitogen-activated protein kinases {Oryza sativa}, *partial (20%)*
TC215960	5404961 - 5406410	1.3e-77	2	None	*similar to* UP|Q6RFY4 (Q6RFY4) MAP3Ka, *partial (19%)*
CA484226	5405802 - 5406512	7.9e-54	1	None	None
TC207985	5405983 - 5407361	8.2e-40	4	None	*similar to* unknown protein {Oryza sativa;}, *partial (8%)*
AL826330	5423954 - 5424863	1.5e-24	1	None	*homologue to* myb-like protein {Oryza sativa}, *partial (15%)*

Overlapping and/or TC sequences with the same annotation are shown grouped together by background colour.

[a] The TIGR TaGI v9.0 sequences were aligned to the TIGR rice genome assembly 2004. "TC" sequences represent contigs while the other names represent Genbank accession numbers of singleton ESTs.

[b] The numbers represent the position (in base pairs) of the orthologous rice sequences in the TIGR rice genome assembly 2004 chromosome 11.

[c] The BLASTn e-values were derived from the alignments of the TIGR TaGI sequences and rice genome sequences in the Gramene Genome Browser.

[d] The number of ESTs used in the assembly of the TC sequences from TIGR TaGI v9.0.

[e] The map positions are as reported in the GrainGenes-SQL database.

[f] The TIGR annotation describes the best BLASTx hit for each entry. The nomenclature is as described on the website (www.tigr.org) and the percentage of the TC sequence aligned with the tentative annotation is shown in parentheses.

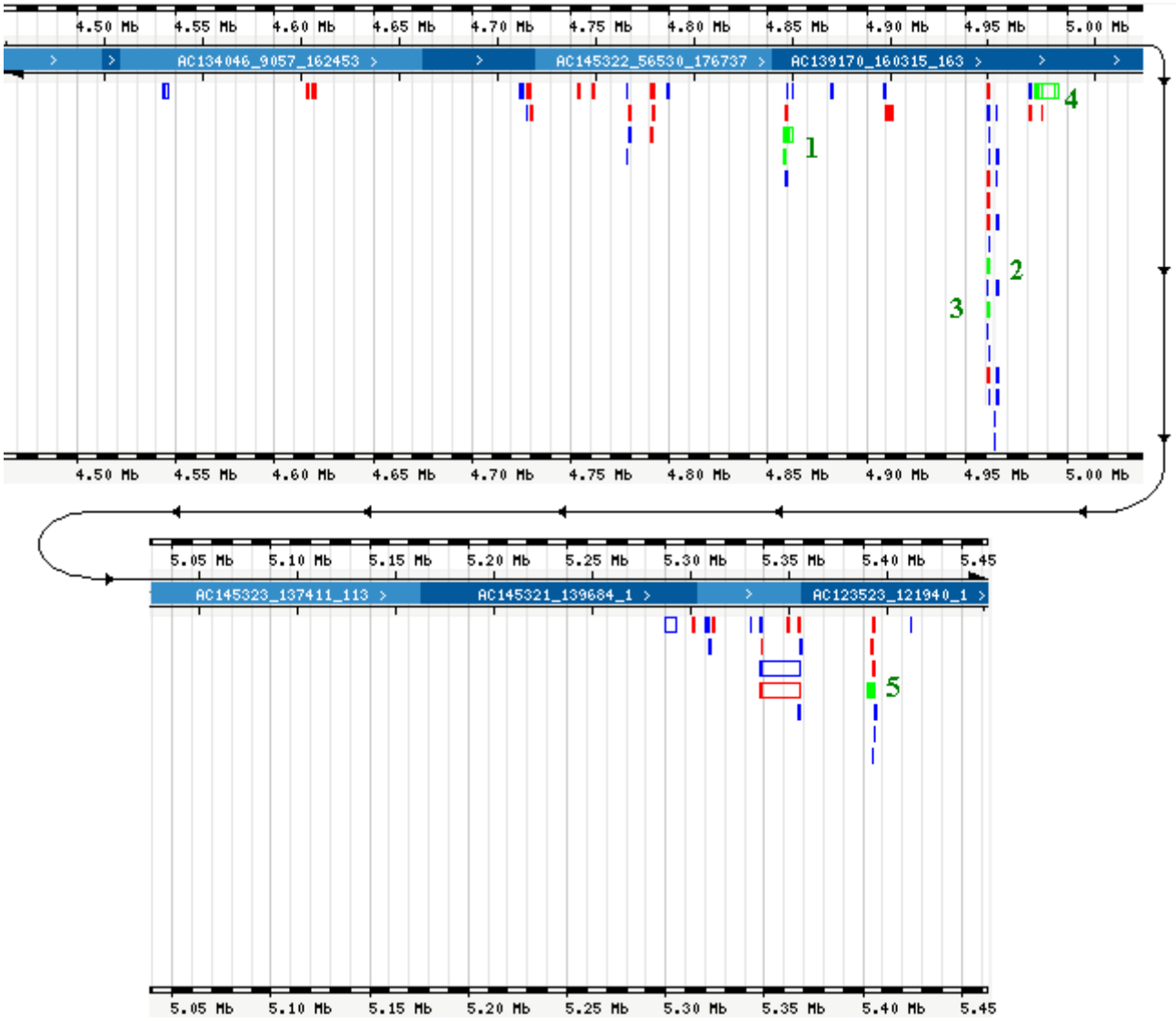


Figure 5.4 Putative wheat orthologs at the *PDI* locus of rice

The 1Mb genomic region of rice chromosome 11 encompassing the *PDI* gene of rice. The rice genome BAC contig is displayed in blue ("DNA(contigs)") with the individual BACs shown as light and dark blue sections. The putatively orthologous sequences from the TIGR TaGI database (Table 5.1) are shown aligned beneath the BACs ("Wheat_GI"). Blue blocks are orthologous EST singletons, red are orthologous TC sequences and green are orthologous TC sequences containing ESTs from probes mapped in the wheat genome, with the numbers referring to the probes used to map them (Table 5.2). This figure was generated from the Genome Browser interface of Gramene (www.gramene.org) and edited to highlight certain features. The scale (in Mb) is shown above and below the figure.

5.3.2 Identification of the physical loci of the putative wheat orthologs

To determine whether any of the above 33 TC sequences were assembled from ESTs sequenced from cDNA probes that had been used for physical mapping experiments available in the GrainGenes-SQL database, the 763 EST accession numbers were entered into the "Mapped Loci for EST-derived Probes" query in the GrainGenes-SQL database (http://wheat.pw.usda.gov/cgi-bin/westsql/map_locus.cgi; accessed 10/04). This analysis identified 9 such ESTs that had been sequenced from 5 cDNA probes physically mapped to wheat chromosomal 'bins' (Figure 5.4, Table 5.2). Further examination revealed that the nine ESTs represented the 5' and 3' sequences of the five different cDNA probes from the TIGR TaGI database, with an absence of the 5' EST from probe BQ280847 (explaining only 9 sequences from 5 cDNA probes, instead of the expected 10). However, some of the 5' and 3' ESTs were not always present in the same TCs; specifically:

- The 5' and 3' ESTs from the probe BF200779 were present in two similar but separate TCs (TC192774 and TC192772) (with a 400bp overlap with 90% identity) at the same locus (probe 1 in Figure 5.4). Thus, it appears that these two TCs probably represent the same gene, or homeologues of it.
- The ESTs from probes BE398523 and BQ280847 are found in a single TC (TC220284), suggesting that these two probes have been used to map the same or homeologous genes.

Assessment of the chromosomal loci of the cDNAs for these probes revealed all of them to be localised to the group 4 chromosomes of wheat (Table 5.2). However, most strikingly, the probes BE398523 and BQ280847, specific for TC220284, i.e., a *PDI* gene of wheat, and hence expected to map to the same locus, are localised to two different loci on chromosome 4AL (Table 5.2; Figure 5.5). A closer inspection of this mapping data was hence warranted, and revealed that the blots (available at http://wheat.pw.usda.gov/cgi-bin/westsql/contig.cgi) for probes BF200779, BQ280847, BE499456 and BM135436 were of sufficient quality to support their determined loci, however, that for BE398523 was particularly poor, possibly explaining the discrepancies between the data obtained from the two probes for the same gene (Appendix V).

Table 5.2 Reported chromosomal loci of the TIGR TaGI TCs

Probe#[a]	TIGR TC (v9.0)	Mapped ESTs	Mapped Loci[b]
1	TC192772	**PROBE:** BF200779 **5'EST:** BF200779	4AL12-0.43-0.59 C-4BS4-0.37 4DS3-0.67-0.82
	TC192774	**3'EST:** BQ166177	
2	TC220284	**PROBE:** BE398523 **5'EST:** BE398523 **3'EST:** BQ167770	4AL5-0.66-1.00* C-4BL1-0.71 4DS1-0.53-0.67
3		**PROBE:** BQ280847 **5'EST:** ABSENT **3'EST:** CD490794	4AL12-0.43-0.59
4	TC207667	**PROBE:** BE499456 **5'EST:** BE499456 **3'EST:** CD454116	4BS8-0.57-0.81 4DS3-0.67-0.82
5	TC207984	**PROBE:** BM135436 **5'EST:** BM135436 **3'EST:** BQ161582	4AL12-0.43-0.59 4DS3-0.67-0.82

[a] These numbers refer to the mapped TCs displayed in Figure 5.4.

[b] These loci are as reported in the GrainGenes-SQL database. These loci are illustrated in Figure 5.7.

5.4 THE CHROMOSOMAL 'BIN' POSTIONS OF THE *PDI* GENES IN WHEAT

As the physical mapping data for TC220284, one of the *PDI* genes of wheat, was questionable (as discussed above), mapping work was conducted to clearly establish the physical chromosomal 'bin' loci for all three *PDI* genes. The AS-primer pairs PA1F/PA1R, PB3F/PB3R and PD2F/PD2R established earlier (Section 3.4) were used to amplify sections of the three genes encoding sequences of the cDNA types *wPDI1*, *wPDI2* or *wPDI3*, respectively, from the 42 nullisomic/tetrasomic lines developed from *T. aestivum* cv. Chinese Spring. The amplifications produced expected products of ~230bp from all lines except for the two lines nullisomic for 4A with PA1F/PA1R, products of ~650bp from all lines except those nullisomic for 4B with PB3F/PB3R, and products of ~240bp from all lines except those nullisomic for 4D with PD2F/PD2R (Figure 5.5).

The *PDI* AS-PCRs were then extended to the ditelosomic and deletion lines of the group 4 chromosomes for determining more precise chromosomal locations. The

PA1F/PA1R primer pair amplified the expected product in all of the ditelosomic and deletion lines assayed, the PB3F/PB3R amplified the expected product in all of the lines except the deletion lines 4BS4 and 4BS8, and the PD2F/PB2R pair amplified the expected product in all of the lines except the ditelosomic line DT4DL and the deletion lines 4DS1 and 4DS3 (Figures 5.5 and 5.6). All three AS-PCRs amplified PCR products from the group 1 chromosome NT lines (Appendix IV-1), suggesting that the fourth *PDI* locus, reportedly from chromosome 1B (Section 1.5.5; Ciaffi *et al*., 1999), if present, was not co-amplified by these primers.

The chromosomal 'bins' for each of the three *PDI* genes suggested by these results are summarised in Table 5.3 and Figure 5.7. The *PDI* gene is deduced to exist on 4AL (C-0.59) (Table 5.3; Figure 5.7), due to our Southern blotting and cloning analysis revealing only 3 full-length *PDI* genes in wheat (Chapter 3) and Ciaffi *et al*. (1999) revealing the *PDI* gene on chromosome 4A to be absent in the 4AS ditelosomic line.

Table 5.3 Summary of the physical mapping of the *PDI* genes

Primer Pair	Lines producing no PCR product	Deduced Chromosomal 'bin'[1]
PA1F/PA1R	N4AT4B, N4AT4D	4AL (C-0.59)[2]
PB3F/PB3R	N4BT4A, N4BT4D, 4BS4, 4BS8	4BS (0.51-0.81)
PD2F/PD2R	N4DT4A, N4DT4B, DT4DL, 4DS1, 4DS3	4DS (0.67-1.00)

[1] These 'bins' are illustrated in Figure 5.7.

[2] Partially based on the localisation of a *PDI* gene to the long arm of chromosome 4A in an earlier study (Ciaffi *et al*., 1999).

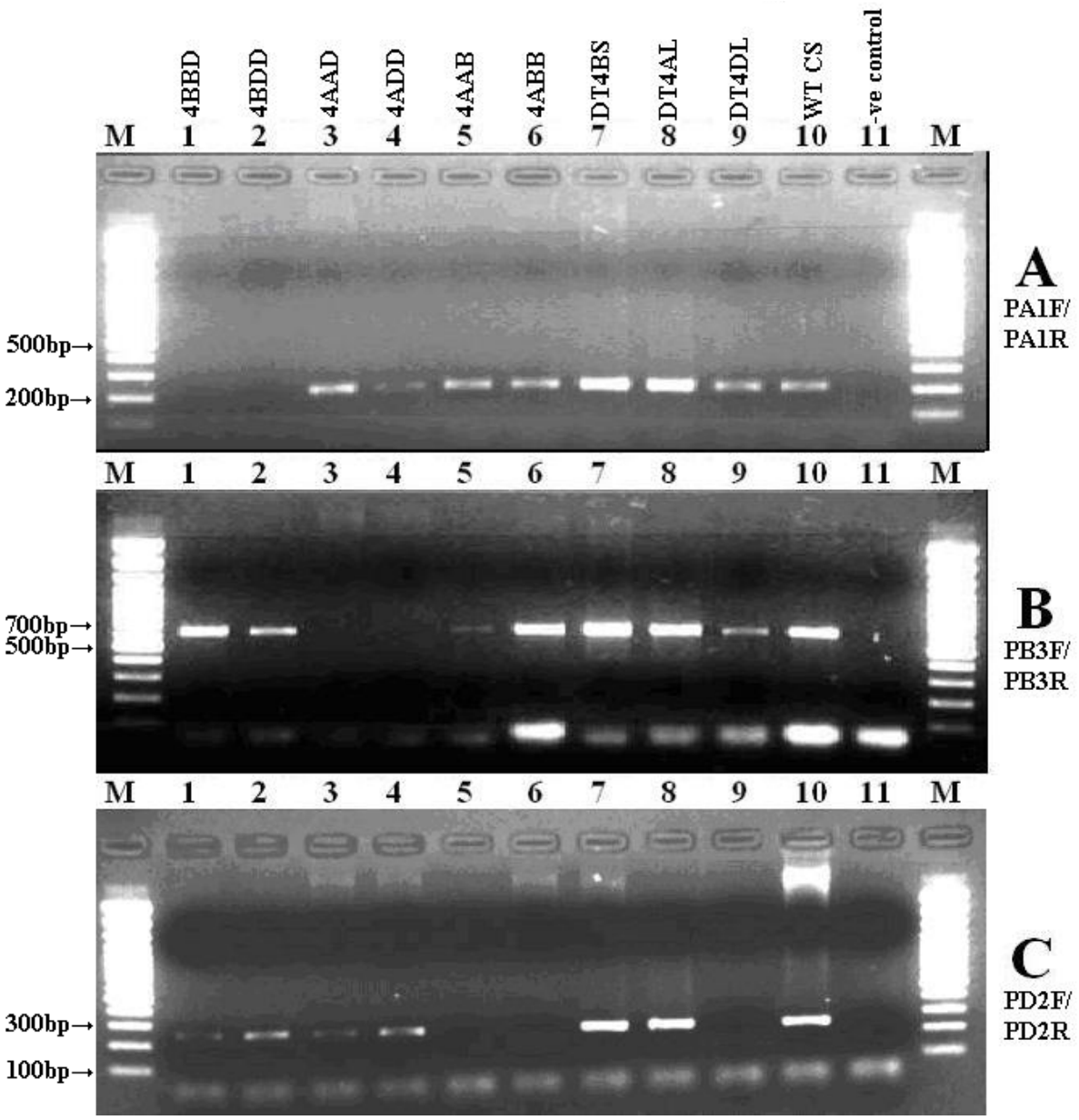


Figure 5.5 Nullisomic/tetrasomic and ditelosomic localisation of *PDI* genes

The AS-PCR products generated from the amplification of *PDI* genes from each of the genomes wheat. Primers were designed based on cDNA sequences *wPDI1* (A), *wPDI2* (B) and *wPDI3* (C) (Johnson *et al.*, 2001). Lanes: M: molecular weight marker; 1: 4BBD; 2: 4BDD; 3: 4AAD; 4: 4ADD; 5: 4AAB; 6: 4ABB; 7: DT4BS; 8: DT4AL; 9: DT4DL; 10: Chinese Spring gDNA; 11: dH_2O negative control.

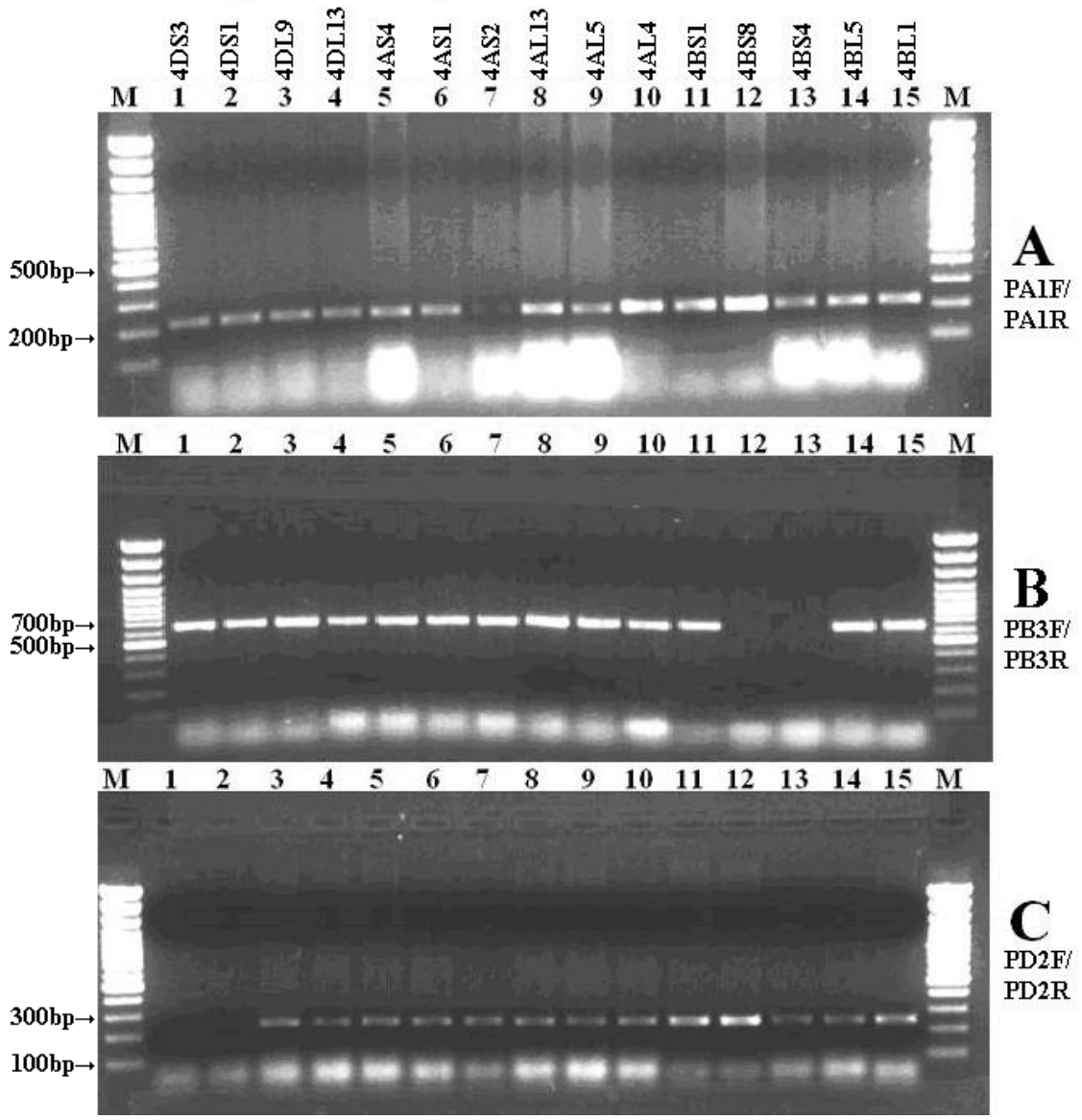


Figure 5.6 Deletion mapping of *PDI* genes

The AS-PCR products generated by amplification of each of the *PDI* genes from chromosome 4 deletion lines of *T. aestivum* cv. Chinese Spring. Positions of the deletions are illustrated in Figure 5.7. Lanes: M: molecular weight marker; 1: 4DS3; 2: 4DS1; 3: 4DL9; 4: 4DL13; 5: 4AS4; 6: 4AS1; 7: 4AS2; 8: 4AL13; 9: 4AL5; 10: 4AL4; 11: 4BS1; 12: 4BS8; 13: 4BS4; 14: 4BL5; 15: 4BL1.

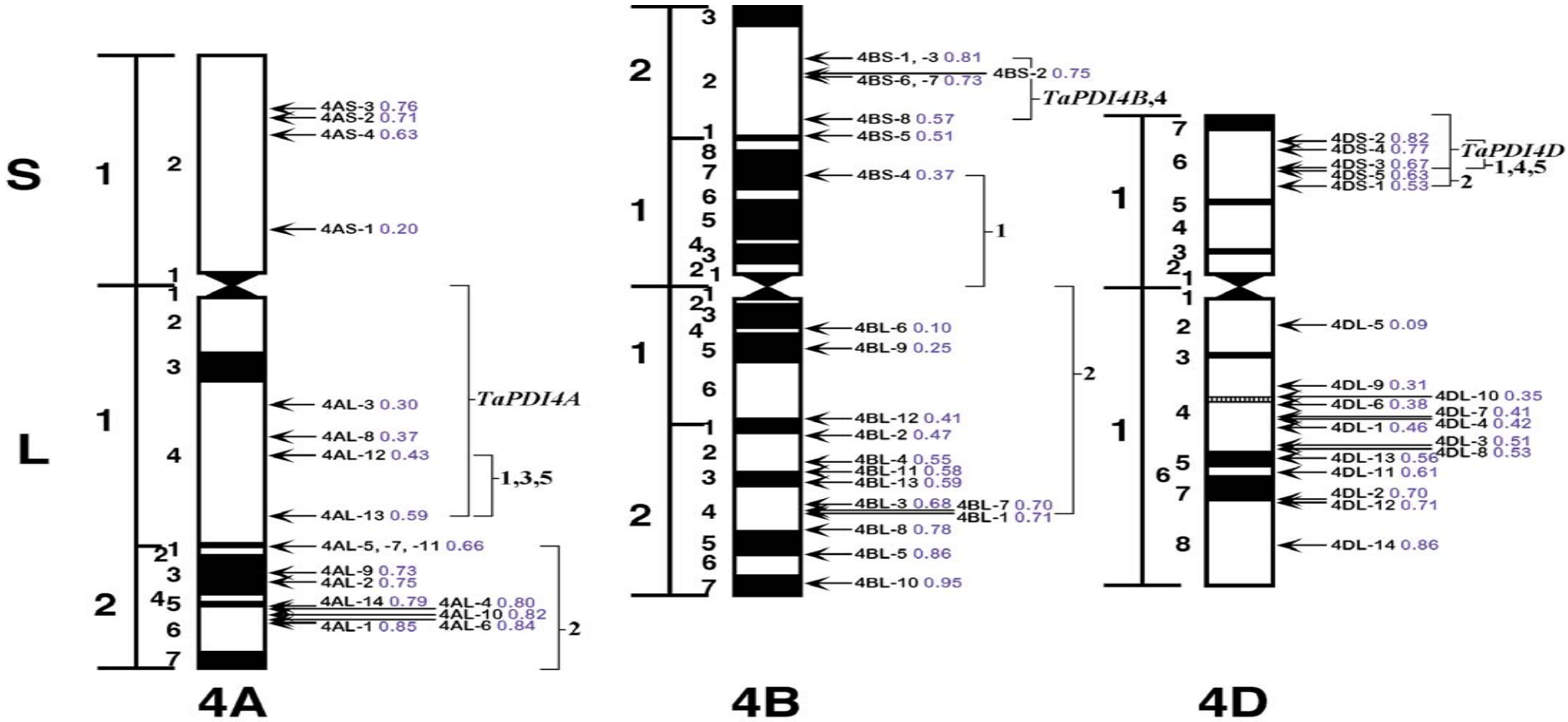


Figure 5.7 Mapped locations of *PDI* genes and TC sequences

The locations of the three *PDI* cDNA clones on the group 4 chromosomes of *T. aestivum* as determined by AS-PCR in the current study are shown as *wPDI1*, *wPDI2* and *wPDI3*. The reported positions of the probes BF200779 (1), BE398523 (2), BQ280847 (3), BE499456 (4) and BM135436 (5) (Table 5.2) are shown. The deletion stocks are shown by name with arrows indicating the breakpoints with the fraction of the remaining chromosome arm next to them. The deletion stock figure was accessed from the Wheat Genetics Resource Page (http://www.ksu.edu/wgrc/Germplasm/Deletions/grp4.html; accessed 10/04).

5.5 INVESTIGATION OF RICE QTLS AT THE *PDI* LOCUS

Comparative mapping of QTLs on the short arm of rice chromosome 11 with the position of the rice PDI gene in the TIGR Gramene Rice Genome Assembly sequence map was undertaken as a preliminary investigation into whether any of the known QTLs map to this area of the rice genome. The individual QTL maps available via the cMAP interface of the Gramene website (accessed 12/04) were compared with markers on the Genome Assembly map and revealed several with correspondence between the two maps, permitting the comparison of QTL positions and the position of the *PDI* gene (bp 4,908,940 to 4,912,305; see above). Altogether twelve QTL maps contained markers that had correspondence to the sequence map and at least one QTL covering the region with the rice *PDI* (Table 5.4). A number of these QTLs covered large sections of the short arm of rice chromosome 11, and, in some cases, the *PDI* locus was located in close proximity of their flanking markers, eg. male fertility restoration, sheath blight disease resistance, days to heading and leaf width. In addition, a number of pathogen resistance QTLs were located at the *PDI* locus; of particular note was the small (1cM) blast resistance QTL (AQCT007) centred on a ~1Mb region containing the *PDI* gene (Table 5.4). Potentially related to these pathogen resistance QTLs was the 8.8cM QTL for H_2O_2 content (AQCW017), also centred on the *PDI* locus.

Table 5.4 Candidate QTLs at the *PDI* locus of rice

QTL Map[a]	QTL[b]	QTL Position (cM)[c]	*PDI* flanking markers[d]	QTL map position (cM)[e]	Genome assembly position (bp)[f]
Rice CNRRI Zh97B/Mil46 RI QTL 2002-11	**Male fertility restoration** (AQCG004)	19.3-37.8	RG118 RM202	19.3 35.1	4,421,099 8,788,252
Rice-CNHZAU Zh97/Ming63 RI QTL 2002-11	**Spikelet number** (AQAB053)	100.1-127.8	RG118 G44	127.8 97.2	4,421,099 9,780,636
	Plant height (AQY004)	97.2-140.7			
	Grain yield (AQAG007)	100.1-127.8			
Rice-JNIAR Aki/Kosh DH QTL 2003a-11	**Cold tolerance** (CQP8)	0-27.4	S1609 G320	0 38.3	1,492,164 6,618,368
Rice-IRRI Mil23/Aki RI QTL 2003-11	**Filled grain percentage** (AQAW019)	0-20.9	XNpb189A (G189) C477	4.8 22.9	2,023,755 5,407,117
Rice-JRGP Nip/Kas F2 QTL 2000-11	**Leaf senescence** (CQN42)	0-64.2	V79 (RG118) C410	25.3 28.2	4,421,099 5,366,267
	Spikelet weight (CQN61)	0-110.1			
	H_2O_2 content (AQCW017)	19.8-28.6			
Rice-JNIG W1944/Peik QTL 2002-11	**Seed dormancy** (CQAH43; QF034; CQAH42; AQF023; AQF017; CQAH45; AQF041)	29.1-44.5	RZ141 G320	29.1 69.5	4,064,334 6,618,368
	Grain Shattering(AQF005)	29.1-44.5			
	Leaf length/width ratio (AQF153)	29.1-44.5			
	Ratooning ability (AQF083)	29.1-44.5			
	Anther length (AQF073)	29.1-44.5			
Rice-Brazil BG90-2/RS16 QTL 2002-11	**100-seed weight** (AQE046)	0-40.2	RM167 RM202	20.4 34.8	4,058,377 8,788,252
	Tiller number (AQE090)	0-40.2			
Rice-Zejiang IR64/Azu DH QTL 2001-11	**Plant height** (CQZ21)	33.8-37.5	RG118 G44	33.8 63.8	4,421,099 9,780,241
Rice-IGCAS Jas85/Lem F2 QTL 2000-11	**Sheath blight disease resistance** (CQAN8; CQAN4)	96.6-119.7	RG118 G44	119.7 96.6	4,421,099 9,780,241
Rice-Cornell IR64/IRG105 QTL 2003-11	**Days to heading** (AQED005)	24-55.1	RM167 RM202	24 39.7	4,058,377 8,788,252
Rice-CIRAD IR64/Azu DH QTL 2003-11	**Blast disease resistance** (AQCT007)	59.4-60.4	CSU50 C496	59.4 61.4	4,618,790 5,647,522
Rice-IRRI Lem/Teq RI QTL 2001-11	**Brown plant hopper resistance** (AQAP011; AQAP012; QAP034; AQAP013)	49.1-52.2	RM167 RZ53	38.9 52.2	4,058,377 5,642,686
	Spikelet fertility (AQCU179)	49.1-52.2			
	Days to heading (AQCU055)	49.1-52.2			
	Leaf width (AQCU229)	28-49.1			
	Plant height (AQCU205)	28-49.1			

[a] The QTL maps were accessed from the Gramene website (www.gramene.org).

[b] The accession numbers of the QTLs are given in brackets.

[c] The position of the QTL on the corresponding QTL map.

[d] The molecular markers most closely flanking the *PDI* gene on the genome assembly map with correspondence to markers on the respective QTL maps. The marker names in brackets are marker aliases as listed on the sequence map.

[e] The positions of the flanking markers on the QTL map.

[f] The positions of the flanking markers on the genome assembly sequence map.

5.6 DISCUSSION

The role and importance of PDI in the efficient deposition of the disulfide-bonded protein bodies of wheat has yet to be firmly established. As there is evidence suggesting an important role for this enzyme in the correct deposition of the disulfide-bonded storage proteins of rice in the form of abnormal storage protein deposition in a rice mutant, esp2, which lacks *PDI* expression (Section 1.5.6), the identification of a locus (or loci) in wheat that is orthologous to the *esp2* locus of the rice would suggest a potentially important wheat locus (or loci) involved in storage protein deposition. Thus, we compared genes present at the *esp2*/*PDI* locus of rice with the position of their orthologs in wheat to determine whether orthology exists between the *PDI* loci in wheat and rice.

5.6.1 The *esp2* locus and *PDI* gene of rice are both located on chromosome 11S

As the esp2 mutant lacked *PDI* expression (Section 1.5.6) it was conceivable that the mutation was directly to *PDI* or its *cis*-acting regulatory sequences and thus, the *PDI* and *esp2* loci would map to similar positions. To investigate this possibility, initially, a rice BAC (AC139170) was identified that contained a *PDI* gene with similar intron/exon structure and exon sequences (~76% identity) to a *PDI* gene in wheat (Figure 5.1). This rice BAC was located on the Rice Gramene TIGR chromosome 11 assembly of the rice genome, supported by the work of Wu *et al.* (2002) which estimated the genetic map position of this BAC to be 28.6 cM from the proximal marker on the short arm of chromosome 11. This position compared with the presence of the *esp2* marker on the Hokkaido Morphological 2000 map, also on chromosome 11, 17 cM from the proximal marker on this map. However, different proximal markers were used in these analyses, ruling out a direct comparison between the genetic loci of *esp2* and the *PDI* gene of rice. The identification of markers on the JRGP RFLP genetic map, with correspondence to the Morph 2000 map containing the *esp2* marker and the Genome assembly sequence map containing the *PDI* gene revealed that markers located at similar positions on the JRGP RFLP map were located distal to the *esp2* marker and the rice BAC containing the *PDI* gene. The absence of genetic markers with correspondences to any available genetic maps distal to the *esp2* marker prevented the

identification of similar flanking markers on both maps, however, taken together, this data does reveal that the *esp2* marker and *PDI* gene in rice are both located on the terminal sections of the short arm of chromosome 11, distal to the S20163S marker on the Genome Assembly map, or on the terminal 5.2 Mb section of the short arm of chromosome 11. Thus, these similar map positions, taken together with the absence of *PDI* expression in the esp2 mutant, strongly suggest that the mutation in *esp2* is directly to the *PDI* gene or regulatory sequences flanking it. As such, a comparison of genes at the *PDI* locus of rice with those at the *PDI* loci of wheat was undertaken, to identify whether orthology has been conserved at these two loci.

5.6.2 The *PDI* gene of rice is flanked by a number of putative wheat orthologs

An investigation to ascertain whether the rice genome region containing the *PDI* gene contained potential orthologues of wheat, involving comparisons of sequences in the TIGR TaGI database that were highly similar to a 1Mb region of the rice genome flanking the *PDI* gene, revealed a high number of putatively orthologous TCs, many clustered at single loci along the rice genome (Figure 5.4), suggesting that they represented a single gene or its homeologues in the wheat genome. This assertion is supported by the tentative annotations attributed to such multiple TCs aligning with a single rice locus generally have the same, or similar putative protein products. For example, there are 24 TCs or singletons in this comparison that are highly similar to the *PDI* gene of rice and all have the same tentative annotation of *PDI* by BLASTx analysis (Table 5.1). Such a large number of similar TCs almost certainly reflect the methods used to construct the TaGI database, whereby overlapping ESTs require >94% identity over 40bp to be placed in a single TC sequence. This can lead to the assembly of: (i) single, 'hybrid' TCs derived from ESTs sequenced from >1 highly similar genes or homeoalleles; and (ii) many, partial TC sequences or singletons from a single type of cDNA failing to be assembled as a single sequence due to poor sequencing quality of the ESTs. Hence, the large numbers of putatively orthologous TCs probably represent far fewer actual expressed genes. This is also supported by the observation that most of the similar TCs at each locus are assembled by few (<10) ESTs (Table 5.1). Determination of the 'true' sequences of all putative orthologues would be required to clarify this issue. However, even with these limitations, the TIGR TaGI database

proved extremely useful in identifying a number of putative wheat orthologues at the *PDI* locus of rice.

5.6.3 Orthology between the *PDI* loci of wheat and rice

In order to identify the chromosomal loci of the putatively orthologous TC sequences in the wheat genome, the individual ESTs used in their assembly were analysed to determine whether they were sequenced from cDNAs used as physical mapping probes as part of the US Wheat EST project. This revealed that 9 of the ESTs had been sequenced from 5 cDNA probes that had been localised to chromosomal 'bins' (Table 5.2). Of these 5 cDNAs, two of them appeared to encode PDI (probes 2 and 3) while probe 1 appeared to encode a homologue of the protein of unknown function from *A. thaliana*, F12K21.21, probe 4 appeared to encode a homologue of B32E, a barley seed protein that is repressed by abscisic acid (Liu *et al*. 1992) and probe 5 appeared to encode a homologue of a rice protein with a sequence similar to mitogen-activated protein kinases (Tarchini *et al*. 2000). Interestingly, according to the wEST-SQL database all 5 probes mapped to the group 4 chromosomes of wheat. On closer inspection, the two PDI probes (probes 2 and 3) were reported to map to different 'bins' on the group 4 chromosomes, with probe 3 mapping to a single locus on chromosome 4AL and probe 2 mapping to three 'bins', one each on 4AL (at a different bin to that identified for probe 3), 4BL and 4DS. Further, the 4BL loci for probe 2 is in disagreement with the findings of Ciaffi *et al.* (1999), who localised the *PDI* genes of wheat to chromosome 4AL, 4BS and 4DS. To clarify the chromosomal 'bin' loci for each of the *PDI* genes, we used the nullisomic/tetrasomic, ditelosomic and deletion stocks of Chinese Spring to localise the *PDI* genes of wheat to chromosomes 4AL (in agreement with the reported loci for probe 3), 4BS (in agreement with the findings of Ciaffi *et al.* (1999)) and 4DS. Inspection of the mapping data generated for all of the probes provided further support for the incorrect localisation of probe 2, as this Southern blot, in particular, was of considerably low quality (Appendix V).

Omitting the data generated from probe 2**,** the reported loci in the GrainGenes SQL database for the remaining 4 probes revealed good agreement with our own physical mapping results, with the exception of probe 1 localising more centromeric on 4BS, while probe 4 and our mapping of *wPDI2* revealed a 'bin' distal to this. Whether this

apparent breakdown in co-linearity is due to a translocation of this gene during the evolution of hexaploid wheat, or due to experimental error in the mapping data, would require further investigation. It must also be noted that the Southern blot data on this website (http://wheat.pw.usda.gov/cgi-bin/westsql/contig.cgi) shows that only probe 1 hybridised to 3 fragments, as is generally observed Southern blots of the hexaploid common wheat, while probes 4 and 5 hybridised to 2 bands each and probe 3, a *PDI*-specific probe, to only one. It is unclear whether there are only two genes encoding the cDNAs used as probes 4 and 5 in the wheat genome, with an absence of the gene in the A or B genomes, respectively, or whether the third homeoallele lacked sufficient homology to hybridise to these particular probes; the latter appears to be the case for the *PDI* probe 3, which only hybridised to a single gene, as this probe is a partial cDNA, complementary to part of the 3'untranslated region of the *PDI* genes, areas known to be more divergent than the coding sequences (Johnson *et al.* 2001; Chapter 3). Further investigation of the genes encoding the other 3 probes would be required to determine their true copy number in *T. aestivum* cv. Chinese Spring as well as the most likely tetraploid (*T. turgidum*) and diploid (*T. urartu* or *T. monococcum, and Ae. speltoides*) progenitors of wheat.

The data generated here establishes firmly that several genes present at the *PDI* locus in rice, which appear to be linked to storage protein deposition are also present on the group 4 chromosomes of wheat, thereby establishing synteny between these two loci. Further support for orthology between the *PDI* loci of rice and wheat was also identified during the analysis of QTLs that map to rice chromosome 11. Identification of a QTL (AQCW017) for H_2O_2 production, covering a ~1Mb region of the rice chromosome 11 containing the *PDI* gene, was an interesting finding, in light of the close genetic linkage between both of the mapped wheat *PDI* genes and *GERMIN* loci which encode oxalate oxidases that are involved in H_2O_2 production in plants (Chapter 4). Although no putative rice *Germin* genes were identified during our preliminary analysis of this 1Mb region, further analysis is required. The assertion of such orthology (as above) is also supported by La Rota and Sorrells (2004) who did large scale comparisons between the rice genomic data and the mapped probes of wheat, revealing that large sections of the group 4 chromosomes of wheat are syntenic with rice chromosome 11, and the findings of Singh *et al*. (2004) who found that the long arm of rice chromosome 11 contains the most orthology to the group 4 chromosomes of wheat. A number of candidate QTLs for

a wide range of traits were also identified encompassing the *PDI* locus in rice (Table 5.4), that could provide future research targets for dissecting the role of PDI in grain development. Taken together, these results provide the foundation for the presence of a hypothetical QTL involved in storage protein deposition at the *PDI* loci in wheat. Further genetic analyses involving the analysis of gross protein body morphology, wheat mutants exhibiting aberrant storage protein deposition or more subtle differences in the protein bodies that may impact on dough rheology would be required to establish if such a QTL is present at these loci in wheat.

CHAPTER 6

CHARACTERISATION AND PHYSICAL MAPPING OF *CYCLOPHILIN A* GENES AND IDENTIFICATION OF NEW CLASSES OF CYCLOPHILINS IN WHEAT

List of Figures and Tables in Chapter 6

Chapter 6 Molecular Characterisation of the *Cyclophilin* Gene Family in Wheat

ABSTRACT

The cyclophilins, a class of 'foldase' enzymes that catalyse the rate-limiting step of peptidyl-prolyl *cis-trans* isomerization, are up-regulated during wheat endosperm development and suggested to be important for folding of the storage proteins, thus influencing wheat quality. However, little information exists on the types of cyclophilins expressed, the genes encoding these and their possible functions in wheat endosperm. We have characterised three isoforms of genes encoding cyclophilin A from eight wheat cultivars, using previously isolated cDNAs. The genes are small, intronless, comprise a small multi-gene family, lack any inter-cultivar polymorphisms and localise to chromosomal arms 6AS, 6BS and 6DS, in a region where genes for other quality traits are localised, the locus at 6AS possibly having duplicated genes. Further, cDNAs encoding two novel, endosperm-expressed classes of cyclophilins have been isolated, one being potentially plastid-localised and the other, a nuclear protein with cyclophilin-like domains. *In-silico* analyses have further led to identification of another form of cyclophilin A and a potentially ER-localised, endosperm-expressed cyclophilin B. The plastid and ER-localised forms are of particular relevance to events occurring during endosperm maturation. The results thus provide valuable data and molecular tools for isolation and analysis of these genes, to address their roles and any association with wheat quality traits.

6.1 INTRODUCTION

The significance of the amount and properties of the storage proteins of wheat to the end-use quality of wheat cultivars has been discussed in Section 1.4.2. Thus, an understanding of genetic factors that influence the synthesis and/or deposition of these proteins into protein bodies, which ultimately become the elastic gluten network, may provide potential targets for breeding programs to increase the quality of wheat cultivars. To address this broad aim, a number of studies have focused on identifying genes and proteins that are upregulated during the development of the wheat endosperm that may influence such agronomically important traits (Section 1.4.3). These genomic and proteomic analyses have identified a number of genes that are upregulated in the developing endosperm, including the expected storage proteins and interestingly, a number of genes encoding molecular chaperones, such as BiP, and foldase enzymes such as PDI and members of the PPIase family, the cyclophilins (Section 1.6). Although these chaperones/foldases may play general, housekeeping roles in this tissue, their marked up-regulation in the developing endosperm suggests that they may (additionally) play important roles associated with the specialised events occurring in this tissue, particularly the deposition of protein bodies. Such a hypothesis is supported by the observation that the gliadins do not appear to aggregate randomly after their synthesis, but are maintained in soluble forms before transport to the vacuole and integration into highly organised protein bodies (Section 1.4.3).

The characterization and genetic mapping of the protein disulfide isomerase genes (Chapters 3-5) has provided the basis for further functional analyses of this potentially important gene family. A similar approach would prove useful in elucidating the role of the cyclophilin family of proteins in wheat and, more specifically, in the development of the endosperm and quality traits. Analysis of cyclophilin expression profiles in the developing wheat endosperm has revealed high levels of *cyclophilin* mRNA as early as 5 DPA, suggesting an important role for these enzymes in the development of this tissue, although the particular class of *cyclophilin* involved was not identified in this study (Section 1.6.3; Grimwade *et al.*, 1996). Previous work undertaken by us to address this issue led to the identification of three cDNA sequence classes from a wheat endosperm library, which were highly similar to the *cyclophilin A* sequences in other plant species and likely to be cytoplasmic (or at least lacked the typical localisation

signals for ER or other organelles) (Section 1.6.3; Johnson *et al.*, 2001); to our knowledge, this is the only class of *cyclophilins* reported and completely sequenced so far from wheat. To gain further understanding of this and any other classes of PPIase enzymes and their roles in the developing wheat endosperm, this chapter describes the characterisation of the *cyclophilin A* gene family in *T. aestivum*. This is an important step in the design of molecular markers for these genes, which can be utilised to assess the association of these genes with any quantitative trait loci (QTL) important to the wheat industry. Further, the information generated from the *cyclophilin A* members was utilised to identify new members of this class as well as other novel classes of cyclophilins expressed in wheat.

6.2 SEQUENCING THE *CYCLOPHILIN A* GENES OF WHEAT

6.2.1 Confirming the specificity of *cyclophilin A* AS-primers

Alignment of the previously reported *cyclophilin* cDNA sequences revealed single nucleotide polymorphisms unique to each sequence, which were used to design a single pair of allele-specific primers for each gene, namely, C1F1/C1R1, C2F1/C2R1 and C3F1/C3R1 for *wCyp1*, *wCyp2* and *wCyp3*, respectively. To ensure the allele-specificity of these primers each pair was used to amplify the corresponding sections from the three cDNA clones, *wCyp1*, *wCyp2* and *wCyp3*. The results revealed the expected sized products, only from the corresponding cDNA when the optimised annealing temperatures were used (Table 2.5). Specifically, a ~520bp product was amplified from *wCyp1* with the C1F1/C1R1 primer pair, a ~780bp product from *wCyp2* with the C2F1/C2R1 primer pair and a ~500bp product from *wCyp3* with the C3F1/C3R1 primer pair (Figure 6.1).

6.2.2 Characterisation of partial *cyclophilin A* genes from *T. aestivum*

These allele-specific primers were then used to amplify the three corresponding *cyclophilin A* genes from the gDNA of eight cultivars used to establish genetic mapping populations (Section 1.8.4.1), *T. aestivum* cvs. Cranbrook, Halberd, Egret, Sunstar, Sunco, Tasman, CD87 and Katepwa. The results indicated products of the same size as that predicted from the respective cDNAs (Figure 6.2), suggesting an absence of introns, from at least the amplified sections of all three genes. Direct sequencing of the

partial of the partial *cyclophilin A* genes, henceforth termed *TaCYP18-1* (*wCyp1*) (Genbank Accession# AY456123), *TaCYP18-2* (*wCyp2*) (Genbank Accession# AY456122) and *TaCYP18-3* (*wCyp3*) (Genbank Accession# AY456124) revealed each to be identical to their corresponding cDNA (Figure 6.3), further confirming the allele-specificity of the primers and an absence of introns in at least these sections of the respective genes.

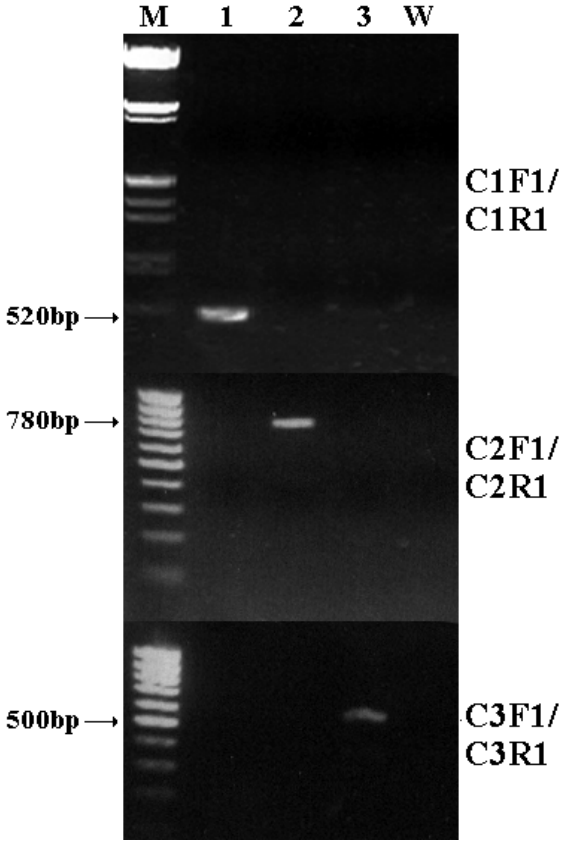


Figure 6.1 Confirmation of the allele-specificity of the *cyclophilin A* primers

The allele-specific amplification of sections of the *T. aestivum cyclophilin* cDNA clones *wCyp1* (1), *wCyp2* (2) and *wCyp3* (3). M= molecular weight marker (C1F1/C1R1 = λ *Eco*RI/*Hind*III, C2F1/C2R1 and C3F1/C3R1 = 100bp ladder); W= no template negative control.

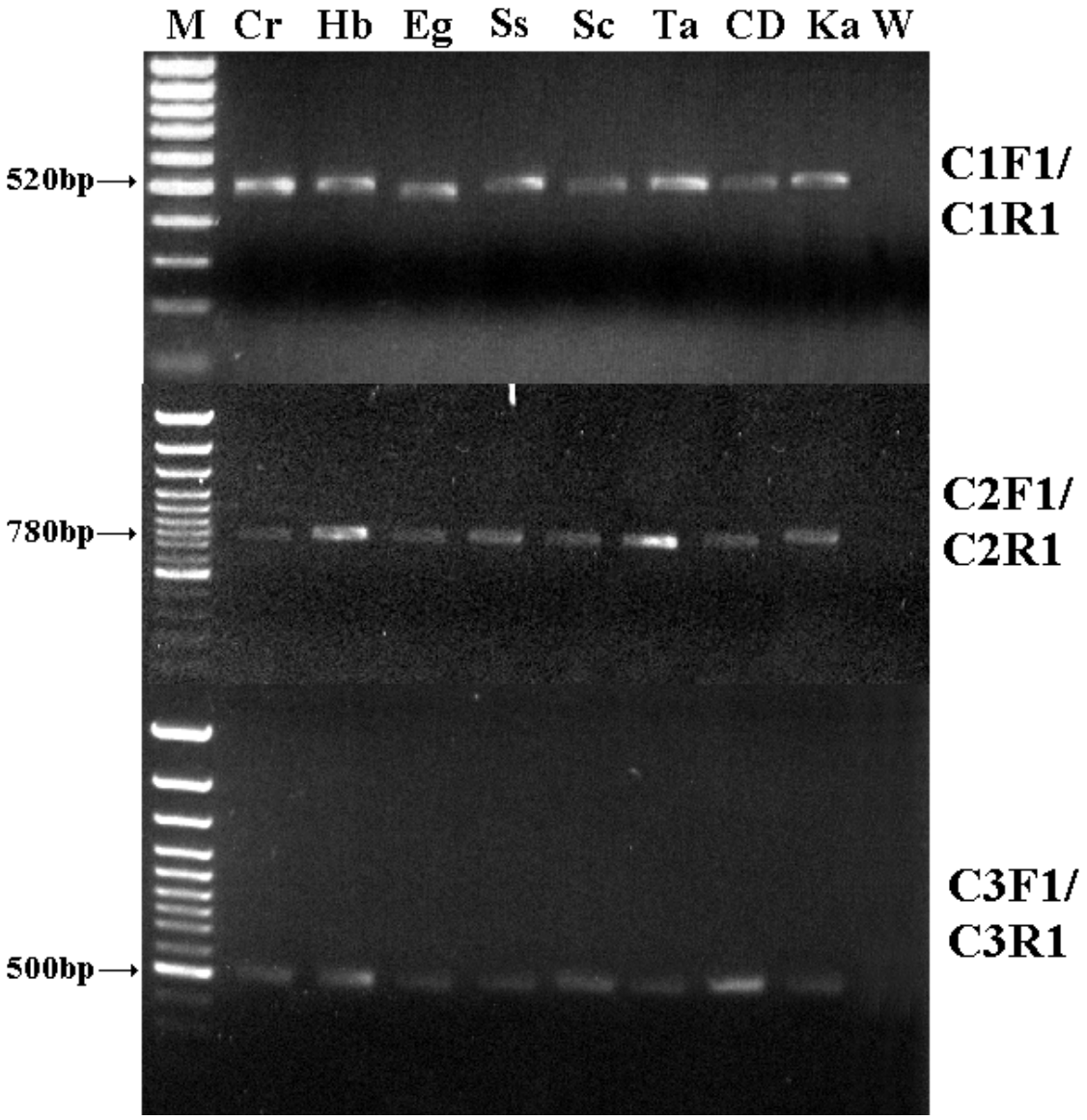


Figure 6.2 Amplification of the *cyclophilin A* genes from *T. aestivum*

The amplified sections of the three *cyclophilin A* genes from the parental cultivars used in the development of genetic mapping populations. M: 100bp molecular weight marker, Cr: Cranbrook, Hb: Halberd, Eg: Egret, Ss: Sunstar, Sc: Sunco, Ta: Tasman, CD: CD87, Ka: Katepwa, W: no template negative control.

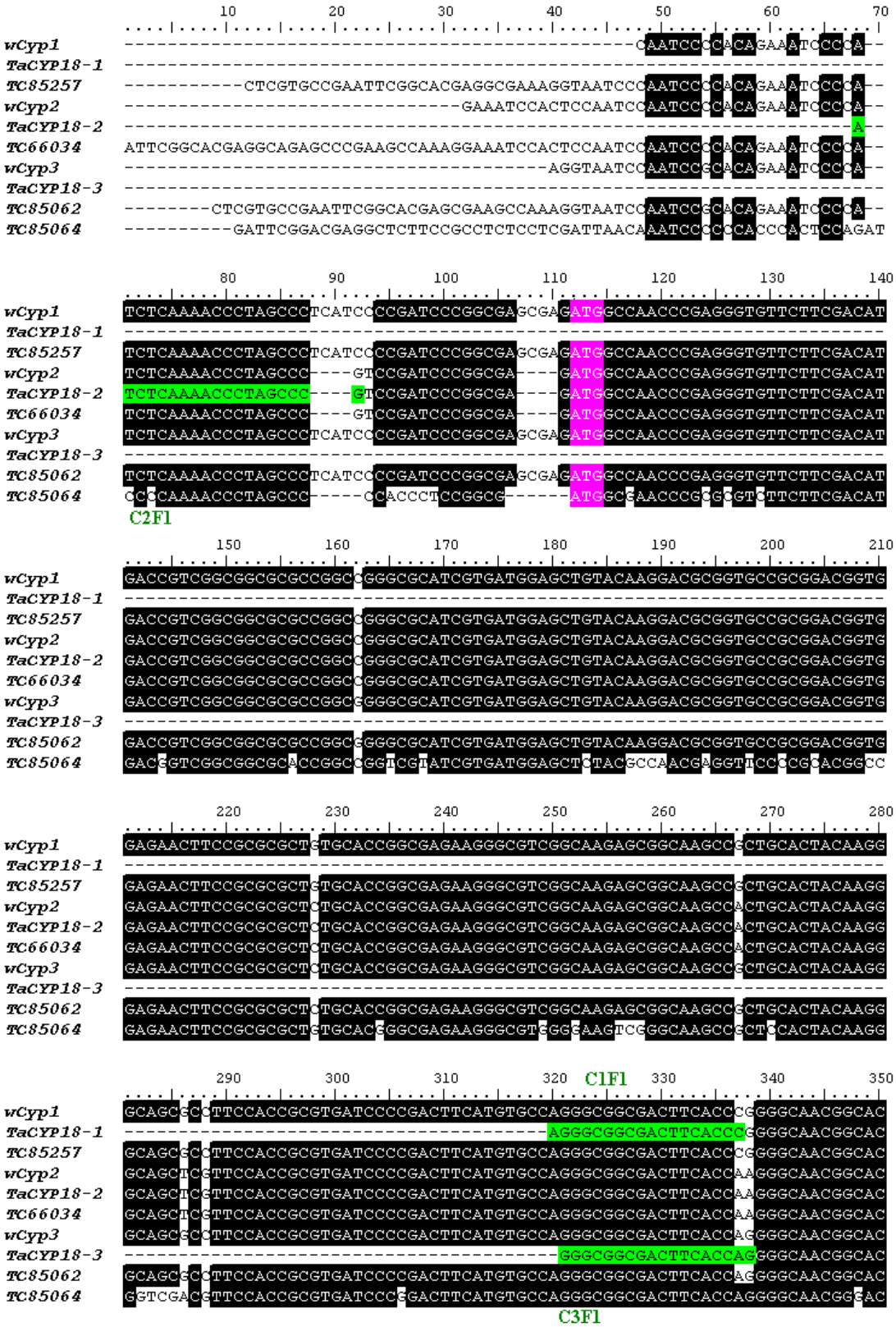


Figure 6.3 To be continued.

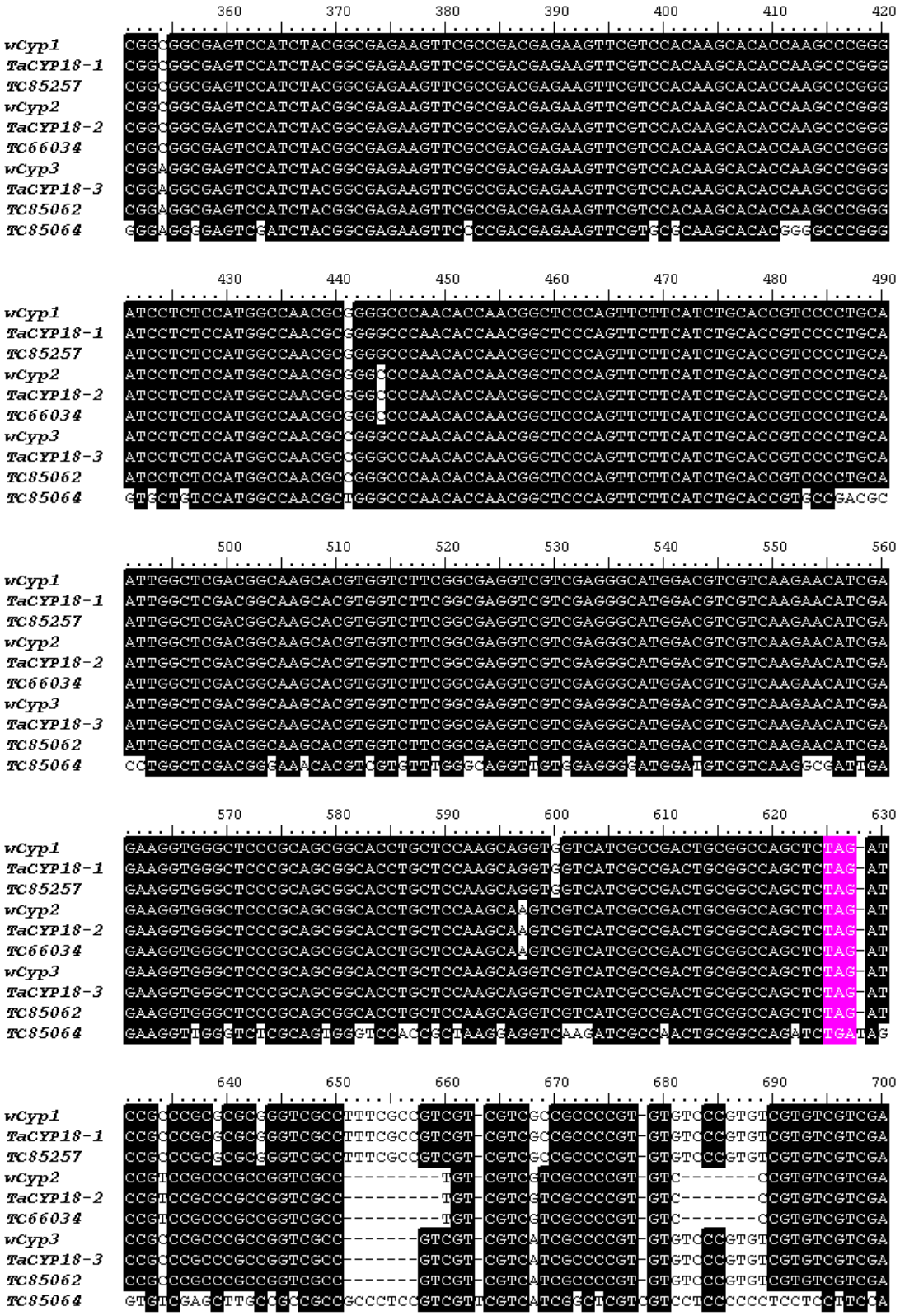


Figure 6.3 To be continued.

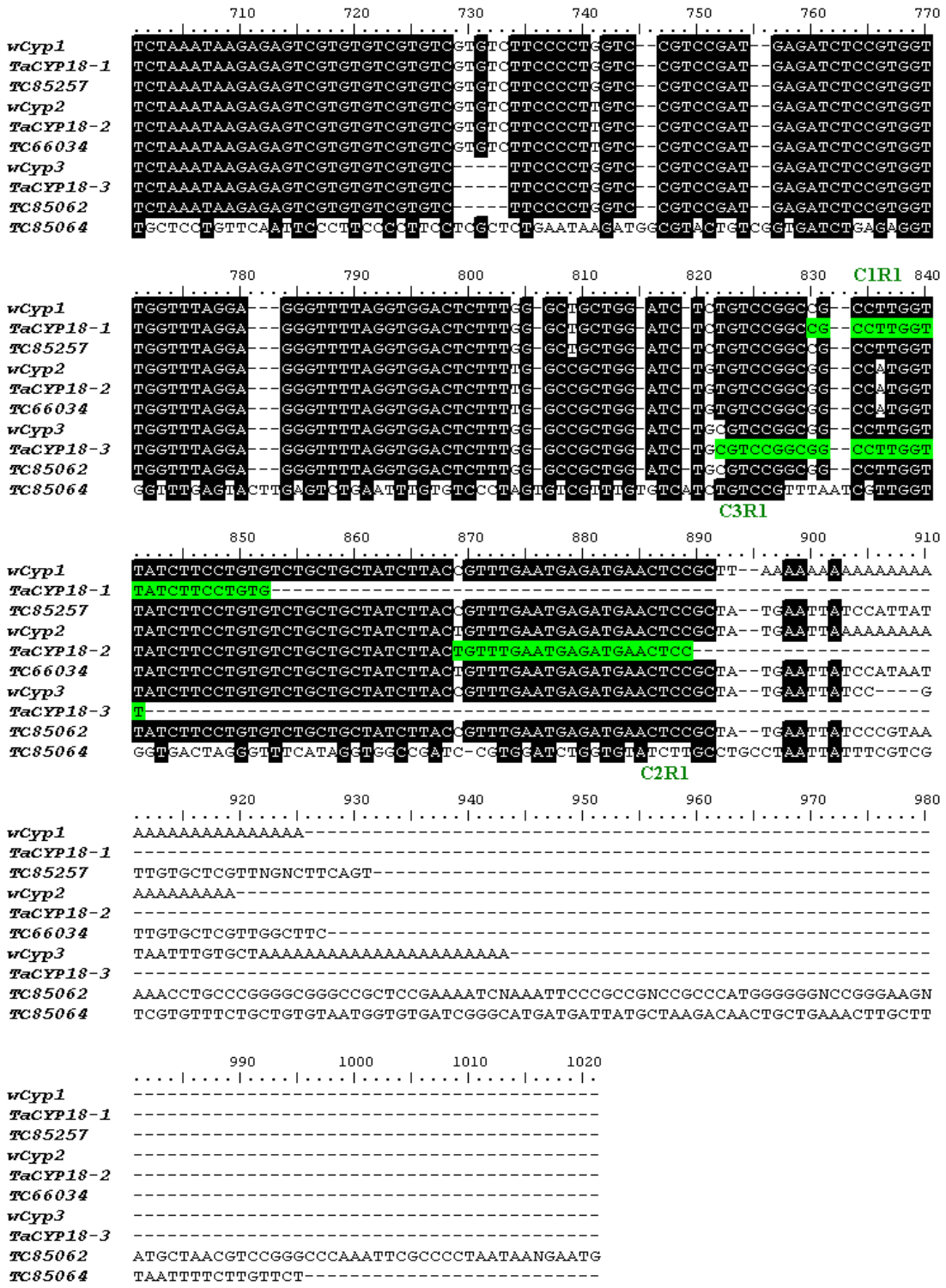


Figure 6.3 ***Cyclophilin A*** **genes and transcripts of wheat**

An alignment of the partial *cyclophilin A* genes of *T. aestivum*, the putative *cyclophilin A* TC sequences isolated from the TIGR TaGI database and the previously reported wheat *cyclophilin A* cDNA clones (Johnson *et al*. 2001). All eight *T. aestivum* sequences were found to be identical and are thus, shown as single sequences. The AS-primers used to amplify the genes are highlighted in green and the 'start'(ATG) and 'stop'(TAG/TGA) codons are highlighted in pink.

6.3 PHYSICAL MAPPING OF THE THREE *CYCLOPHILIN A* GENES

The chromosomal locus of each of the above *cyclophilin A* genes was investigated using allele-specific PCR on the 42 possible nullisomic/tetrasomic aneuploid stocks of *T. aestivum* cv. Chinese Spring. The *wCyp1*-specific primer pair C1F1/C1R1, used to amplify *TaCYP18-1,* revealed the expected PCR product in all of the nullisomic/tetrasomic lines, except 6AAB and 6ABB (Appendix IV-2; Figure 6.4A). Amplifications conducted on the ditelosomic lines for the short arms of these chromosomes and lines with partial chromosome deletions (Table 2.2), once again revealed the presence of the expected product in the three ditelosomic lines DT6AS, DT6BS and DT6DS and in all of the deletion lines except 6DS4 and 6DS2 (Figure 6.4A). The *wCyp2*-specific primer pair, C2F1/C2R1, used to specifically amplify *TaCYP18-2*, revealed the expected PCR product in all of the nullisomic/tetrasomic lines except 6AAD and 6ADD (Appendix IV-2), in the ditelosomic lines DT6AS, DT6BS and DT6DS and in all of the deletion lines except 6BS2, 6BS3 and 6BS5 (Figure 6.4B). Likewise, amplifications with the *wCyp3*-specific primer pair C3F1/C3R1 revealed the expected products in all of the nullisomic/tetrasomic lines except 6BBD and 6BDD (Appendix IV-2), in the ditelosomic lines DT6AS, DT6BS and DT6DS and in all of the deletion lines except 6AS1 (Figure 6.4C). These results thus suggest the three *cyclophilin A* genes, *TaCYP18-1, TaCYP18-2* and *TaCYP18-3* are located on the distal sections of chromosomes 6DS, 6BS and 6AS respectively (Figure 6.5).

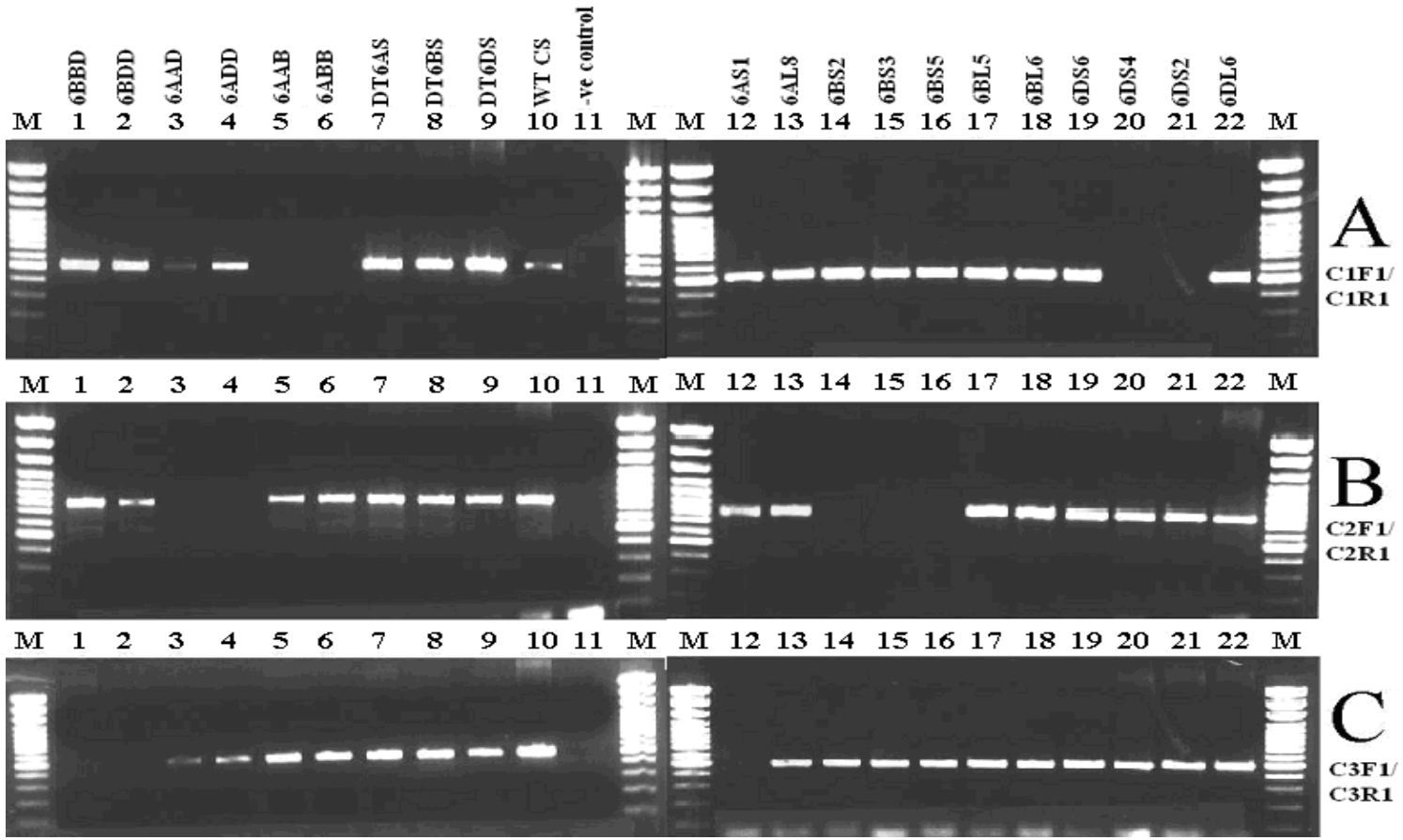


Figure 6.4 Physical mapping of the *cyclophilin A* genes of wheat

Allele-specific PCR amplification of gDNAs of selected nullisomic/ tetrasomic, ditelosomic and deletion lines of chromosome 6 of Chinese Spring. Primers details are provided in Table 2.5. Gel A: C1F1/C1R1; Gel B: C2F1/C2R1; Gel C: C3F1/C3R1. Lanes: M: 100bp ladder; 1: 6BBD; 2: 6BDD; 3: 6AAD; 4: 6ADD; 5: 6AAB; 6: 6ABB; 7: DT6AS; 8: DT6BS; 9: DT6DS; 10: wild type Chinese Spring gDNA; 11: no-template negative control; 12: 6AS1; 13: 6AL8; 14: 6BS2; 15: 6BS3; 16: 6BS5; 17: 6BL5; 18: 6BL6; 19: 6DS6; 20: 6DS4; 21: 6DS2; 22: 6DL6.

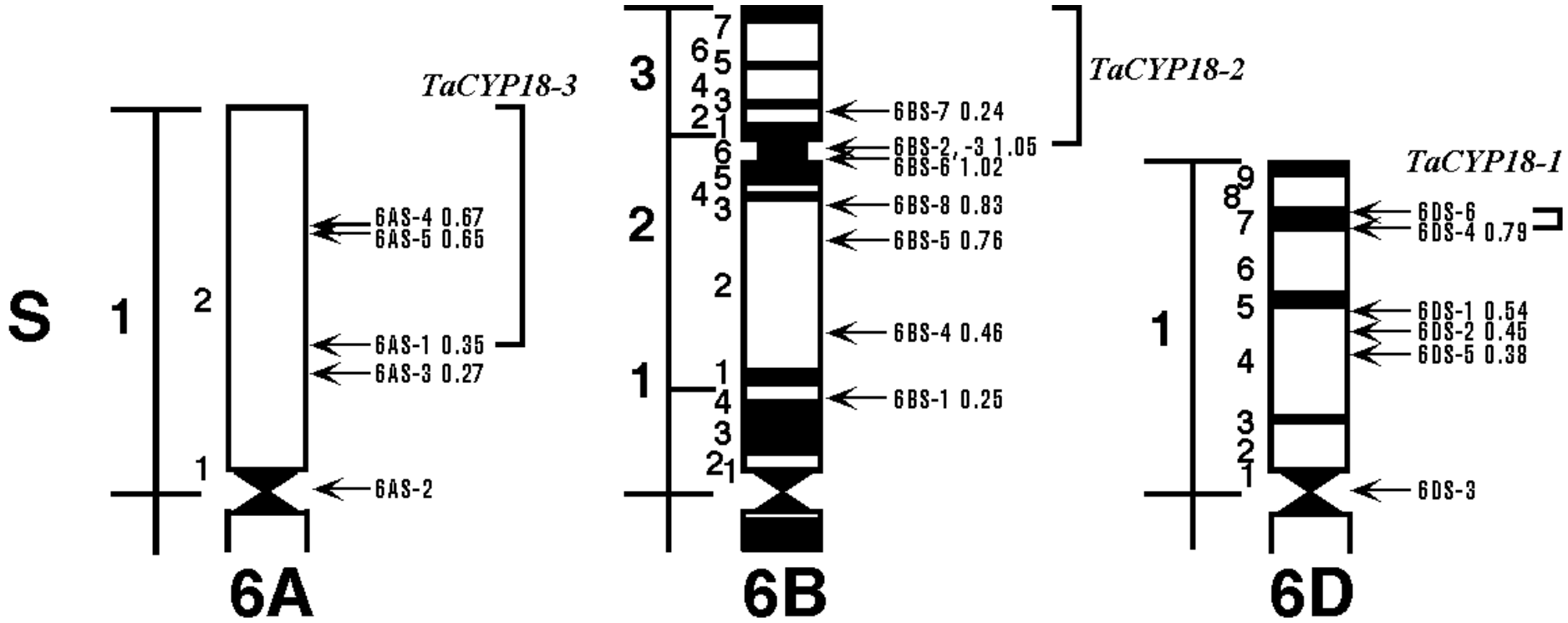


Figure 6.5 Chromosomal loci of the *cyclophilin A* genes of *T. aestivum*

The physical chromosomal loci of the three characterised *cyclophilin A* genes of wheat on the short arms of chromosome 6 (chromosome figure and deletion breakpoints from Endo and Gill, 1996).

6.4 SOUTHERN BLOT OF *CYCLOPHILIN A* IN *T. aestivum*

In order to investigate whether the *cyclophilin A* gene family was limited to the three genes isolated and characterised in the previous sections (Section 6.2 and 6.3) a Southern blot was conducted. Analysis of the Southern blot of *T. aestivum* cv. Chinese Spring gDNA probed with a *cyclophilin A* cDNA probe revealed the presence of eight fragments, of sizes of approximately 1.6, 1.8, 2.3, 3.0, 4.8, 6.2, 9.0 and >10kb (a few of these being of higher intensity than the others) in the *Bam*HI digest, and four fragments, of sizes of approximately 5.0, 6.4, 8.7 and >10kb, in the *Hind*III digest (Figure 6.6), a curious observation, considering there are no internal *Hind*III sites in the three cDNAs and, thus, suggesting the presence of a fourth *cyclophilin A* gene with an internal *Bam*HI site, similar to that in the three cDNAs.

6.5 IDENTIFYING THE *CYCLOPHILINS* ON THE SOUTHERN BLOT

To address whether the four bands represented (at least) four *different cyclophilin A* genes (for one/some of which we had perhaps not found the corresponding cDNA), or whether there were any gene duplications, sections corresponding to the four *Hind*III bands (Figure 6.6) were cut out from another agarose gel containing a *Hind*III digest of gDNA of *T. aestivum* cv. Chinese Spring. The DNA purified from all of these bands produced amplification products of approximately 800bp using the PPI5/PPI3 consensus primers (Figure 6.7), which were then subjected to a quick, preliminary RFLP analysis. As the characterised sections of the *cyclophilin A* genes revealed an absence of introns (Section 6.2.2) the restriction maps of the cDNAs, which revealed RFLPs for *Nco*I and *Eag*I (Figure 2.5), were used to identify each of the corresponding genes amplified from the fractionated gDNA. Thus, the PCR product originating from the 5kb size fraction showed an RFLP pattern similar to that of the cDNA *wCyp1* (i.e. 363bp and 446bp products in the *Nco*I digest, a 655bp product in the *Eag*I digest and 385bp and 270bp products in the *Eag*I/*Nco*I digest), the PCR product from the 6.4kb fraction showed an RFLP pattern similar to that of the cDNA *wCyp2* (i.e. 355bp and 375bp products in the *Nco*I digest, a 700bp product in the *Eag*I digest and 375bp and 270bp products in the *Eag*I/*Nco*I digest) and the PCR products from *both* the 8.7kb and 10kb fractions were similar to *wCyp3* (i.e. 363bp and 434bp products in the *Nco*I and *Eag*I/*Nco*I digests and no digestion with *Eag*I) (Figure 6.7).

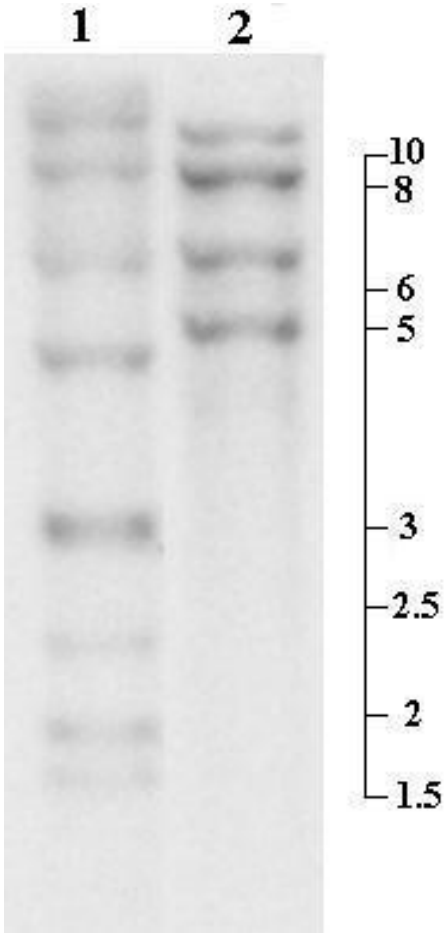


Figure 6.6 ***Cyclophilin A*** **Southern blot of** ***T. aestivum***

The Southern blot of *T. aestivum* cv. Chinese Spring gDNA, digested with *Bam*HI (lane 1) and *Hind*III (lane 2) and probed with a 797bp *cyclophilin A* probe amplified from the cDNA clone *wCyp3* with the primers PPI5/PPI3 (Table 2.7). The DNA molecular weight marker (kb) is shown on right.

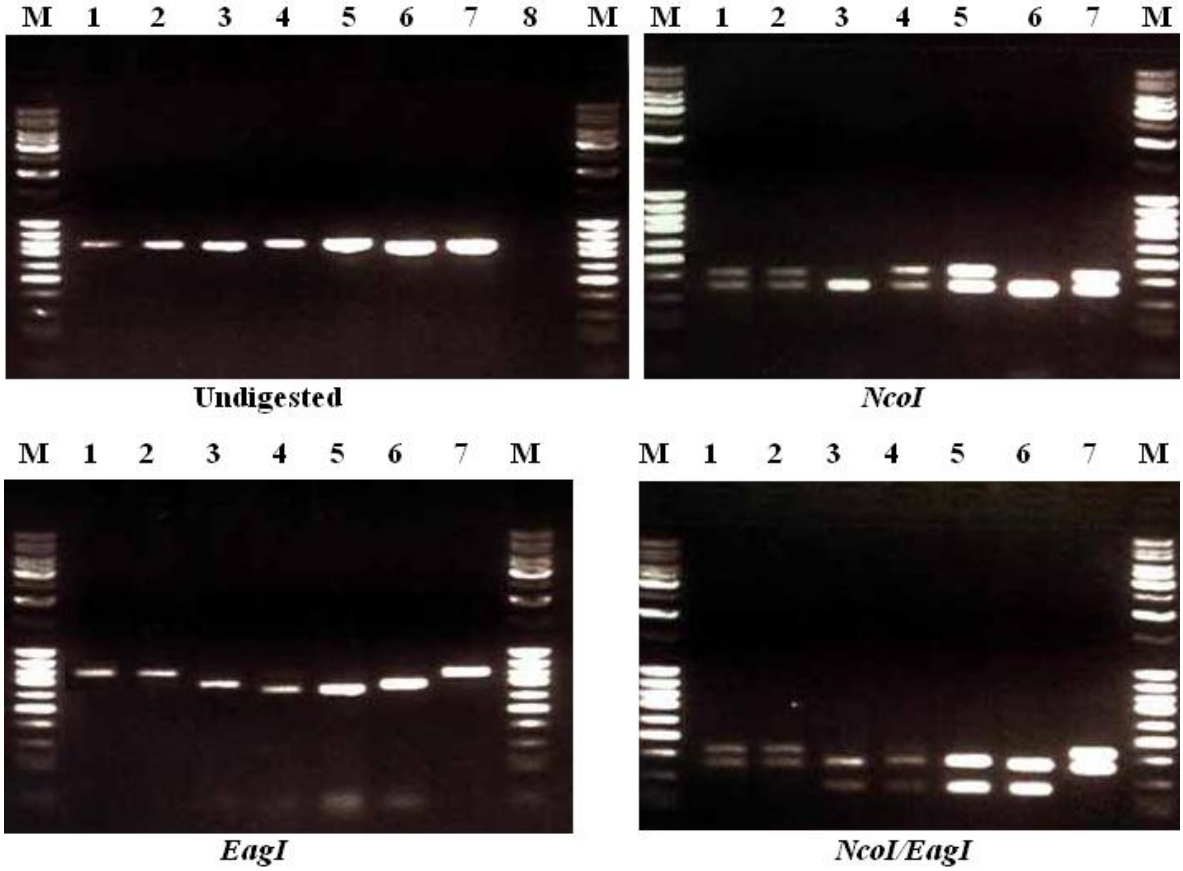


Figure 6.7 RFLP analysis of amplified *cyclophilin A* genes

RFLP patterns of *cyclophilin A* genes, amplified with the PPI5/PP3 consensus primers, from the gel sections corresponding to the four hybridizing fragments observed in the *Hind*III digest (Figure 6.6). The PCR products are shown undigested and after *Nco*I, *Eag*I and *Nco*I/*Eag*I digestion. Lanes: M: 100bp DNA ladder; 1: 10-15 kb gel fraction; 2: 8-9 kb gel fraction; 3: 6-7 kb gel fraction; 4: 4.5-5.5kb gel fraction; 5: *wCyp1*; *6*: *wCyp2*; 7: *wCyp3*; 8: no-template negative control.

6.6 *IN SILICO* IDENTIFICATION OF NOVEL *CYCLOPHILIN* GENES

In an effort to identify and characterise additional members of the *cyclophilin* gene family in wheat a query of an EST database containing sequences from a wheat endosperm cDNA library (Section 2.19.1) revealed 46 ESTs containing the term 'cyclophilin' in their most significant BLASTn 'hit'. An alignment of these sequences revealed that 40 of these were >95% identical to the *cyclophilin A* cDNA clones, the differences being present at the 3' ends, possibly due to low quality sequence data (raw data not shown). Of the remaining 6 clones, one was similar to another class of PPIase, the FKBPs, and as such was not studied further. The inserts of the five remaining clones were amplified with vector-based primers, revealing sizes of <300bp for clones 1299 and 1300, ~1000bp for clones 2629 and 4260 and ~2200bp for clone 2463 (Figure 6.8); the first two clones were disregarded as probably partial cDNAs and the remaining three (clones 2629, 4260 and 2463) were characterised further.

6.6.1 Identification of putative chloroplast-localised *cyclophilin* genes

Full length sequencing of clones 2629 and 4260 revealed 942bp and 995bp *cyclophilin*-like sequences, henceforth referred to as *TaCYP20-1* (AY217751) and *TaCYP20-2* (AY217753), respectively. Both contained an open reading frame (ORF) of 738bp, translation of which revealed basic proteins of 245 amino acids acids (Figure 6.9), with estimated molecular weights of approximately 25.9 kDa, isoelectric points (pI) of 9.8 (TaCYP20-1) and 10.1 (TaCYP20-2) and the presence of the conserved tryptophan residue required for cyclosporin A binding (Section 1.6.2; Bossard *et al.*, 1991). The clones exhibited 88.3% DNA sequence identity to each other, most of the differences consisting of a longer 3' untranslated region (UTR) in clone *TaCYP20-2*, insertion/deletions in the putative 5' and 3' UTRs, and SNPs throughout the putative ORFs. Both clones exhibited limited DNA sequence identity (51-53%) to the corresponding sections of the *cyclophilin A* cDNAs, but a higher degree of similarity (75%) to the appropriate sections of their putative amino acid sequences, the major differences being the presence of a 74 amino acid extension at the N-termini, three additional amino acids (NPV) at this end, and the lack of a seven amino acid insertion, characteristic of plant cyclophilins (Section 1.6.3), in both TaCYP20-1 and TaCYP20-2 (Figure 6.9). BLASTx analyses of the TaCYP20-1 sequence identified the chloroplast-

localised TLP20 from *A. thaliana* (Q9ASS6) and a peptide sequence (AAG03106) translated from the partial gene encoding the assumed rice orthologue of TLP20, as the most highly significant matches in the non-redundant ("nr") protein database. Further evidence suggesting the chloroplast localization of these proteins was provided by the software TargetP, used to determine the most probable intracellular location of proteins based on their N-terminal amino acid sequences, which placed them in reliability class 1 for chloroplast localisation. Another program, PSORT, also predicted the chloroplast stroma as the most probable intracellular location of these proteins. Analysis for conserved motifs revealed the presence of a potential ATP/GTP binding site (P-loop), GNGTGGKS, at position 156-163 in both the TaCYP20-1 and TaCYP20-2 sequences (Figures 6.9 and 6.10).

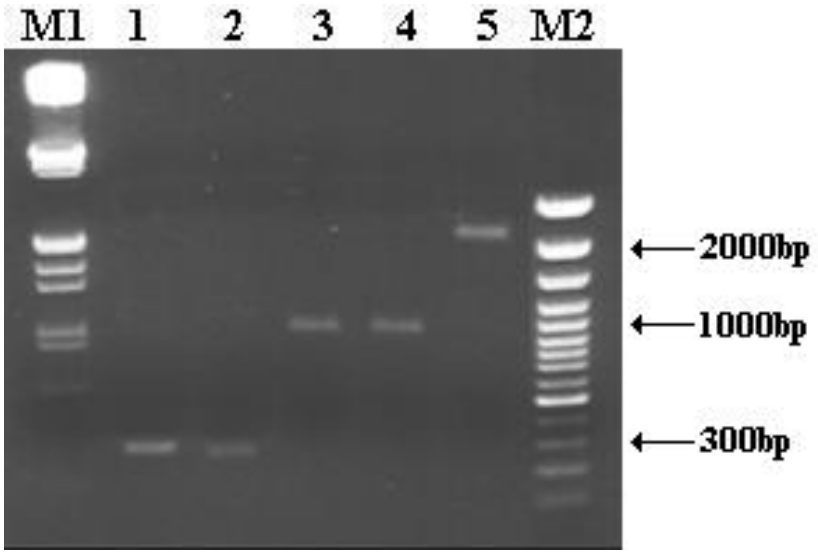


Figure 6.8 Amplification of novel *cyclophilins* from a wheat endosperm library

The PCR products amplified from the candidate, novel *cyclophilin* plasmid clones isolated from a wheat endosperm cDNA library. M1: λ *Eco*RI/*Hind*III; 1: Clone 1299; 2: Clone 1300; 3: Clone 2629; 4: Clone 4260; 5: Clone 2463; M2: 100bp ladder.

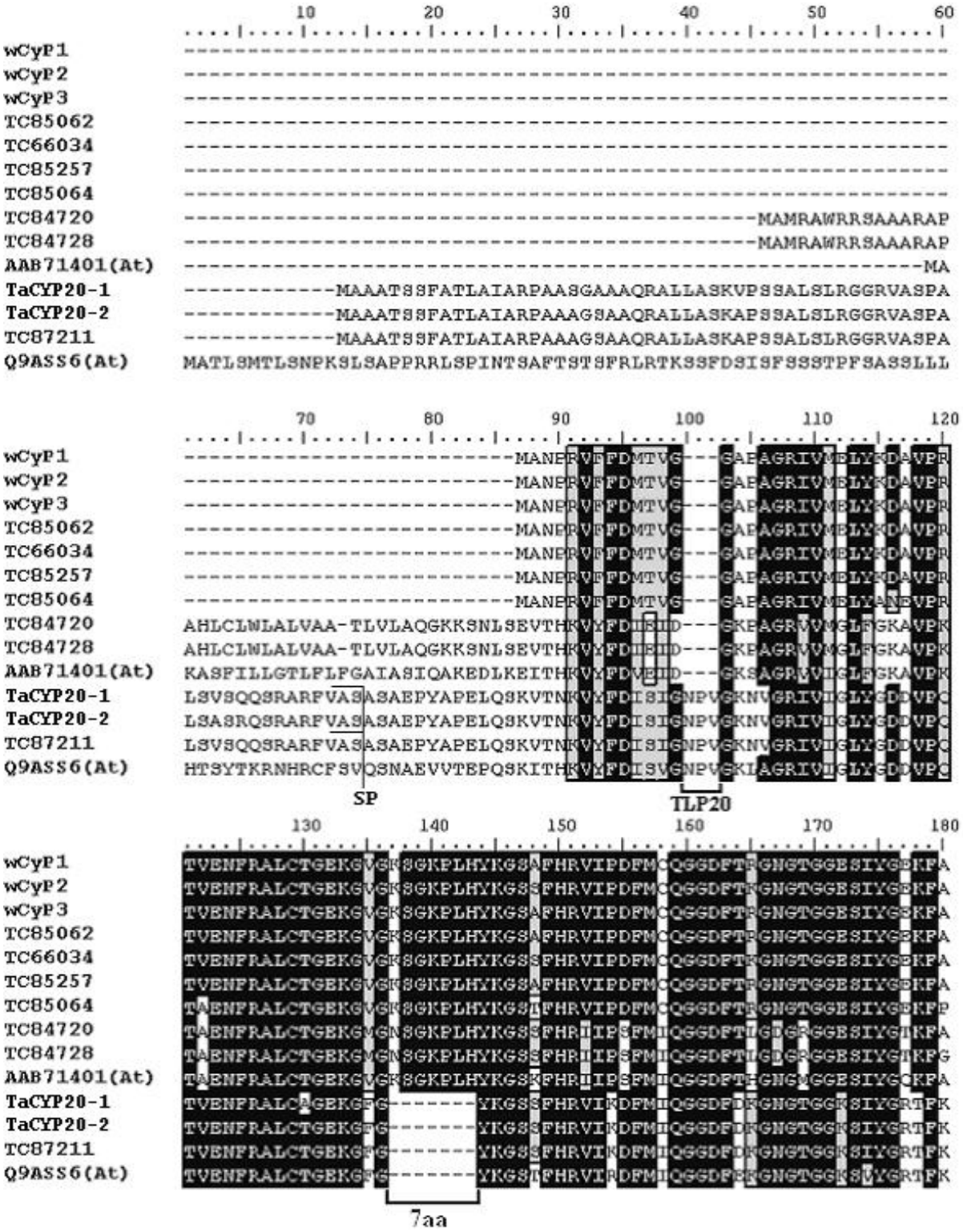


Figure 6.9 continued.

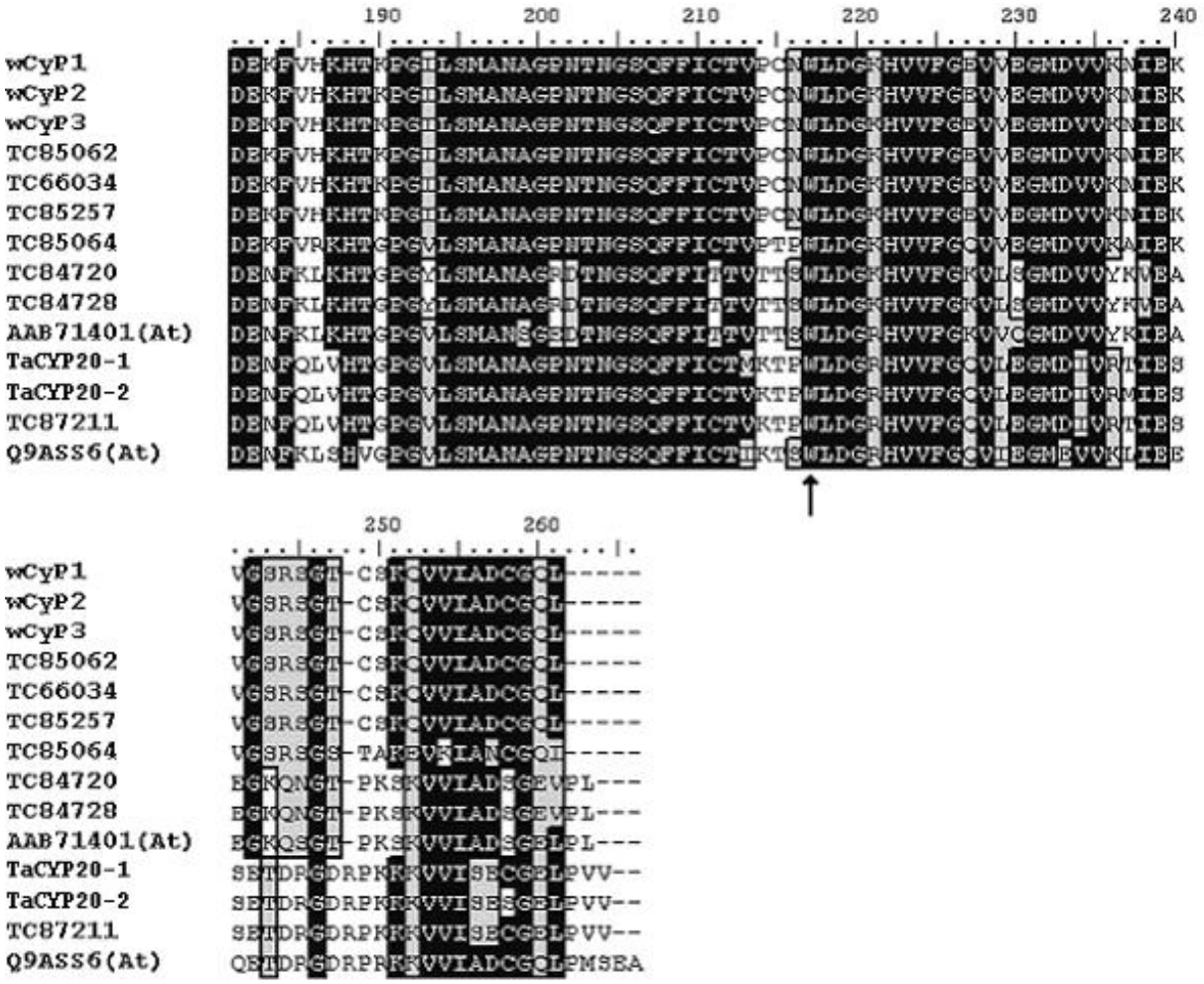


Figure 6.9 Alignment of the putative protein products of the *cyclophilin* cDNAs and relevant *cyclophilin* TC sequences

The wheat cyclophilin A sequences wCyp1, wCyp2 and wCyp3 (Johnson *et al.*, 2001) correspond to the TIGR TC sequences TC85062, TC66034 and TC85257, respectively. TC85064 represents a novel putative wheat *cyclophilin A* protein. TC84720 and TC84728 are putative wheat cyclophilin B sequences, shown aligned with an ER-resident cyclophilin (AAB71401) from *Arabidopsis thaliana* (At) (Saito *et al.*, 1999). The putative wheat chloroplast thylakoid lumen cyclophilins TaCYP20-1 and TaCYP20-2 are aligned with TC87211 and a TLP20 chloroplast cyclophilin from *A. thaliana* (Q9ASS6) (Schubert *et al.*, 2002). The 7 amino acid insertion (7aa), typical of plant cyclophilin A's, the TLP20 'fingerprint' insert of three amino acids in the chloroplast cyclophilins, and the NPV (TLP20) are highlighted. The tryptophan residue essential for binding to cyclosporin A is indicated with an arrow.

6.6.2 Identification of a putative nuclear-localised *cyclophilin* gene

Full length sequencing of the clone 2463, isolated as mentioned above (Section 6.6), revealed a 2205bp insert, henceforth referred to as *TaCYP-70* (AY217752), that exhibited limited (~42-44%) DNA sequence identity to the comparable sections of the three *cyclophilin A* cDNAs, with BLASTn analysis revealing highest identity (85%) to a stretch of 115bp in an uncharacterised 516bp maize cDNA sequence (AY111069). *TaCYP-70* contained a putative 1911bp ORF, potentially encoding a 636 amino acid, 69.9 kDa protein product with a predicted pI of 11.1 and also containing the conserved tryptophan residue (Figure 6.9). The comparable N-terminal section of wCyp-70 exhibited 61-63% similarity to the putative proteins of the three *cyclophilin A* cDNAs and its remaining central and C-terminal sections contained serine rich (41.7% of residues 229-324 and 38.6% of residues 586-629) and arginine rich (25.4% of residues 450-622) sections (Figure 6.10). A BLASTp search of the non-redundant (nr) protein database identified some similarity (~60%) to the cyclophilin-like domain (204 amino acids) and the C-terminal (56 amino acids) of an uncharacterised 570 amino acid putative protein from *A. thaliana*. PSORT localised TaCYP-70 to the nucleus with a high level of confidence (0.987), primarily based on the presence of five putative nuclear localization signals (NLS), fitting the motif outlined in the ProSite database (Figure 6.10), however, TargetP failed to provide a significant result, probably due to this program not predicting nuclear proteins.

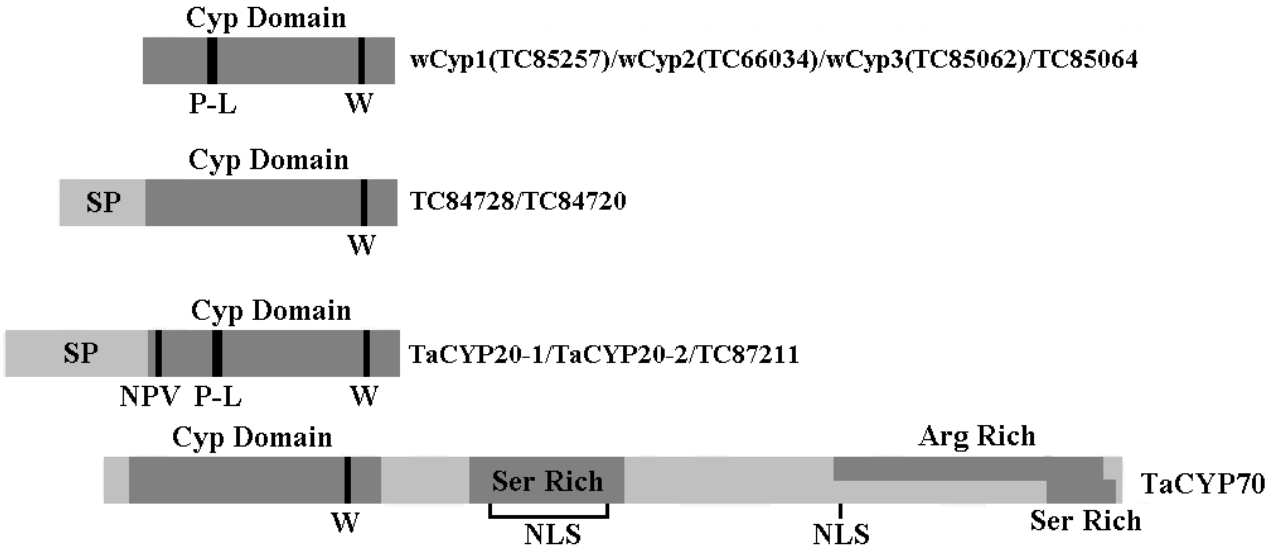


Figure 6.10 Characteristic features of the putative wheat cyclophilins

The putative proteins encoded by the cDNA clones isolated from a wheat endosperm cDNA library or the TIGR tentative consensus sequences. Cyp Domain: The cyclophilin-like domain of the archetypal cyclophilin A; P-L: ATP/GTP-binding P-loop; W: the tryptophan residue essential for binding to cyclosporin A; SP: putative signal peptide; NPV: the fingerprint motif of the chloroplast localized cyclophilin, TLP20; NLS: nuclear localizing signal; Arg/Ser Rich: arginine or serine rich domains.

6.6.3 Screening of the TIGR *T. aestivum* gene index for *cyclophilins*

In order to identify any other *cyclophilin A* gene(s) or members of the *cyclophilin* superfamily that could hybridize to the probe used for the Southern blot, we searched the TIGR *T. aestivum* gene index (TaGI), assembled from ESTs expressed in different tissues and cultivars and under various environmental conditions (413,955 ESTs as at version 6.0, 1/2003). A BLASTn search of this database with the *wCyp1* cDNA sequence identified eight putative *cyclophilin* TCs (*TC66034*, *TC85062*, *TC85257*, *TC85064*, *TC85065*, *TC85263*, *TC85068* and *TC85069*). Alignment of these sequences revealed *TC85257*, *TC66034* and *TC85062* corresponded to the three cDNAs *wCyp1*, *wCyp2* and *wCyp3* and *TC85065* represented a novel, full length sequence (alignment not shown). However, the four other TC sequences identified in this search were truncated at their 5'or 3'ends relative to the *wCyp* cDNA sequences (alignment not shown) and thus, potentially represent partial sequences assembled separately due to the quality of their respective EST sequences. Thus, to increase the chances of identifying only full-length *cyclophilin* sequences a tBLASTx search using the *wCyp1* cDNA sequence was conducted. This search revealed several tentative consensus sequences (TCs) putatively encoding proteins with high similarity (>68%) to that encoded by the query sequence *wCyp1* (Table 6.1), their alignments falling into three main groups: (i) *cyclophilin A*-like sequences; (ii) a *TaCYP20*-like sequence; and (iii) a new class (Table 6.1; Figure 6.9).

Among the first group, three TCs (*TC85257*, *TC66034* and *TC85062*) corresponded to our three *cyclophilin A* sequences (Johnson *et al.*, 2001), however, a fourth novel sequence (*TC85064*) exhibited only 59-62% DNA identity to these, most of the divergence being at the 5'and 3' UTRs (Figure 6.3). Its translation revealed a putative protein of the same size as the other wheat cyclophilin A's, i.e. 171 amino acids, with 87-88% amino acid identity and ~93% similarity to these, the presence of the seven amino acids characteristic of plant cyclophilins and the tryptophan residue essential for cyclosporin A binding (Liu *et al.*, 1991; Section 1.6.2) (Figure 6.9).

The second group was represented by a single 1054bp long sequence, *TC87211*, which had a high identity (95% and 99%) to the *TaCYP-20* cDNAs, its putative protein product lacking the plant cyclophilin-specific 7 amino acids. The presence of a single

TC within the TaGI, when at least two distinct genes of this class, *TaCYP20-1* and *TaCYP20-2*, are expressed in the endosperm, can be explained by the method used to construct the TaGI database: overlapping ESTs only need 94% identity over 40bp to be placed in a single contig, therefore, the closely related homeoalleles of hexaploid wheat will commonly be grouped together as hybrid TCs. In support of this, all but 6 of the 35 SNPs between the ESTs used to assemble *TC87211* reported in the TC description can be explained by differences between *TaCYP20-1* and *TaCYP20-2*, the other six SNPs representing either true inter-cultivar polymorphism or polymorphism present in an additional so-far uncharacterized homeoallele.

The third group was represented by the two closely related contigs, *TC84720* and *TC84728*, which were 892bp and 990bp long, respectively, the differences in length being at their 5' and 3' ends, and had 92% DNA identity and >99% amino acid similarity over comparable sections. Once again, these two sequences may represent 'hybrid TCs', constructed from more than two distinct cDNA classes. Comparison of these TCs with the three *cyclophilin A* sequences revealed 53-54% DNA identity and BLASTn query of the GenBank database with both TCs revealed several sequences with high DNA identity (86-99%) over appropriate lengths, including cDNA clones from wheat, rice and maize, the latter encoding the short peptide previously sequenced from a cyclophilin B protein from the ER of maize (Sheldon and Venis, 1996). The ORFs in *TC84720* and *TC84728* encoded 213 amino acid putative protein products consisting of a cyclophilin-like domain with a novel N-terminal extension and the characteristic 7-amino acid insertion, however a recognizable C-terminal ER-retention signal, such as KDEL (Denecke *et al.*, 1992) or VEKPFAIAKE (Arber *et al.*, 1992) was absent (Figures 6.9 and 6.10). Interestingly, a BLASTx query with these identified a 201 amino acid ER-localised protein from *Arabidopsis thaliana*, encoded by *CYP5* (*AtCYP19-4*) (AAB71401) (Saito *et al.*, 1999), the alignment revealing a slightly longer putative signal peptide in the wheat proteins but high sequence conservation throughout the rest of the sequence (88% identity, 96% similarity) (Figure 6.6), providing further evidence that these two TCs may encode cyclophilin B-like proteins. PSORT identified these proteins as "outside", suggesting targeting to the ER, followed by secretion, while TargetP localized them to the mitochondria with a score of 0.848, but at a lower confidence level (of 3) due to a significant, albeit lower score (0.429) that these proteins are found within the secretory pathway. The absence of the typical ER-retention signal,

KDEL (Denecke *et al.*, 1992), at the C-terminal of cyclophilin B's may explain why these software packages did not predict the ER as their a potential intracellular location, as PSORT uses this particular motif as strong evidence of an ER-residency.

As the tBLASTx search of the TIGR TaGI database was conducted in 1/2003 on version 6.0, the current database (version 9.0, 9/2004) was queried to identify the current TC sequences corresponding to the TC sequences described above. This search revealed that all of the above-described cyclophilins are present in version 9.0 of the database and their current TC sequences are presented in Table 6.1. Notably, the presence of the *TaCYP20-2* and *TaCYP20-1* in the GenBank database has led to the splitting of *TC87211* into two distinct TC sequences in the current database, *TC192597* and *TC222490*, corresponding to the two putative chloroplast cyclophilins. Further, the novel *cyclophilin A TC85064* is now represented by *TC211285*, which was assembled with only 8 ESTs; thus this sequence would have been filtered out as probably representing a TC generated from low-quality EST sequences.

Table 6.1 Cyclophilin-related TC sequences identified in a tBLASTx search of the TIGR TaGI database

Group	TIGR TaGI v6.0 TC# (# of ESTs)*	Tentative classification	TIGR TaGI v9.0 TC# (# of ESTs)*
1	*TC66034* (241) *TC85062* (385) *TC85257* (267) *TC85064* (12)	Cyclophilin A (cytoplasmic)	*TC206054* (259) *TC190359* (374) *TC206056* (232) *TC211285* (8)
2	*TC87211* (35)	TLP20 orthologue (chloroplast thylakoid lumen)	*TC192597* (27) *TC222490* (24)
3	*TC84720* (27) *TC84728* (88)	Cyclophilin B (ER)	*TC205877* (47) *TC205883* (97)

*The # of ESTs describes the number of EST sequences used to assemble the contig sequence.

6.7 DISCUSSION

The potential role of cyclophilins in the developing wheat endosperm has been discussed in relation to the folding of the storage proteins in this tissue (Section 1.6.4). This chapter describes another important step in elucidating such a role, i.e., characterisation of the gene family encoding the cytosolic cyclophilin As, with respect to their sequences, any differences therein, and the copy number, to allow their physical and genetic mapping, thereby determining any association with quantitative traits and their potential use in marker-assisted selection by wheat breeders.

6.7.1 The *cyclophilin A* genes of wheat are intronless and located on the group 6 chromosomes

Our earlier isolation of three classes of the putatively-cytoplasmic *cyclophilin A* cDNAs from a wheat endosperm library (Johnson *et al.*, 2001) raised questions in our mind as to the role(s) of these enzymes in this tissue and any association with QTLs for agronomically important traits. To address this, characterisation of these genes was undertaken in the same eight cultivars that had been used to establish genetic maps in wheat (Kammholz *et al.*, 2001). One of the first aims was to identify any introns, as these would serve as sources of greater sequence divergence, which would be helpful for creating markers for the genetic mapping process. However, sequencing of the genomic copies (Section 6.2.2) revealed an absence of introns, similar to the *cyclophilin A* genes in rice (Buchholz *et al.*, 1994) and the cytoplasmic *cyclophilins* of *Arabidopsis* (Chou and Gasser, 1997). Absence of inter-cultivar polymorphisms in the coding sequences of these genes in all eight cultivars has further prevented the design of molecular markers for these genes at present, and investigations into the flanking regions would be essential for identification of any polymorphisms and establishment of tightly linked markers, to enable the analysis of any genetic association of these genes with any QTLs.

Despite these limitations, the specificity of the three allele-specific primer pairs was successfully exploited in determining the chromosomal loci of the three genes, the absence of PCR products in the specific nullisomic/tetrasomic, ditelosomic and deletion

stocks revealing *TaCYP18-1, TaCYP18-2* and *TaCYP18-3* to be located on the distal sections of chromosomes 6DS, 6BS and 6AS, respectively. This region has been found to be rich in other genes of interest including those for polyphenol oxidase (*Ppo*) and gliadins (*Gli-2*) (Li *et al.*, 1999; Weng *et al.*, 2000). The short arm of chromosome 6 has also been implicated in high protein content in durum wheat, although this QTL appears to be more towards the centromere (Joppa *et al.*, 1997).

6.7.2 The *cyclophilin A* genes may form a small gene family in wheat

While the physical mapping results (Section 6.3) and the identification of only three types of cDNAs (Johnson *et al.* 2001) suggested three *cyclophilin A* genes in bread wheat, it was important to ascertain the size of this gene family due to reports of *cyclophilins* comprising multigene families in several diploid plant species (Marivet *et al.*, 1992; Buchholz *et al.*, 1994; Saito *et al.*, 1995; Romano *et al.*, 2004). The observations of four hybridizing bands in the *Hind*III digests of gDNA (Figure 6.6), the lack of introns (ruling out intron-based *Hin*dIII sites), and absence of internal *Hind*III sites in the three cDNAs together suggest the presence of four genes encoding cyclophilin A in wheat. The observation of eight major hybridising fragments in the *Bam*HI digest, an enzyme expected to cut the three cDNAs and the corresponding genomic copies internally once, also supports the existence of one other gene, also containing an internal *Bam*HI site. Interestingly, preliminary restriction mapping (Figure 6.7) revealed that *two* of the four genes shared their RFLP patterns for *Eag*I and *Nco*I with *wCyp3*. This finding raises various possibilities, e.g., (i) that the *cyclophilin A* gene on chromosome 6AS (*TaCYP18-3*) is present in (at least) two copies, located on two different *Hind*III fragments but having identical maps for the internal *Bam*HI, *Eag*I and *Nco*I sites and thus perhaps identical sequences; or (ii) one of these is a variant *cyclophilin A* gene, with the above restriction sites and adequate homology to the probe but an otherwise divergent sequence (explaining why it may not have been amplified from the gDNAs by the allele-specific primers) and/or lack of expression in the endosperm (explaining why a corresponding cDNA was not found by us; and/or (iii) this could be a pseudogene, as reported for human *cyclophilins* (Haendler and Hofer, 1990). Sequencing of these two PCR products would be required to address these issues.

6.7.3 Identification of variants of *cyclophilin A* in wheat

The search for a variant, fourth *cyclophilin A* sequence was conducted using an endosperm-specific EST database (Section 6.6) and the TIGR TaGI database assembled from wheat sequences publicly-available in the GenBank databases. The sequences identified in the wheat endosperm library database (Section 6.6) were either *wCyp1*-, *wCyp2*- or *wCyp3*-like sequences, or the *TaCYP-20* and *TaCYP-70* clones; the latter two appeared to be too divergent (53-54% and 47% identity) to represent the fourth hybridising fragment. However, these clones did represent novel *cyclophilins* and were thus characterised further and are discussed below (Sections 6.7.4 and 6.7.5).

The search of the TIGR TaGI database for a variant *cyclophilin A* with a BLASTn query of the TaGI database with *wCyp1* (Section 6.6.3) intriguingly revealed four cyclophilin A-like cDNA sequences containing full length ORFs, however this search also revealed a number of partial sequences. A tBLASTx search with *wCyp1* filtered out these partial sequences as the absence of full-length ORFs in the partial sequences, which the tBLASTx program translates (and then back-translates) would have greatly reduced their similarity to the *wCyp1* query sequence in this context. Thus, the tBLASTx search of the TaGI database identified the four *cyclophilin A*-like TCs identified in the BLASTn search (above) and also identified three other TCs encoding putative chloroplast and ER-localised cyclophilins (Table 6.1) (discussed below; Sections 6.7.4 and 6.7.6). The fourth, novel *cyclophilin A*-like sequence identified in the BLASTn and tBLASTx searches, *TC85064*, was investigated as it could have potentially represented the fourth hybridising fragment observed on the Southern blot (Figure 6.6) as the putative protein encoded by this TC sequence was highly similar (~93%) to cyclophilin A. The twelve ESTs used to construct this TC were from seedling and meiotic anther cDNA libraries, possibly explaining the absence of this type of cDNA from our extensive analyses of >40 *cyclophilin* clones from an endosperm library (Johnson *et al.*, 2001; and unpublished). As such, *TC85064* could represent a different class of cyclophilin A with tissue-specific expression, however, further analysis suggests that it is unlikely to represent the extra bands on the Southern blot, due to its limited DNA sequence identity (~66%) to the probe used, the high stringency hybridisation conditions and the lack of an internal *Bam*HI site in it, required to explain the additional bands. Further, it would not be amplified by the consensus primers used for amplifying

the *cyclophilin A* genes from the *Hind*III fragments or the allele-specific primers used for mapping studies, and it does not share the RFLP patterns for *Eag*I and *Nco*I with *wCyp3*. Thus, while it is possible that divergent *cyclophilin A*-like sequences with other expression patterns exist in wheat, the most likely explanation for the Southern blot and physical mapping results together appears to be that the *cyclophilin A*'s expressed in the endosperm comprise a small multigene family with three homeologous loci on chromosome 6, that at 6AS having (at least) two gene copies.

6.7.4 Putative plastid *cyclophilins* are expressed in the wheat endosperm

The isolation of novel *cyclophilin* sequences from a wheat endosperm cDNA library putatively localising to the chloroplast (Section 6.6.1) and identification of related TCs in the TaGI database (Section 6.6.3) are significant findings, as the starch biosynthetic pathways which determine important wheat quality traits are localized to the amyloplast in this tissue. The presence of a TLP20 'fingerprint', i.e., the 'NPV' insertion, in proteins encoded by *TaCYP20-1*, *TaCYP20-2* and *TC87211* suggest strongly that these transcripts encode orthologs of *A. thaliana*, rice and spinach TLP20 cyclophilin, which has been shown to be responsible for most of the PPIase activity in the chloroplast lumen and proposed to be an important catalyst of protein folding here (Edvardsson *et al.*, 2003). Thus, the plastid-localised wheat cyclophilins may be involved in the folding and/or stabilization of enzymes involved in starch biosynthesis pathways and play a role in influencing starch quality traits; further investigations of the corresponding genes and their expression patterns will allow their role(s) to be determined more clearly.

6.7.5 A putative nuclear SR-cyclophilin is expressed in the wheat endosperm

The presence and sequence of the six NLS's in the putative amino acid sequence of *TaCYP70* (Section 6.6.2) suggests that this is a novel multidomain protein with a CLD. A multidomain cyclophilin, AtCYP63, containing a similar pattern of functional motifs, including an N-terminal CLD, NLS signals and a RS-rich region at the C-terminal has recently been identified in *Arabidopsis* and is suggested to be an SR-protein, involved in constitutive and alternative splicing of RNA (Romano *et al.*, 2004). The relatively high sequence identity (40%) and similarity (60%) between TaCYP70 and AtCYP63

and the similar pattern of functional motifs suggests a similar function for TaCYP70 in wheat. The TIGR TaGI database did not reveal *TaCYP70* or other proteins containing CLDs, most likely due to the methods used to query the database, wherein such multidomain proteins are unlikely to be considered sufficiently similar to *wCyp1* to be revealed by tBLASTx.

6.7.6 Putative *cyclophilin B* genes are expressed in wheat

The identification of two TC sequences in the TaGI database that potentially encode ER-localised cyclophilin B's (Section 6.6.3) is also an interesting finding, as the presence of such cyclophilins in the wheat endosperm ER has been suggested to be potentially important for the folding of storage proteins (Shewry *et al.*, 1995; Shewry and Tatham, 1997). Like the *A. thaliana* CYP5 (Saito *et al.*, 1999), these putative cyclophilin B's lack recognizable ER-retention signals such as the frequently found 'KDEL' (Denecke *et al.*, 1992) or the 'VEKPFAIAKE' (Arber *et al.*, 1992), suggesting an alternative process for ER-retention of plant cyclophilin B's.

This chapter has provided useful information which can be used to further elucidate the role of cyclophilins in wheat. The characterisation of the *cyclophilin A* gene family will permit the design of gene-specific molecular markers for quantitative genetic analyses of these genes in agronomically important traits. Further, the identification of novel, putatively organelle-specific *cyclophilin* sequences will permit the isolation and characterisation of these genes in future studies to gain a further understanding of their role in the developing endosperm.

CHAPTER 7

CONCLUSIONS AND FUTURE DIRECTIONS

A major aim of this study was to isolate and characterise the gene families encoding PDI in *T. turgidum* subsp. *durum* and *Ae. tauschii* to determine their copy number and to identify genome-specific sequences to facilitate the isolation of the corresponding genes from *T. aestivum*. Upon achieving this aim, we were able to utilise the information to isolate and characterise the *PDI* genes of *T. aestivum* and assess these genes for inter-cultivar polymorphism to develop molecular markers for those genes where inter-cultivar polymorphism was identified and integrate them into existing genetic maps. Further, the knowledge gained on the structure of the *PDI* genes in wheat permitted an analysis of the rice orthologue of this gene, which appears to play an important role in storage protein deposition in that species, and allowed a preliminary investigation to determine the level of orthology between these loci. Another major aim was to characterise the gene family encoding other foldase enzymes in wheat, the cyclophilins, through analysis of the gene copy number and sequence and physical mapping of these genes. Further, cDNA library screening and data-mining of EST databases allowed the identification of several novel, putative organelle-bound members of this gene family. Taken together, the characterisation of these gene families has provided extensive insight into their potential roles in the developing endosperm and provided important data and directions for further research into these genes.

7.1 CONCLUSIONS

7.1.1 The *PDI* gene family of wheat

Investigations into the hexaploid bread wheat, *T. aestivum* (AABBDD) and the tetraploid progenitor of the A and B genomes, *T. turgidum* subsp. *durum* (AABB) and the diploid progenitor of the D genome, *Ae. tauschii* (DD) provided insights into the organisation of this gene family in wheat.

- Southern blot analysis of the *PDI* gene family in the hexaploid *T. aestivum* cv. Wyuna (AABBDD), *T. turgidum* subsp. *durum* (AABB) and *Ae. tauschii* (DD) revealed the presence of a single hybridising fragment per genome (Section 3.2).

- PCR amplification and cloning of the *PDI* genes from the gDNA of *T. turgidum* subsp. *durum* and *Ae. tauschii* supports the presence of a single *PDI* gene per genome, as clones of two distinct sequence types were isolated from the tetraploid, while all of the clones isolated from the diploid had the same sequence (Section 3.3).

- The *PDI* genes have a 10 exon / 9 intron structure, confirming and extending the previously reported partial sequences of the *PDI* genes *TtPDI4A* and *TtPDI4B* of *T. turgidum* subsp. *durum* and providing the first report of the sequence of the *PDI* gene, *AetPDI4D*, from the D genome of *Ae. tauschii* (Section 3.3).

- The inter-genome polymorphisms identified in the study of the above progenitors enabled the design of allele-specific (AS) primers for the targeted amplification and characterisation of sections of *PDI* genes from the individual genomes of *T. aestivum* revealing extremely high sequence conservation (>97% at the intron and exon level) between the *PDI* genes in hexaploid wheat (*TaPDI4A*, *TaPDI4B* and *TaPDI4D*) and those in the corresponding genomes of *T. turgidum* subsp. *durum* and *Ae. tauschii* (Section 3.4). Further, these genes are 100% identical to the corresponding sequences in the three cDNA clones previously isolated by us from an endosperm library (Section 1.5.5), revealing that all three *PDI* genes are expressed in this tissue.

- The report of a fourth (third in tetraploid wheat) gene on chromosome 1B of some cultivars of *T. aestivum* and *T. turgidum* subsp. *durum* (Section 1.5.5), could not be confirmed with certain probes in Southern blotting (Section 3.2) or in the cloning and sequencing of *PDI* genes in these species (Sections 3.3 and 3.4). However, the appearance of the reported band with a longer probe indicates this may be a partial gene with only it 5’ sequence present (Section 3.6.2). This conclusion is supported by the amplification of a PCR product from the 5’ section of the *PDI* gene and the absence of a product using primers specific for the 3’ section (Section 3.6.3).

- The *PDI* gene family of wheat thus appears to consist of a single expressed gene in each of the A, B and D genomes of wheat, each with a conserved intron/exon structure and conserved DNA sequence at the exon level (>95%) and the reported gene on chromosome 1B, if present, appears to be a partial gene consisting of only the 5' section of the gene.

- The sequencing of sections of *TaPDI4A*, *TaPDI4B* and *TaPDI4D* in the eight cultivars, previously used in the development of genetic maps (Section 1.8.4.1) revealed that:
    - the intron 5 of *TaPDI4A* contained a G/T SNP that was polymorphic between three of the four characterised crosses (Section 4.2.1);
    - the intron 9 of *TaPDI4B* contained an A/T SNP in that was polymorphic in CD87 x Katepwa (Section 4.2.2); and
    - that there was a complete absence of inter-cultivar polymorphism in *TaPDI4D* after sequencing of over 80% of this gene (Section 4.2.3).

- The design of molecular markers for *TaPDI4A* and *TaPDI4B* using certain RFLPs introduced by the SNPs described above (Sections 4.3 and 4.4), allowed linkage analysis and integration of these genes into genetic maps and revealed, that as expected, these genes are linked to markers on chromosome 4A and 4B, respectively (Section 4.6). Further, both genes were found to be closely linked to *germin* genes, known to be important in plant germination and pathogen defense. The integration of the *PDI* genes into the wheat genetic maps represents an important step in assessing the involvement of these genes in quantitative traits in these crosses.

- Preliminary evidence establishing orthology between the *PDI* loci of wheat and *esp2* locus of rice and thus a hypothetical QTL for storage protein deposition at the *PDI* loci of wheat was established as follows:

- The putative rice *PDI* orthologue of wheat *PDI* is located on chromosome 11S and encodes a protein 94% identical to that encoded by *TtPDI4A* (Section 5.2.1).
- This rice *PDI* gene and the *esp2* marker of rice are both present on the distal section of the chromosome 11S (Section 5.2.2) suggesting that the mutation in esp2 is directly to this *PDI* gene or to flanking regulatory elements.
- A 1Mb section of rice chromosome 11S, encompassing the *PDI* locus, contains sequences putatively orthologous to 76 ESTs or TC sequences in the TIGR TaGI database (Section 5.3.1), of which 5 had been physically mapped to the group 4 chromosomes in wheat often in the same 'bin' as the *PDI* genes of wheat, however some discrepancies were present (Section 5.3.2).
- Physical mapping of the *PDI* genes of wheat to clarify the above discrepancies confirmed and refined their locations to 'bins' on chromosomes 4AL, 4BS and 4DS (Section 5.4).
- Thus, the presence of genes flanking *PDI* genes of wheat *and* the *PDI* gene of rice, which plays a role in storage protein deposition, provides good preliminary evidence for orthology between these loci and thus, for the presence of a QTL for storage protein deposition at the *PDI* loci of wheat.

7.1.2 The *cyclophilin* gene family of wheat

Prior to this research, the structure and organisation of *cyclophilin* genes in wheat was unknown. The current research has provided insight into a number of members of this gene family in wheat, especially the *cyclophilin A* genes.

- The *cyclophilin A* genes of wheat, *TaCYP18-1*, *TaCYP18-2* and *TaCYP18-3*, appear to be intronless and are highly conserved in sequence between the three genomes, with no polymorphisms identified also between the eight cultivars investigated (Section 6.2).

- The *cyclophilin A* genes map to the distal sections of chromosomes 6AS, 6BS and 6DS using the N/T, ditelosomic and deletion stocks of common wheat, an area containing other genes that play a role in determining wheat quality, including the gliadins and polyphenol oxidase (Section 6.3).

- The *cyclophilin A* gene family of common wheat appears to consist of four members, and the fourth gene may be a duplicate of the *cyclophilin A* gene on chromosome 6AS, i.e. *TaCYP18-3* (Sections 6.4 and 6.5).

- The wheat genome also encodes novel, organelle-bound *cyclophilins*, including putative nuclear (*TaCYP-70*), plastid (*TaCYP20-1* and *TaCYP20-2*) and ER members (Section 6.6).

7.2 FUTURE DIRECTIONS

The findings of the current study have provided much of the groundwork to permit investigations into functional associations between the *PDI* genes and important quality traits of wheat. Further, the identification of novel cyclophilins, especially a putative ER-resident cyclophilin B, which may play an important role in the folding of the storage proteins, will permit investigations into the role(s) of these enzymes in the developing endosperm. The following research directions are suggested to gain further insight into these areas.

The development of molecular markers for *TaPDI4A* and *TaPDI4B* will allow their contribution to quantitative traits to be assessed. However, the low level of polymorphism in the D genome *PDI* gene, *TaPDI4D*, has thus far prevented the development of a marker for this gene. The results of the current research has provided some avenues to pursue in the development of such a marker, for example,

- The comparison of the *PDI* gene of *Ae. tauschii* with that of *T. aestivum* revealed a 34bp insertion/deletion in intron 1 (Section 3.4.3). This polymorphism may be exploited as a simple size difference in the PCR product, amplifying this intron to differentiate the two molecular alleles, i.e. with or

without the insertion/deletion. The mapping population generated from the synthetic hexaploid (Altar84/*Ae. tauschii*) x Opata M85 (Section 1.8.4.1) would likely exhibit this polymorphism, as the D genomes in this cross were donated by *Ae. tauschii* and *T. aestivum*.

- The sequence generated from *TaPDI4D* could be used to generate an AFLP marker by digesting gDNA with *Pst*I/*Mse*I, which produces more polymorphic fragments in wheat than other enzyme combinations (Langridge *et al.* 2001), and using an anchored primer based on the sequence of *TaPDI4D* coupled with an AFLP primer with few selective bases. Such a method could identify polymorphisms flanking the *TaPDI4D* gene and thus be genetically tightly linked.

One other clear area of future research would be QTL studies on dough rheological properties in the populations polymorphic for the *PDI* molecular markers. As PDI conceivably has an important role in protein body formation, and the mapping populations segregate for traits directly influenced by the storage proteins (Table 1.2), the mapping populations could be assessed for their elasticity, dough strength and/or extensibility. This data could be analysed to determine whether QTLs for such traits co-segregate with the *PDI* markers, providing genetic evidence for a role for the *PDI* genes in determining these traits.

The effect of absence of *PDI* expression in rice on protein body morphology in the endosperm raises the question whether such a mutation would have a similar effect on storage protein folding and protein body formation in wheat. The hexaploid nature of common wheat provides a high level of genetic redundancy, suggesting that all three *PDI* genes would need to be knocked out or any naturally occurring null-mutants would need to be identified, to observe the effect of such mutations on protein body morphology. However, analysis on the *waxy* gene in common wheat, which is responsible for amylose synthesis in starch, shows an unequal involvement of the three homeoalleles in determining amylose content (Yamamori *et al.* 2000). Thus, it will be interesting to identify (or develop by mutatgenesis) *PDI* single/double/triple-null mutants, to determine whether there are any subtle or quantitative effects of these on protein body formation.

The lack of inter-cultivar polymorphism in the *cyclophilin A* genes is most likely due to the functional conservation of the coding sequences in these intronless genes (Section 6.2.2). Thus, to identify polymorphisms in these genes for developing molecular markers, the non-coding flanking regions, which may not be under the same level of selection pressure that the coding sequences are under, need to be explored. As discussed above, the sequence data generated from the *cyclophilin* genes may be used generate a marker using an AFLP approach, with a *cyclophilin*-specific primer in conjunction with a traditional AFLP primer specific for the *Pst*I/*Mse*I restriction site.

Although the cyclophilins appear to be involved in a number of cellular activities (Section 1.6.4), the 'foldase' activity of these enzymes was the primary interest for the current research, especially in the context of protein body formation. However, the cytosolic cyclophilin A's do not seem to have access to the storage proteins during the formation of the protein bodies, as this process occurs in the ER and vacuoles. Thus, our identification of plastid- and ER-localised cyclophilins, which could conceivably play important roles in the folding of enzymes in the starch biosynthesis pathways and protein body formation, respectively, represents an important first step in elucidating the roles of these cyclophilins in the developing endosperm. Studies into the expression of these plastid and ER forms, to determine whether they are up-regulated during endosperm development, needs to be conducted. These studies could involve Northern blot analysis, however the close sequence homology between the *cyclophilin* classes would necessitate the analysis of their sequence to ensure that probes specific for each class did not cross-hybridise with the other classes of *cyclophilin* expressed in the endosperm. Alternatively, a quantitative real-time reverse transcriptase (RT-)PCR approach could be used, such as that used to study the expression of acetyl-coenzyme A carboxylase genes in wheat (Podkowinski *et al.* 2003), with primers designed that are specific to each class of *cyclophilin*.

Similar to the proposed studies of *PDI* null-mutants described above, an absence of *cyclophilin B* expression on the deposition of the storage proteins would be of great interest. As described above for the *PDI* genes, this analysis would involve the identification of any naturally occurring *cyclophilin B* null-mutants or mutagenesis to

engineer such mutants, followed by analysis of the effects of such mutations on the deposition of the storage proteins.

Overall, this project adds substantial new data to the field of wheat molecular genetics, particularly factors potentially affecting the folding of proteins in the wheat endosperm and thus, wheat grain quality.

REFERENCES

Abbas-Terki, T., Donze, O., Briand, P. A., and Picard, D. (2001). Hsp104 interacts with Hsp90 cochaperones in respiring yeast. Mol Cell Biol. **21**: 7569-7575.

Ahn, S. and Tanksley, S. D. (1993). Comparative linkage maps of the rice and maize genomes. Proc Natl Acad Sci USA. **90**: 7980-7984.

Anderson, O. D., and Blechl, A. E. (1998). Structure of the wheat storage protein genes. Proceedings of the 9th International Wheat Genetics Symposium (Saskatchewan, Canada, University Extension Press).

Anderson, S. K., Gallinger, S., Roder, J., Frey, J., Young, H. A. and Ortaldo, J. R. (1993). A cyclophilin-related protein involved in the function of natural killer cells. Proc Natl Acad Sci USA. **90**: 542-546.

Andreeva, L., Heads, R., and Green, C. J. (1999). Cyclophilins and their possible role in the stress response. Int J Exp Pathol. **80**: 305-315.

Antes, S., and Wieser, H. (2001). Effects of high and low molecular weight glutenin subunits on rheological dough properties and breadmaking quality of wheat. Cereal Chem. **78**: 157-159.

Appels, R., Gras, P., Clarke, B. C., Anderssen, R., Wesley, I., and Békés, F. (2001). Molecular genetic studies on processing traits of wheat flour. Euphytica. **119**: 49-54.

Arber, S., Krause, K. H. and Caroni, P. (1992). S-cyclophilin is retained intracellularly via a unique COOH-terminal sequence and colocalizes with the calcium storage protein calreticulin. J Cell Biol. **116**:113-125.

Arumuganathan, K., and Earle, E. (1991). Nuclear DNA content of some important plant species. Plant Mol Biol Rep. **9**: 208-218.

Badaeva, D. E., Freibe, B., and Gill, B. S. (1996). Genome differentiation in *Aegilops*. 1. Distribution of highly repetitive DNA sequences on chromosomes of diploid species. Genome. **39**: 293-306.

Bai, G. H., Kolb, F. L., Shaner, G., and Domier, L. L. (1999). Amplified fragment length polymorphism markers linked to a major quantitative trait locus controlling scab resistance in wheat. Phytopathology. **89**: 343-348.

Bailey, P. C., McKibbin, R. S., Lenton, J. R., Holdsworth, M. J., Flintham, J. E., and Gale, M. D. (1999). Genetic map locations for orthologous Vp1 genes in wheat and rice. Theor Appl Genet. **98**: 281-284.

Batey, I. L., Hayden, M. J., Cai, S., Sharp, P. J., Cornish, G. B., Morell, M. K. and Appels, R. (2001). Genetic mapping of commercially significant starch characteristics in wheat crosses. Aust J Agric Res. **52**: 1287-1296.

Beckwith, A. C., Wall, J. S., and Dimmler, R. J. (1963). Amide groups as interaction sites in wheat gluten proteins: effects of amide-ester conversion. Arch Biochem Biophys. **103**: 319-330.

Belton, P. S., Gil, A. M., and Tatham, A. S. (1994). 1H NMR relaxation times studies of the hydration of the barley protein C-hordein. J Chem Soc Farad Trans. **90**: 1099.

Belton, P. S., Colquhoun, I. J., Field, J., Grant, A., Shewry, P. R., Tatham, A. S., and Wellner, N. (1995). FTIR and NMR studies on the hydration of a high Mr subunit of glutenin. Int J Biol Macromol. **17**: 74.

Belton, P. S., Gil, A. M., Grant, A., Alberti, E., and Tatham, A. S. (1998). Proton and carbon NMR measurements of the effects of hydration on the wheat protein omega gliadin. Spectrochim Acta. **54**: 955-966.

Belton, P. S. (1999). On the elasticity of wheat gluten. J Cereal Sci. **29**: 103-107.

Bennetzen, J. L. and Ramakrishna, W. (2002). Numerous small rearrangements of gene content, order and orientation differentiate grass genomes. Plant Mol Biol. **48**: 821-827.

Bennetzen, J. L. and Ma, J. (2003). The genetic colinearity of rice and other cereals on the basis of genomic sequence analysis. Curr Opin Plant Biol. **6**: 128-133.

Bergsma, D. J., Eder, C., Gross, M., Kersten, H., Sylvester, D., Appelbaum, E., Cusimano, D., Livi, G. P., McLaughlin, M. M., and Kasyan, K. (1991). The cyclophilin multigene family of peptidyl-prolyl isomerases. Characterization of three separate human isoforms. J Biol Chem. **266**: 23204-23214.

Berna, A. and Bernier, F. (1999). Regulation by biotic and abiotic stress of a wheat germin gene encoding oxalate oxidase, a H2O2-producing enzyme. Plant Mol Biol. **39**: 539-549.

Bietz, J. A., Shepherd, K. A., and Wall, J. S. (1975). Single-kernel analysis of glutenin: use in wheat genetics and breeding. Cereal Chem. **52**: 513-532.

Blake, N. K., Lehfeldt, B. R., Hemphill, A., Shan, X., and Talbert, L. E. (1998). DNA sequence analysis suggests a monophyletic origin of the wheat B genome. Proceedings of the 9th International Wheat Genetics Symposium (Saskatchewan, Canada, University Extension Press).

Blake, N. K., Lehfeldt, B. R., Lavin, M., and Talbert, L. E. (1999). Phylogenetic reconstruction based on low copy DNA sequence data in an allopolyploid: The B genome of wheat. Genome. **42**: 351-360.

Bonierbale, M. W., Plaisted, R. L. and Tanksley, S. D. (1988). RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics. **120**: 1095-1103.

Bossard, M. J., Koser, P. L., Brandt, M., Bergsma, D. J. and Levy, M. A. (1991). A single Trp-121 to Ala-121 mutation in human cyclophilin alters cyclosporin A affinity and peptidyl-prolyl isomerase activity. Biochem Biophys Res Commun. **176**: 1142-1148.

Botstein, D., White, R. L., Skolnick, M. and Davis, R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet. **32**:314-331.

Boyko, E. V., Gill, K. S., Mickelson-Young, L., Nasuda, S., Raupp, W. J., Ziegle, J. N., Singh, S., Hassawi, D. S., Fritz, A. K., Namuth, D., Lapitan, N. L. V. and Gill, B. S.

(1999). A high-density genetic linkage map of *Aegilops tauschii*, the D-genome progenitor of bread wheat. Theor Appl Genet. **99**:16–26.

Breiman, A., Fawcett, T. W., Ghirardi, M. L., and Mattoo, A. K. (1992). Plant organelles contain distinct peptidylprolyl cis,trans-isomerases. J Biol Chem. **267**: 21293-21296.

Breiman, A., and Graur, D. (1995). Wheat evolution. Israel J of Plant Sci. **43**: 85-98.

Bryan, G. J., Stephenson, P., Collins, A., Kirby, J., Smith, J. B., and Gale, M. D. (1999). Low levels of DNA sequence variation among adapted genotypes of hexaploid wheat. Theor Appl Genet. **99**: 192-198.

Buchholz, W. G., Harrishaller, L., Derose, R. T., and Hall, T. C. (1994). Cyclophilins are encoded by a small gene family in rice. Plant Molec Biol. **25**: 837-843.

Bulleid, N. J., and Freedman, R. B. (1988). Defective co-translational formation of disulphide bonds in protein disulphide-isomerase-deficient microsomes. Nature. **335**: 649-651.

Buonocore, F., Bertini, L., Ronchi, C., Békés, F., Caporale, C., Lafiandra, D., Gras, P., Tatham, A. S., Greenfield, J. A., Halford, N. G., and Shewry, P. R. (1998). Expression and functional analysis of Mr 58000 peptides derived from the repetitive domain of high molecular weight subunit 1Dx5. J Cereal Sci. **27**: 209-215.

Cai, H., Wang, C. C., and Tsou, C. L. (1994). Chaperone-like activity of protein disulfide isomerase in the refolding of a protein with no disulfide bonds. J Biol Chem. **269**: 24550-24552.

Çakir, M., Appels, R., Carter, M., Loughman, R., Francki, M., Li, C., Johnson, J., Bhave, M., Wilson, R., McLean, R. and Barclay, I. (2003). Accelerated wheat breeding using molecular markers. Proceedings of the 10th International Wheat Genetics Symposium (Paestum, Italy, University Extension Press).

Caliskan, M. and Cuming, A.C. (1998). Spatial specificity of H_2O_2-generating oxalate oxidase gene expression during wheat embryo germination. Plant J. **15**: 165-171.

Callebaut I. and Mornon J. P. (1995). Trigger factor, one of the *Escherichia coli* chaperone proteins, is an original member of the FKBP family. FEBS Lett. **374**:211-215.

Campbell, K. G., Bergman, C. J., Gualberto, D. G., Anderson, J. A., Giroux, M. J., Hareland, G., Fulcher, R. G., Sorrells, M. E., and Finney, P. L. (1999). Quantitative trait loci associated with kernel traits in a soft x hard wheat cross. Crop Sci. **39**: 1184-1195.

Causse, M. A., Fulton, T. M., Cho, Y. G., Ahn, S. N., Chunwongse, J., Wu, K., Xiao, J., Yu, Z., Ronald, P. C., Harrington, S. E., Second, G., McCouch, S. R. and Tanksley, S. D. (1994). Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics. **138**: 1251-1274.

Chalmers, K. J., Campbell, A. W., Kretschmer, J., Karakousis, A., Henschke, P. H., Pierens, S., Harker, N., Pallotta, M., Cornish, G. B., Shariflou, M. R., *et al.* (2001). Construction of three linkage maps in bread wheat, *Triticum aestivum*. Aust J Agric Res. **52**: 1089-1119.

Chatrenet, B., and Chang, J. Y. (1992). The folding of hirudin adopts a mechanism of trial and error. J Biol Chem. **267**: 3038-3043.

Chen, F. Q., and Hayes, P. M. (1994). Nucleotide sequence and developmental expression of duplicated genes encoding protein disulfide isomerase in barley (*Hordeum vulgare* L.). Plant Physiol. **106**: 1705-1706.

Chen, M., SanMiguel, P., de Oliveira, A. C., Woo, S.-S., Zhang, H., Wing, R. A., and Bennetzen, J. L. (1997). Microcolinearity in sh2-homologous regions of the maize, rice, and sorghum genomes. Proc Natl Acad Sci USA. **94**: 3431-3435.

Chou, I. T. and Gasser, C. S., (1997). Characterisation of the cyclophilin gene family of *Arabidopsis thaliana* and phylogenetic analysis of known cyclophilin proteins. Plant Mol Biol. **35**: 873-892.

Ciaffi, M., Dominici, L., Tanzarella, O. A., and Porceddu, E. (1999). Chromosomal assignment of gene sequences coding for protein disulphide isomerase (PDI) in wheat. Theor Appl Genet. **98**: 405-410.

Ciaffi, M., Dominici, L., Umana, E., Tanzarella, O. A., and Porceddu, E. (2000). Restriction fragment length polymorphism (RFLP) for protein disulfide isomerase (PDI) gene sequences in *Triticum* and *Aegilops* species. Theor Appl Genet. **101**: 220-226.

Ciaffi, M., Paolacci, A. R., Dominici, L., Tanzarella, O. A., Porceddu, E. (2001). Molecular characterization of gene sequences coding for protein disulfide isomerase (PDI) in durum wheat (*Triticum turgidum* ssp. durum). Gene. **265**: 147-56.

Clarke, B. C., Hobbs, M., Skylas, D. and Appels, R. (2000). Genes active in developing wheat endosperm. Funct Integr Genomics. **1**: 44 - 55.

Clarke, B. C., Larroque, O. R., Bekes, F., Somers, D., and Appels, R. (2001). The frequent classes of expressed genes in wheat endosperm tissue as possible sources of genetic markers. Aust J Agric Res. **52**: 1181-1193.

Colley, N. J., Baker, E. K., Stamnes, M. A., and Zuker, C. S. (1991). The cyclophilin homolog ninaA is required in the secretory pathway. Cell. **67**: 255-263.

Coughlan, S. J., Hastings, C., and Winfrey Jr, R. J. (1996). Molecular characterisation of plant endoplasmic reticulum. Identification of protein disulfide-isomerase as the major reticuloplasmin. Eur J Biochem. **235**: 215-224.

Creighton, T. E. (1979). Intermediates in the refolding of reduced ribonuclease A. J Mol Biol. **129**: 411-431.

Creighton, T. E., and Goldenberg, D. P. (1984). Kinetic role of a meta-stable native-like two-disulphide species in the folding transition of bovine pancreatic trypsin inhibitor. J Mol Biol. **179**: 497-526.

Crompton, M. (1999). The mitochondrial permeability transition pore and its role in cell death. Biochem J. **341**: 233-249.

Dai, Y., and Wang, C. (1997). A mutant truncated protein disulfide isomerase with no chaperone activity. J Biol Chem. **272**: 27572-27576.

Danielson, P. E., Forss-Petter, S., Brow, M. A., Calavetta, L., Douglass, J., Milner, R. J. and Sutcliffe, J. G. p1B15: a cDNA clone of the rat mRNA encoding cyclophilin. DNA. **7**:261–267.

Daud, H. M., and Gustafson, J. P. (1996). Molecular evidence for *Triticum speltoides* as a B-genome progenitor of wheat (*Triticum aestivum*). Genome. **39**: 543-548.

Denecke, J., De Rycke, R., and Botterman, J. (1992). Plant and mammalian sorting signals for protein retention in the endoplasmic reticulum contain a conserved epitope. EMBO J. **11**: 2345-2355.

Derkx, P. M. and Madrid, S. M. (2001). The foldase CYPB is a component of the secretory pathway of *Aspergillus niger* and contains the endoplasmic reticulum retention signal HEEL. Mol Genet Genomics. **266**:537-545.

Desilva, M. G., Notkins, A. L., and Lan, M. S. (1997). Molecular characterization of a pancreas-specific protein disulfide isomerase, PDIp. DNA Cell Biol. **16**: 269-274.

D'Ovidio, R., Simeone, M., Masci, S., and Porceddu, E. (1997). Molecular characterization of a LMW-GS gene located on chromosome 1B and the development of primers specific for the Glu-B3 complex locus in durum wheat. Theor Appl Genet. **95**: 1119-1126.

Duina, A. A., Chang, H. C., Marsh, J. A., Lindquist, S., and Gaber, R. F. (1996). A cyclophilin function in Hsp90-dependent signal transduction. Science. **274**: 1713-1715.

DuPont, F. M., Hurkman, W. J., Tanaka, C. K., and Chan, R. (1998). BiP, HSP70, NDK and PDI in wheat endosperm. I. Accumulation of mRNA and protein during grain development. Physiol Plant. **103**: 70-79.

Dvorak, J., McGuire, P. E., and Cassidy, B. (1988). Apparent sources of the A genomes of wheats inferred from polymorphism in abundance and restriction fragment length of repeated nucleotide sequences. Genome. **30**: 680-689.

Dvorak, J., di Terlizzi, P., Zhang, H.-B., and Resta, P. (1993). The evolution of polyploid wheat: identification of the A genome donor species. Genome. **36**: 21-31.

Dvorak, J., Luo, M. C., Yang, Z. L., and Zhang, H. B. (1998). The structure of the *Aegilops tauschii* genepool and the evolution of hexaploid wheat. Theor Appl Genet. **97**: 657-670.

Eckermann, P. J., Verbyla, A. P., Cullis, B. R., and Thompson, R. (2001). The analysis of quantitative traits in wheat mapping populations. Aust J Agric Res. **52**: 1195-1206.

Edvardsson, A., Eshaghi, S., Vener, A. V., and Andersson, B. (2003). The major peptidyl-prolyl isomerase activity in thylakoid lumen of plant chloroplasts belongs to a novel cyclophilin TLP20. FEBS Lett. **542**: 137-141.

Effertz, R. J., Anderson, J. A. and Francl, L. J. (1998). QTLs associated with resistance to chlorosis induction by *Pyrenophora tritici-repentis* in adult wheat. Can J Plant Pathol. **20**: 438-439.

Elliott, J. G., Oliver, J. D., and High, S. (1997). The thiol-dependent reductase ERp57 interacts specifically with N-glycosylated integral membrane proteins. J Biol Chem. **272**: 13849-13855.

Emanuelsson, O., Nielsen, H., Brunak, S. and von Heijne, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol. **300**: 1005-1016.

Endo, T. R., and Gill, B. S. (1996). The deletion stocks of common wheat. J Heredity. **87**: 295-307.

Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C. J., Hofmann, K. and Bairoch, A. (2002). The PROSITE database, its status in 2002. Nucleic Acids Res. **30**: 235-238.

Feinberg, A.P. and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Biochem. **132**: 6-13.

Feng, W., Bedows, E., Norton, S. E., and Ruddon, R. W. (1996). Novel covalent chaperone complexes associated with human chorionic gonadotropin beta subunit folding intermediates. J Biol Chem. **271**: 18543-18548.

Ferrari, D. M., and Söling, H. D. (1999). The protein disulphide-isomerase family: unravelling a string of folds. Biochem J. **339**: 1-10.

Ferreira, P. A., Hom, J. T. and Pak, W. L. (1995). Retina-specifically expressed novel subtypes of bovine cyclophilin. J Biol Chem. **270**: 23179-23188.

Fido, R. J., Békés, F., Gras, P. W., and Tatham, A. S. (1997). Effects of alpha-, beta-, gamma- and omega-gliadins on the dough mixing properties of wheat flour. J Cereal Sci. **26**: 271-277.

Field, J., Shewry, P. R., and Miflin, B. (1983). Solubilization and characterization of wheat gluten proteins; correlations between the amount of aggregated proteins and baking quality. J Sci Food Agric. **34**: 370-377.

Fischer, G., and Bang, H. (1985). The refolding of urea-denatured ribonuclease A is catalyzed by peptidyl-prolyl cis-trans isomerase. Biochim Biophys Acta. **828**: 39-42.

Fischer, G., Bang, H., and Mech, C. (1984). Determination of enzymatic catalysis for the cis-trans-isomerization of peptide binding in proline-containing peptides. Biomed Biochim Acta. **43**: 1101-1111.

Freedman, R. B., Klappa, P., and Ruddock, L. W. (2002). Protein disulfide isomerases exploit synergy between catalytic and specific binding domains. EMBO Rep. **3**: 136-140.

Freskgard, P. O., Bergenhem, N., Jonsson, B. H., Svensson, M., and Carlsson, U. (1992). Isomerase and chaperone activity of prolyl isomerase in the folding of carbonic anhydrase. Science. **258**: 466-468.

Friedman, J. and Weissman, I. (1991). Two cytoplasmic candidates for immunophilin action are revealed by affinity for a new cyclophilin: one in the presence and one in the absence of CsA. Cell. **66**: 799-806.

Friedman, J., Trahey, M., and Weissman, I. (1993). Cloning and characterization of cyclophilin C-associated protein: A candidate natural cellular ligand for cyclophilin C. Proc Natl Acad Sci USA. **90**: 6815-6819.

Frigerio, G. and Pelham, H. R. B. (1993). A *Saccharomyces cerevisiae* cyclophilin resident in the endoplasmic reticulum. J Mol Biol. **233**:183-188.

Fulgosi, H., Vener, V., Altschmied, L., Herrmann, R. G., and Andersson, B. (1998). A novel multi-functional chloroplast protein: identification of a 40 kDa immunophilin-like protein located in the thylakoid lumen. EMBO J. **17**: 1577-1587.

Galat, A., and Rivière, S. (1998). Peptidyl-prolyl *cis/trans* isomerases (Oxford, Oxford University Press).

Galat, A. (1999). Variations of sequences and amino acid compositions of proteins that sustain their biological functions: an analysis of the cyclophilin family of proteins. Arch Biochem Biophys. **371**: 149-162.

Galat, A. (2003). Peptidylprolyl cis/trans isomerases (immunophilins): biological diversity - targets - functions. Curr Top Med Chem. **3**: 1315-1347.

Galat, A. (2004). Function-dependent clustering of orthologues and paralogues of cyclophilins. Proteins. **56**: 808-820.

Gale, M. D., and Devos, K. M. (1998). Comparative genetics in the grasses. Proc Natl Acad Sci USA. **95**: 1971-1974.

Gasser, C. S., Gunning, D. A., Budelier, K. A., and Brown, S. M. (1990). Structure and expression cytosolic cyclophilin/peptidyl-prolyl cis-trans isomerase of higher plants and production of active tomato cyclophilin in *Escherichia coli*. Proc Natl Acad Sci USA. **87**: 9519-9523.

Gilbert, H. F. (1997). Protein disulfide isomerase and assisted protein folding. J Biol Chem. **272**: 29399-29402.

Gilbert, S. M., Wellner, N., Belton, P. S., Greenfield, J. A., Siligardi, G., Shewry, P. R., and Tatham, A. S. (2000). Expression and characterisation of a highly repetitive peptide derived from a wheat seed storage protein. Biochim Biophys Acta. **1479**: 135-146.

Gill, B. S., and Kimber, G. (1974). Giesma C-banding and the evolution of wheat. Proc Natl Acad Sci USA. **71**: 4086-4090.

Gill, B. S., and Appels, R. (1988). Relationships between *Nor*-loci from different *Triticeae* species. Plant Syst Evol. **160**: 77-99.

Gill, K. S., Gill, B. S., Endo, T. R., and Boyko, E. V. (1996). Identification and high-density mapping of gene-rich regions in chromosome group 5 of wheat. Genetics. **143**: 1001-1012.

Godoy, A. V., Lazzaro, A. S., Casalongué, C. A., and San Segundo, B. (2000). Expression of a *Solanum tuberosum* cyclophilin gene is regulated by fungal infection and abiotic stress conditions. Plant Sci. **152**: 123-134.

Goff, S. A., Ricke, D., Lan, T. H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., *et al.* (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). Science. **296**: 92-100.

Goldberger, R. F., Epstein, C. J., and Anfinsen, C. B. (1963). Acceleration of reactivation of reduced bovine pancreatic ribonuclease by a microsomal system from rat liver. J Biol Chem. **238**: 628-635.

Göthel, S. F., and Marahiel, M. A. (1999). Peptidyl-prolyl cis-trans isomerases, a superfamily of ubiquitous folding catalysts. Cell Mol Life Sci. **55**: 423-436.

Graveland, A., Bosveld, P., Lichtendonk, W. J., Marseille, J. P., Moonen, J. H. E., and Scheepstra, A. (1985). A model for the molecular structure of the glutenins from wheat flour. J Cereal Sci. **3**: 1-16.

Grimwade, B., Tatham, A. S., Freedman, R. B., Shewry, P. R., and Napier, J. A. (1996). Comparison of the expression patterns of genes coding for wheat gluten proteins and proteins involved in the secretory pathway in developing caryopses of wheat. Plant Molec Biol. **30**: 1067-1073.

Haendler, B., Hofer-Warbinek, R. and Hofer, E. (1987). Complementary DNA for human T-cell cyclophilin. EMBO J. **6**:947–950.

Haendler, B. and Hofer, E. (1990). Characterization of the human cyclophilin gene and of related processed pseudogenes. Eur J Biochem. **190**: 477-482.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser. **41**: 95-98.

Handschumacher, R. E., Harding, M. W., Rice, J., and Drugge, R. J. (1984). Cyclophilin: a specific cytosolic binding protein for cyclosporin A. Science. **226**: 544-547.

Harding, M. W., Handschumacher, R. E. and Speicher, D. W. (1986). Isolation and amino acid sequence of cyclophilin. J Biol Chem. **261**:8547–8555.

Harding, M. W., Galat, A., Uehling, D. E., and Schreiber, S. L. (1989). A receptor for the immunosuppressant FK506 is a *cis-trans* peptidyl-prolyl isomerase. Nature. **341**: 758-760.

Harker, N., Rampling, L. R., Shariflou, M. R., Hayden, M. J., Holton, T. A., Morell, M. K., Sharp, P. J., Henry, R. J. and Edwards, K. J. (2001). Microsatellites as markers for Australian wheat improvement. Aust J Agric Res. **52**: 1121-1130.

Hasel, K. W. and Sutcliffe, J. G. (1990). Nucleotide sequence of a cDNA coding for mouse cyclophilin. Nucleic Acids Res. **18**:4019.

Hayano, T., Hirose, M., and Kikuchi, M. (1995). Protein disulfide isomerase mutant lacking its isomerase activity accelerates protein folding in the cell. FEBS Lett. **377**: 505-511.

Hayano, T., and Kikuchi, M. (1995). Molecular cloning of the cDNA encoding a novel protein disulfide isomerase-related protein (PDIR). FEBS Lett. **372**: 210-214.

Hayman, G. T., and Miernyk, J. A. (1994). The nucleotide and deduced amino acid sequences of a peptidyl-prolyl cis-trans isomerase from Arabidopsis thaliana. Biochim Biophys Acta. **1219**: 536-538.

Hennig, L., Christner, C., Kipping, M., Schelbert, B., Rucknagel, K. P., Grabley, S., Kullertz, G., and Fischer, G. (1998). Selective inactivation of parvulin-like peptidyl-prolyl cis/trans isomerases by juglone. Biochemistry. **37**: 5953-5960.

Herzberg, O., and Moult, J. (1991). Analysis of the steric strain in the polypeptide backbone of protein molecules. Proteins. **11**: 223-229.

High, S., Lecomte, F. J., Russell, S. J., Abell, B. M., and Oliver, J. D. (2000). Glycoprotein folding in the endoplasmic reticulum: a tale of three chaperones? FEBS Lett. **476**: 38-41.

Hohmann, U., Endo, T. R., Herrmann, R. G. and Gill, B. S. (1994). Characterization of deletions in common wheat induced by an *Aegilops cylindrica* chromosome: detection of multiple chromosome re-arrangements. Theor Appl Genet. **91**: 611-617.

Horibe, T., Yosho, C., Okada, S., Tsukamoto, M., Nagai, H., Hagiwara, Y., Tujimoto, Y. and Kikuchi, M. (2002). The chaperone activity of protein disulfide isomerase is affected by cyclophilin B and cyclosporin A in vitro. J Biochem. **132**: 401-407.

Hsia, C. C., and Anderson, O. D. (2001). Isolation and characterization of wheat omega-gliadin genes. Theor Appl Genet. **103**: 37-44.

Huang, S., Sirikhachornkit, A., Su, X. J., Faris, J., Gill, B., Haselkorn, R., and Gornicki, P. (2002). Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum*/*Aegilops* complex and the evolutionary history of polyploid wheat. Proc Natl Acad Sci USA. **99**: 8133-8138.

Iwai, N. and Inagami, T. (1990). Molecular cloning of a complementary DNA to rat cyclophilin-like protein mRNA. Kidney Int. **37**:1460–1465.

Jaaska, V. (1980). Electrophoretic survey of seedling esterases in wheats in relation to their phylogeny. Theor Appl Genet. **56**: 273-284.

Jaaska, V. (1997). Isoenzyme differences between the wild diploid and tetraploid wheats. Genet Resources and Crop Evol. **44**: 137-146.

Jackson, E. A., Holt, L. M., and Payne, P. E. (1983). Characterisation of high molecular weight gliadin and low molecular-weight glutenin subunits of wheat endosperm by two dimensional electrophoresis and the chromosomal location of their controlling genes. Theor Appl Genet. **66**: 29-37.

Jefferies, S. P., Pallotta, M. A., Paull, J. G., Karakousis, A., Kretschmer, J. M., Manning, S., Islam, A., Langridge, P., and Chalmers, K. J. (2000). Mapping and validation of chromosome regions conferring boron toxicity tolerance in wheat (*Triticum aestivum*). Theor Appl Genet. **101**: 767-777.

Johnson, J. C., Clarke, B. C., and Bhave, M. (2001). Isolation and characterisation of cDNAs encoding protein disulphide isomerases and cyclophilins in wheat. J Cereal Sci. **34**: 159-171.

Jones, B. L., and Mak, A. S. (1983). The amino-acid sequence of a purothionin from *Triticum monococcum*, a diploid wheat. Eur J Biochem. **133**: 621-625.

Joppa, L. R., Du, C. H., Hart, G. E. and Hareland, G. A., (1997). Mapping gene(s) for grain protein in tetraploid wheat (*Triticum turgidum* l.) using a population of recombinant inbred chromosome lines. Crop Sci. **37**: 1586-1589.

Kammholz, S. J., Campbell, A. W., Sutherland, M. W., Hollamby, G. J., Martin, P. J., Eastwood, R. F., Barclay, I., Wilson, R. E., Brennan, P. S., and Sheppard, J. A. (2001). Establishment and characterisation of wheat genetic mapping populations. Aust J Agric Res. **52**: 1079-1088.

Keck, B., Köhler, P., and Wieser, H. (1995). Disulphide bonds in wheat gluten: cystine peptides derived from gluten proteins following peptic and thermolytic digestion. Z Lebensm Unters Forsch. **200**: 432-439.

Keller, B., and Feuillet, C. (2000). Colinearity and gene density in grass genomes. Trends Plant Sci. **5**: 246-251.

Keller, M., Keller, B., Schachermayr, G., Winzeler, M., Schmid, J. E., Stamp, P., and Messmer, M. M. (1999a). Quantitative trait loci for resistance against powdery mildew in a segregating wheat x spelt population. Theor Appl Genet. **98**: 903-912.

Keller, M., Karutz, C., Schmid, J. E., Stamp, P., Winzeler, M., Keller, B., Messmer, M. M. (1999b). Quantitative trait loci for lodging resistance in a segregating wheat x spelt population. Theor Appl Genet. **98**: 1171-1182.

Kemmink, J., Dijkstra, K., Mariani, M., Scheek, R. M., Penka, E., Nilges, M., and Darby, N. J. (1999). The structure in solution of the b domain of protein disulfide isomerase. J Biomol NMR. **13**: 357-368.

Kerby, K., Kuspira, J., Jones, B. L., and Lockhart, G. L. (1990). Biochemical data bearing on the origin of the B genome in the polyploid wheats. Genome. **33**: 360-368.

Kieffer, L. J., Seng, T. W., Li, W., Osterman, D. G., Handschumacher, R. E. and Bayney, R. M. (1993). Cyclophilin-40, a protein with homology to the p59 component of the steroid receptor complex. J Biol Chem. **268**: 12303-12310.

Kilian, A., Chen, J., Han, F., Steffenson, B., and Kleinhofs, A. (1997). Towards map-based cloning of the barley stem rust resistance genes Rpg1 and rpg4 using rice as an intergenomic cloning vehicle. Plant Mol Biol. **35**: 187-195.

Klappa, P., Ruddock, L. W., Darby, N. J., and Freedman, R. B. (1998a). The b' domain provides the principal peptide-binding site of protein disulfide isomerase but all domains contribute to binding of misfolded proteins. EMBO J. **17**: 927-935.

Klappa, P., Stromer, T., Zimmermann, R., Ruddock, L. W., and Freedman, R. B. (1998b). A pancreas-specific glycosylated protein disulphide-isomerase binds to misfolded proteins and peptides with an interaction inhibited by oestrogens. Eur J Biochem. **254**: 63-69.

Köhler, P., Belitz, H. D., and Wieser, H. (1993). Disulphide bonds in wheat gluten: further cystine peptides from high molecular weight (HMW) and low molecular weight (LMW) subunits of glutenin and from gamma-gliadins. Z Lebensm Unters Forsch. **196**: 239-247.

Koivunen, P., Helaakoski, T., Annunen, P., Veijola, J., Raisanen, S., Pihlajaniemi, T., and Kivirikko, K. I. (1996). ERp60 does not substitute for protein disulphide isomerase as the beta-subunit of prolyl 4-hydroxylase. Biochem J. **316**: 599-605.

Kojima, T., Nagaoka, T., Noda, K. and Ogihara, Y. (1998). Genetic linkage map of ISSR and RAPD markers in Einkorn wheat in relation to that of RFLP markers. Theor Appl Genet. **96**: 37-45.

Konishi, Y., Ooi, T., and Scheraga, H. A. (1982). Regeneration of ribonuclease A from the reduced protein. Rate-limiting steps. Biochemistry. **21**: 4734-4740.

Kreis, M., Forde, B. G., Rahman, S., Miflin, B., and Shewry, P. R. (1985). Molecular evolution of the seed storage proteins of barley, rye and wheat. J Mol Biol. **183**: 499-502.

Kreis, M., and Shewry, P. R. (1989). Unusual features of seed protein structure and evolution. BioEssays. **10**: 201-207.

Kurata, N., Moore, G., Nagamura, Y., Foote, T., Yano, M., Minobe, Y. and Gale, M. D. (1994). Conservation of genome structure between rice and wheat. Biotechnology (N.Y.). **12**: 276-278.

Laboissiere, M. C., Sturley, S. L., and Raines, R. T. (1995). The essential function of protein-disulfide isomerase is to unscramble non-native disulfide bonds. J Biol Chem. **270**: 28006-28009.

LaMantia, M., Miura, T., Tachikawa, H., Kaplan, H. A., Lennarz, W. J., and Mizunga, T. (1991). Glycosylation site binding protein and protein disulfide isomerase are identical and essential for cell viability in yeast. Proc Natl Acad Sci USA. **88**: 4453-4457.

Lane, B. G., Dunwell, J. M., Ray, J. A., Schmitt, M. R., and Cuming, A. C. (1993). Germin, a protein marker of early plant development, is an oxalate oxidase. J Biol Chem. **268**: 12239-12242.

Lane BG. (2002). Oxalate, germins, and higher-plant pathogens. IUBMB Life. **53**: 67-75.

Lang, K., Schmid, F. X., and Fischer, G. (1987). Catalysis of protein folding by prolyl isomerase. Nature. **329**: 268-270.

Langridge, P., Lagudah, E. S., Holton, T. A., Appels, R., Sharp, P. J., and Chalmers, K. J. (2001). Trends in genetic and genome analyses in wheat: a review. Aust J Agric Res. **52**: 1043-1077.

LaRota, M. and Sorrells, M. E. (2004). Comparative DNA sequence analysis of mapped wheat ESTs reveals the complexity of genome relationships between rice and wheat. Funct Integr Genomics. **4**: 34–46.

Lawrence, G. J., and Payne, P. I. (1983). Detection by gel electrophoresis of oligomers formed by the association of high-molecular-weight glutenin protein subunits of wheat endosperm. J Exp Bot. **34**: 256-267.

Lebeau, M. C., Massol, N., Herrick, J., Faber, L. E., Renoir, J. M., Radanyi, C., and Baulieu, E. E. (1992). P59, an hsp 90-binding protein. Cloning and sequencing of its cDNA and preparation of a peptide-directed polyclonal antibody. J Biol Chem. **267**: 4281-4284.

Levanony, H., Rubin, R., Altschuler, Y., and Galili, G. (1992). Evidence for a novel route of wheat storage proteins to vacuoles. J Cell Biol. **119**: 1117-1128.

Li, C. P., and Larkins, B. A. (1996). Expression of protein disulfide isomerase is elevated in the endosperm of the maize floury-2 mutant. Plant Molec Biol. **30**: 873-882.

Li, W. L., Faris, J. D., Chittoor, J. M., Leach, J. E., Hulbert, S., Liu, D. J., Chen, P. D. and Gill, B.S. (1999). Genomic mapping of defense response genes in wheat. Theor Appl Genet. **98**: 226-233.

Lilie, H., McLaughlin, S., Freedman, R., and Buchner, J. (1994). Influence of protein disulfide isomerase (PDI) on antibody folding in vitro. J Biol Chem. **269**: 14290-14296.

Lilienfeld, F. A., and Kihara, H. (1934). Genome analysis of *Triticum* and *Aegilops.* V. *Triticum timopheevii* Zhuk. Cytologia. **6**: 87-122.

Lippuner, V., Chou, I. T., Varian Scott, S., Ettinger, W. F., Theg, S. M., and Gasser, C. S. (1994). Cloning and characterization of chloroplast and cytosolic forms of cyclophilin from *Arabidopsis thaliana*. J Biol Chem. **269**: 7863-7868.

Liu, B. H. (1998). Statistical genomics: linkage, mapping and QTL analysis. CRC Press, New York.

Liu, J., Farmer, J. D., Jr., Lane, W. S., Friedman, J., Weissman, I., and Schreiber, S. L. (1991). Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP-FK506 complexes. Cell. **66**: 807-815.

Liu, R., Olsen O. A., Kreis M. and Halford N.G. (1992). Molecular cloning of a novel barley seed protein gene that is repressed by abscisic acid. Plant Mol Biol. **18**: 1195-1198.

Lodish, H. F., and Kong, N. (1991). Cyclosporin A inhibits an initial step in folding of transferrin within the endoplasmic reticulum. J Biol Chem. **266**: 14835-14838.

Luan, S., Li, W., Rusnak, F., Assmann, S. M., and Schreiber, S. L. (1993). Immunosuppressants implicate protein phosphatase regulation of K+ channels in guard cells. Proc Natl Acad Sci USA. **90**: 2202-2206.

Luan, S., Lane, W. S., and Schrieber, S. L. (1994). pCyP B: a chloroplast-localized, heat-shock-responsive cyclophilin from fava bean. Plant Cell. **6**: 885-892.

Lubbers, E. L., Gill, K. S., Cox, T. S., and Gill, B. S. (1991). Variation of molecular markers among geographically diverse accessions of *Triticum tauschii*. Genome. **34**: 354-361.

Lyles, M. M., and Gilbert, H. F. (1991). Catalysis of the oxidative folding of ribonuclease A by protein disulfide isomerase: pre-steady state kinetics and the utilisation of the oxidising equivalents of the isomerase. Biochemistry. **30**: 619-625.

Manly, K. F. and Olson, J. M. (1999). Overview of QTL mapping software and introduction to Map Manager QT, Mamm Genome. **10**: 327-334.

Mares, D. J. and Campbell, A. W. (2001). Mapping components of flour and noodle colour in Australian wheat. Aust J Agric Res. **52**: 1297-1309.

Mares, D. J. and Mrva, K. (2001). Mapping quantitative trait loci associated with variation in grain dormancy in Australian wheat. Aust J Agric Res. **52**: 1257-1265.

Mares, M., Mrva, K., Tan, M. and Sharp, P. (2002). Dormancy in white-grained wheat: progress towards identification of genes and molecular markers. Euphytica. **126**: 47-53.

Marino, C. L., Nelson, J. C., Lu, Y. H., Sorrells, M. E., Leroy, P., Tuleen, N. A., Lopes, C. R. and Hart, G. E. (1996). Molecular genetic maps of the group 6 chromosomes of hexaploid wheat (*Triticum aestivum* L. em. Thell). Genome. **39**: 359-366.

Marivet, J., Frendo, P., and Burkard, G. (1992). Effects of abiotic stresses on cyclophilin gene expression in maize and bean and sequence analysis of bean cyclophilin cDNA. Plant Sci. **84**: 171-178.

Marivet, J., Margis-Pinheiro, M., Frendo, P., and Burkard, G. (1994). Bean cyclophilin gene expression during plant development and stress conditions. Plant Mol Biol. **26**: 1181-1189.

Markert, C. L. and Moller, F. (1959). Multiple forms of enzymes: tissue, ontogenetic, and species-specific patterns. Proc Natl Acad Sci USA. **45**: 753-763.

Marty, I., Brugidou, C., Chartier, Y., and Meyer, Y. (1993). Growth-related gene expression in Nicotiana tabacum mesophyll protoplasts. Plant J. **4**: 265-278.

Masci, S. M., Porceddu, E., Colaprico, G., and Lafiandra, D. (1991a). Comparison of the B and D subunits of glutenin encoded at the *Glu-D3* locus in two biotypes of the common wheat cultivar Newton with different technological characteristics. J Cereal Sci. **14**: 35-46.

Masci, S. M., Porceddu, E., and Lafiandra, D. (1991b). Two-dimensional electrophoresis of 1D-encoded B and D glutenin subunits in common wheats with similar omega gliadins. Biochem Genet. **29**: 403-413.

Masci, S. M., Lafiandra, D., Porceddu, E., Lew, E. J.-L., Tao, H. B., and Kasarda, D. D. (1993). D-glutenin subunits: N terminal sequences and evidence for the presence of cysteine. Cereal Chem. **70**: 581-585.

Masci, S., Egorov, T. A., Ronchi, C., Kuzmicky, D. D., Kasarda, D. D., and Lafiandra, D. (1999). Evidence for the presence of only one cysteine residue in the D-type low molecular weight subunits of wheat glutenin. J Cereal Sci. **29**: 17-25.

Matouschek, A., Rospert, S., Schmid, K., Glick, B. S., and Schatz, G. (1995). Cyclophilin catalyzes protein folding in yeast mitochondria. Proc Natl Acad Sci USA. **92**: 6319-6323.

McFadden, E. S., and Sears, E. R. (1946). The origin of Triticum spelta and its free threshing hexaploid relatives. J Hered. **37**: 81-89,107-116.

McMullen, M. D., Byrne, P. F., Snook, M. E., Wiseman, B. R., Lee, E. A., Widstrom, N. W., and Coe, E. H. (1998). Quantitative trait loci and metabolic pathways. Proc Natl Acad Sci USA. **95**: 1996-2000.

Meunier, L., Usherwood, Y. K., Chung, K. T., and Hendershot, L. M. (2002). A subset of chaperones and folding enzymes form multiprotein complexes in endoplasmic reticulum to bind nascent proteins. Mol Biol Cell. **13**: 4456-4469.

Mi, H., Kops, O., Zimmermann, E., Jaschke, A. and Tropschug, M. (1996). A nuclear RNA-binding cyclophilin in human T cells. FEBS Lett. **378**: 201-205.

Miflin, B., Field, J., and Shewry, P. R. (1983). Cereal storage proteins and their effects on technological properties. In Seed Proteins, J. Daussant, J. Mosse, and J. Vaughan, eds. (London, Academic Press), pp. 255-319.

Mita, T., and Matsumoto, M. (1981). Flow properties of aqueous gluten and gluten methyl ester dispersions. Cereal Chem. **58**: 57-61.

Miyashita, N., Mori, N., and Tsunewaki, K. (1994). Molecular variation in chloroplast DNA regions in ancestral species of wheat. Genetics. **137**: 883-889.

Moonen, J. H. E., Scheepstra, A., and Graveland, A. (1982). Use of SDS-sedimentation test and SDS-polyacrylamide gel electrophoresis for screening breeders' samples of wheat for breadmaking quality. Euphytica. **31**: 677-690.

Moonen, J. H. E., Scheepstra, A., and Graveland, A. (1983). The positive effects of the high molecular weight subunits 3 + 10 and 2* of glutenin on the breadmaking quality of wheat cultivars. Euphytica. **32**: 735-742.

Moore, G., Devos, K. M., Wang, Z., Gale, M. D. (1995). Cereal genome evolution. Grasses, line up and form a circle. Curr Biol. **5**: 737-739.

Morgan, T. H. (1911). Random segregation versus coupling in Mendelian inheritance. Science. **34**: 384.

Mori, N., Miyashita, N. T., Terachi, T., and Nakamura, C. (1997). Variation in *cox*II intron in the wild ancestral species of wheat. Hereditas. **126**: 281-288.

Morris, C. F. (1998). Genetic determinants of wheat grain quality. Proceedings of the 9th International Wheat Genetics Symposium (Saskatchewan, Canada, University Extension Press) pp. 249-253.

Morris, R., and Sears, E. R. (1967). The cytogenetics of wheat and its relatives. In: Wheat and wheat improvement, L. P. Quisenberry, and L. P. Reitz, eds. (Madison, WI, American Society of Agronomy), pp. 19-87.

Müller, S., Vensel, W. H., Kasarda, D. D., Köhler, P., and Wieser, H. (1998). Disulphide bonds of adjacent cysteine residues in low molecular weight subunits of wheat glutenin. J Cereal Sci. **27**: 109-116.

Mullis, K. B. and Faloona, F. A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Method Enzymol. **155**: 335-350.

Nakai, K. and Kanehisa, M. (1992). A knowledge base for predicting protein localization sites in eukaryotic cells. Genomics. **14**: 897-911.

Napier, J. A., Richard, G., Turner, M. F., and Shewry, P. R. (1997). Trafficking of wheat gluten proteins in transgenic tobacco plants: gamma-gliadin does not contain an endoplasmic reticulum-retention signal. Planta. **203**: 488-494.

Nestle, F. P., Colwill, K., Harper, S., Powson, T. and Anderson, S. K. (1996). RS cyclophilins: identification of an NK-TR_1-related cyclophilin. Gene. **180**: 151-155.

Noiva, R., Freedman, R. B., and Lennarz, W. J. (1993). Peptide binding to protein disulfide isomerase occurs at a site distinct from the active sites. J Biol Chem. **268**: 19210-19217.

Ohsako, T., Wang, G. Z., and Miyashita, N. T. (1996). Polymerase chain reaction-single strand conformational polymorphism analysis of intra- and interspecific variations in organellar DNA regions of *Aegilops mutica* and related species. Genes Genet Syst. **71**: 281-292.

Oliver, J. D., van_der_Wal, F. J., Bulleid, N. J., and High, S. (1997). Interaction of the thiol-dependent reductase ERp57 with nascent glycoproteins. Science. **275**: 86-88.

Orsi, A., Sparvoli, F., and Ceriotti, A. (2001). Role of individual disulfide bonds in the structural maturation of a low molecular weight glutenin subunit. J Biol Chem. **276**: 32322-32329.

Osborne B. G., Turnbull, K. M., Anderssen R. S., Rahman S., Sharp P. J. and Appels R. (2001). The hardness locus in Australian wheat lines. Aust J Agric Res. **52**: 1275-1286.

Osborne, T. B. (1924). The Vegetable Proteins, 2nd edn (London, Longmans, Green and Co.).

Page, A. P. (1997). Cyclophilin and protein disulfide isomerase genes are co-transcribed in a functionally related manner in *Caenorhabditis elegans*. DNA Cell Biol. **16**: 1335-1343.

Parker, G. D., Chalmers, K. J., Rathjen, A. J. and Langridge, P. (1998). Mapping loci associated with flour colour in wheat (*Triticum aestivum* L.). Theor Appl Genet. **97**: 238-245.

Parker, G. D., Chalmers, K. J., Rathjen, A. J. and Langridge, P. (1999). Mapping loci associated with milling yield in wheat (*Triticum aestivum* L.). Mol Breed. **5**: 561-568.

Parker, G. D., Fox, P. N., Langridge, P., Chalmers, K., Whan, B., and Ganter, P. F. (2002). Genetic diversity within Australian wheat breeding programs based on molecular and pedigree data. Euphytica. **124**: 293-306.

Paterson, A. H., Tanksley, S. D. and Sorrells, M. E. (1991). DNA markers in plant improvement. Adv Agron. **44**: 39-90.

Paterson, A. H., Lin, Y.-R., Li, Z., Schertz, K. F., Doebley, J. W., Pinson, S. R. M., Liu, S.-C., Stansel, J. W., and Irvine, J. E. (1995). Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science. **269**: 1714-1718.

Paterson, A. H. (1996). Making genetic maps. In: Genome mapping in plants. Paterson A. H. (ed.). R.G. Landes Company, Austin. pp. 23-29.

Paterson, A. H., Bowers, J. E., Burow, M. D., Draye, X., Elsik, C. G., Jiang, C. X., Katsar, C. S., Lan, T. H., Lin, Y. R., Ming, R., and Wright, R. J. (2000). Comparative genomics of plant chromosomes. Plant Cell. **12**: 1523-1540.

Patnaik, D., and Khurana, P. (2001). Germins and germin like proteins: an overview. Indian J. Exp. Biol. **39**: 191-200.

Paull, J. G., Chalmers, K. J., Karakousis, A., Kretschmer, J. M., Manning, S., and Langridge, P. (1998). Genetic diversity in Australian wheat varieties and breeding material based on RFLP data. Theor Appl Genet. **96**: 435-446.

Payne, P. I., and Corfield, K. G. (1979). Subunit composition of wheat glutenin proteins, isolated by gel filtration in a dissociating medium. Planta. **145**: 83-88.

Payne, P. I., Corfield, K. G., and Blackman, J. A. (1979). Identification of a high molecular weight subunit of glutenin whose presence correlates with breadmaking quality in wheats of related pedigree. Theor Appl Genet. **55**: 153-115.

Payne, P. I., and Lawrence, G. J. (1983). Catalogue of alleles for the complex gene loci, *Glu-A1, Glu-B1, Glu-D1* which code for high-molecular-weight subunits of glutenin in hexaploid wheat. Cereal Res Commun. **11**: 29-35.

Pereira, M., and Lee, M. (1995). Identification of genomic regions affecting plant height in sorghum and maize. Theor Appl Genet. **90**: 380-388.

Perretant, M. R., Cadalen, T., Charmet, G., Sourdille, P., Nicolas, P., Boeuf, C., Tixier, M. H., Branlard, G., Bernard, S., and Bernard, M. (2000). QTL analysis of bread-making quality in wheat using a doubled haploid population. Theor Appl Genet. **100**: 1167-1175.

Pestova, E., Ganal, M. W., and Röder, M. S. (2000). Isolation and mapping of microsatellite markers specific for the D genome of bread wheat. Genome. **43**: 689-697.

Podkowinski, J., Jelenska, J., Sirikhachornkit, A., Zuther, E., Haselkorn, R. and Gornicki, P. (2003). Expression of cytosolic and plastid acetyl-coenzyme A carboxylase genes in young wheat plants. Plant Physiol. **131**: 763-772.

Price, E. R., Zydowsky, L. D., Jin, M. J., Baker, C. H., McKeon, F. D., and Walsh, C. T. (1991). Human cyclophilin B: a second cyclophilin gene encodes a peptidyl-prolyl isomerase with a signal sequence. Proc Natl Acad Sci USA. **88**: 1903-1907.

Qi, L. L., Echalier, B., Chao, S., Lazo, G. R., Butler, G. E., Anderson, O. D., Akhunov, E. D., Dvorak, J., Linkiewicz, A. M., Ratnasiri, A., Dubcovsky, J., Bermudez-Kandianis, C. E., Greene, R. A., Kantety, R., La Rota, C. M., Munkvold, J. D., Sorrells, S. F., Sorrells, M. E., Dilbirligi, M., Sidhu, D., Erayman, M., Randhawa, H. S., Sandhu, D., Bondareva, S. N., Gill, K. S., Mahmoud, A. A., Ma, X. F., Miftahudin, Gustafson, J. P., Conley, E. J., Nduati, V., Gonzalez-Hernandez, J. L., Anderson, J. A., Peng, J. H., Lapitan, N. L., Hossain, K. G., Kalavacharla, V., Kianian, S. F., Pathan, M. S., Zhang, D. S., Nguyen, H. T., Choi, D. W., Fenton, R. D., Close, T. J., McGuire, P. E., Qualset, C. O. and Gill, B. S. (2004). A chromosome bin map of 16,000 expressed sequence tag

loci and distribution of genes among the three genomes of polyploid wheat. Genetics. **168**: 701-712.

Rahfeld, J. U., Schierhorn, A., Mann, K., and Fischer, G. (1994). A novel peptidyl-prolyl *cis/trans* isomerase from *Escherichia coli*. FEBS Lett. **343**: 65-69.

Ramakrishna, W., Emberton, J., SanMiguel, P., Ogden, M., Llaca, V., Messing, J. and Bennetzen, J. L. (2002). Comparative sequence analysis of the sorghum Rph region and the maize Rp1 resistance gene complex. Plant Physiol. **130**: 1728-1738.

Randhawa, H. S., Dhaliwal, H. S., and Harjit, S. (1997). Diversity for HMW glutenin subunit composition and the origin of polyploid wheats. Cereal Res Commun. **25**: 77-84.

Rassow, J., Mohrs, K., Koidl, S., Barthelmess, I. B., Pfanner, N., and Tropschug, M. (1995). Cyclophilin 20 is involved in mitochondrial protein folding in cooperation with molecular chaperones Hsp70 and Hsp60. Mol Cell Biol. **15**: 2654-2662.

Rebetzke, G. J., Appels, R., Morrison, A. D., Richards, R. A., McDonald, G., Ellis, M. H., Spielmeyer, W. and Bonnett, D. G. (2001). Quantitative trait loci on chromosome 4B for coleoptile length and early vigour in wheat (*Triticum aestivum* L.) Aust J Agric Res. **52**: 1221-1234

Roden, L. T., Miflin, B. J., and Freedman, R. B. (1982). Protein disulphide-isomerase is located in the endoplasmic reticulum of developing wheat endosperm. FEBS Lett. **38**: 121-124.

Röder, M. S., Plaschke, J., Konig, S. U., Borner, A., Sorrells, M. E., Tanksley, S. D., and Ganal, M. W. (1995). Abundance, variability and chromosomal location of microsatellites in wheat. Mol Gen Genet. **246**: 327-333.

Röder, M. S., Korzun, V., Wendehake, K., Plaschke, J., Tixier, M. H., Leroy, P. and Ganal M. W. (1998). A microsatellite map of wheat. Genetics. **149**: 2007-2023.

Romano, P. G. N., Edvardsson, A., Ruban, A. V., Andersson, B., Vener, A. V., Gray, J. E. and Horton, P. (2004a). *Arabidopsis* AtCYP20-2 is a light-regulated cyclophilin-type

peptidyl-prolyl cis-trans isomerase associated with the photosynthetic membranes. Plant Physiol. **134**: 1244-1247.

Romano, P. G. N., Horton, P. and Gray, J. E. (2004b). The *Arabidopsis* cyclophilin gene family. Plant Physiol. **134**: 1268-1282.

Rubin, R., Levanony, H., and Galili, G. (1992). Evidence for the presence of two different types of protein bodies in wheat endosperm. Plant Physiol. **99**: 718-724.

Saito, T., Niwa, Y., Ashida, H., Tanaka, K., Kawamukai, M., Matsuda, H. and Nakagawa, T. (1999). Expression of a gene for cyclophilin which contains an amino-terminal endoplasmic reticulum-targeting signal. Plant Cell Physiol. **40**: 77-87.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor laboratory Press, New York.

Sanger F., Nicklen S., Coulson A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci. USA **74**: 5463-5467.

Sarma, R. N., Gill, B. S., Sasaki, T., Galiba, G., Sutka, J., Laurie, D. A., and Snape, J. W. (1998). Comparative mapping of the wheat chromosome 5A Vrn-A1 region with rice and its relationship to QTL for flowering time. Theor Appl Genet. **97**: 103-109.

Sax, K. (1922). Sterility in wheat hybrids. II. Chromosome behaviour in partially sterile hybrids. Genetics. **7**: 513-522.

Scherens, B., Dubois, E., and Messenguy, F. (1991). Determination of the sequence of the yeast YCL313 gene localized on chromosome III. Homology with the protein disulfide isomerase (PDI gene product) of other organisms. Yeast. **7**: 185-193.

Scholze, C., Peterson, A., Diettrich, B., and Luckner, M. (1999). Cyclophilin isoforms from *Digitalis lanata*. Sequences and expression during embryogenesis and stress. J Plant Physiol. **155**: 212-219.

Schönbrunner, E. R., and Schmid, F. X. (1992). Peptidyl-prolyl cis-trans isomerase improves the efficiency of protein disulfide isomerase as a catalyst of protein folding. Proc Natl Acad Sci USA. **89**: 4510-4513.

Schreiber, S. L. (1992). Immunophilin-sensitive protein phosphatase action in cell signaling pathways. Cell. **70**: 365-368.

Schropp, P., and Wieser, H. (1996). Effects of high molecular weight subunits of glutenin on the rheological properties of wheat gluten. Cereal Chem. **73**: 410-413.

Schubert, M., Petersson, U. A., Haas, B. J., Funk, C., Schroder, W. P., Kieselbach, T. (2002). Proteome map of the chloroplast lumen of *Arabidopsis thaliana*. J Biol Chem. **277**: 8354-8365.

Sears, E. R. (1954). The aneuploids of common wheat. In Research Bulletin 572, University of Missouri College of Agriculture, Agriculture Experimental Station, pp. 3-58.

Shah, M. M., Yen, Y., Gill, K. S., and Baenziger, P. S. (2000). Comparisons of RFLP and PCR-based markers to detect polymorphism between wheat cultivars. Euphytica. **114**: 135-142.

Shani, N., Rosenberg, N., Kasarda, D. D., and Galili, G. (1994). Mechanisms of assembly of wheat high molecular weight glutenins inferred from expression of wild-type and mutant subunits in transgenic tobacco. J Biol Chem. **269**: 8924-8930.

Sheldon, P. S., and Venis, M. A. (1996). Purification and characterization of cytosolic and microsomal cyclophilins from maize (*Zea mays*). Biochem J. **315**: 965-969.

Shewry, P. R., Napier, J. A., and Tatham, A. S. (1995). Seed storage proteins, structures and biosynthesis. Plant Cell. **7**: 945-956.

Shewry, P. R., and Tatham, A. S. (1997). Disulphide bonds in wheat gluten proteins. J Cereal Sci. **25**: 207-227.

Shewry, P. R., Halford, N. G., Tatham, A. S., Popineau, Y., Lafiandra, D., and Belton, P. S. (2003). The high molecular weight subunits of wheat glutenin and their role in determining wheat processing properties. Adv Food Nutr Res. **45**: 219-302.

Shimoni, Y., and Galili, G. (1996). Intramolecular disulfide bonds between conserved cysteines in wheat gliadins control their deposition into protein bodies. J Biol Chem. **271**: 18869-18874.

Shimoni, Y., Segal, G., Zhu, X., and Galili, G. (1995a). Nucleotide sequence of a wheat cDNA encoding protein disulfide isomerase. Plant Physiol. **107**: 281.

Shimoni, Y., Zhu, X., Levanony, H., Segal, G., and Galili, G. (1995b). Purification, characterization, and intracellular localization of glycosylated protein disulfide isomerase from wheat grains. Plant Physiol. **108**: 327-335.

Shorrosh, B. S., and Dixon, R. A. (1991). Molecular cloning of a putative plant endomembrane protein resembling vertebrate protein disulfide-isomerase and a phosphatidylinositol-specific phospholipase C. Proc Natl Acad Sci USA. **88**: 10941-10945.

Shorrosh, B. S., and Dixon, R. A. (1992). Molecular characterization and expression of an alfalfa protein with sequence similarity to mammalian ERp72, a glucose-regulated endoplasmic reticulum protein containing active site sequences of protein disulphide isomerase. Plant J. **2**: 51-58.

Simonetti, M. C., Bellomo, M. P., Laghetti, G., Perrino, P., Simeone, R. and Blanco, A. (1999). Quantitative trait loci influencing free-threshing habit in tetraploid wheats, Genet Resour Crop Evol. **46**: 267-271.

Singh, N.K., Raghuvanshi, S., Srivastava, S. K., Gaur, A., Pal, A. K., Dalal, V., Singh A., Ghazi, I. A., Bhargav, A., Yadav, M., Dixit, A., Batra, K., Gaikwad, K., Sharma, T. R., Mohanty, A., Bharti, A. K., Kapur, A., Gupta, V., Kumar, D., Vij, S., Vydianathan, R., Khurana, P., Sharma, S., McCombie, W. R., Messing, J., Wing, R., Sasaki, T., Khurana, P., Mohapatra, T., Khurana, J. P. and Tyagi, A. K. (2004). Sequence analysis of the long arm of rice chromosome 11 for rice-wheat synteny. Funct Integr Genomics. **4**: 102-117.

Sissons, M. J., Békés, F., and Skerritt, J. H. (1998). Isolation and functionality testing of low molecular weight glutenin subunits. Cereal Chem. **73**: 644-649.

Solovyov, A., Xiao, R.Y. and Gilbert, H.F. (2004). Sulfhydryl oxidation, not disulfide isomerization, is the principal function of protein disulfide isomerase in yeast Saccharomyces cerevisiae. J Biol Chem. **279**: 34095-34100.

Song, J. L., and Wang, C. C. (1995). Chaperone-like activity of protein disulfide-isomerase in the refolding of rhodanese. Eur J Biochem. **231**: 312-316.

Sourdille, P., Snape, J. W., Cadalen, T., Charmet, G., Nakata, N., Bernard, S. and Bernard, M. (2000a). Detection of QTLs for heading time and photoperiod response in wheat using a doubled-haploid population. Genome. **43**: 487-494.

Sourdille, P., Tixier, M. H., Charmet, G., Gay, G., Cadalen, T., Bernard, S., and Bernard, M. (2000b). Location of genes involved in ear compactness in wheat (Triticum aestivum) by means of molecular markers. Mol Breed. **6**: 247-255.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol. **98**:503-517.

Steinmann, B., Bruckner, P., and Superti-Furga, A. (1991). Cyclosporin A slows collagen triple-helix formation in vivo: indirect evidence for a physiologic role of peptidyl-prolyl cis-trans-isomerase. J Biol Chem. **266**: 1299-1303.

Stoller, G., Rucknagel, K. P., Nierhaus, K. H., Schmid, F. X., Fischer, G. and Rahfeld, J. U. (1995). A ribosome-associated peptidyl-prolyl cis/trans isomerase identified as the trigger factor. EMBO J. **14**:4939-4948.

Takahashi, N., Hayano, T. and Suzuki, M. (1989). Peptidyl-prolyl cis-trans isomerase is the cyclosporin A-binding protein cyclophilin. Nature. **337**:473–475.

Takemoto, Y., Coughlan, S. J., Okita, T. W., Satoh, H., Ogawa, M., and Kumamaru, T. (2002). The rice mutant esp2 greatly accumulates the glutelin precursor and deletes the protein disulfide isomerase. Plant Physiol. **128**: 1212-1222.

Takumi, S., Nasuda, S., Liu, Y. G., and Tsunewaki, K. (1993). Wheat phylogeny determined by RFLP analysis of nuclear DNA. I. Einkorn wheat. Jpn J Genet. **68**: 73-79.

Talbert, L. E., Magyer, G. M., Lavin, M., Blake, N. K., and Moylan, S. L. (1991). Molecular evidence for the origin of the S-derived genomes of polyploid *Triticum* species. Am J Bot. **78**: 340-349.1

Talbert, L. E., Blake, N. K., Storlie, E. W., and Lavin, M. (1995). Variation in wheat based on low-copy DNA sequence comparisons. Genome. **38**: 951-957.

Tamás, L., Gras, P. W., Solomon, R. G., Morell, M. K., Appels, R., and Békés, F. (2002). Chain extension and termination as a function of cysteine content and the length of the central repetitive domain in storage proteins. J Cereal Sci. **36**: 313-325.

Tao, H. P., Adalsteins, A. E., and Kasarda, D. D. (1992). Intermolecular disulphide bonds link specific high-molecular-weight glutenin subunits in wheat endosperm. Biochim Biophys Acta. **1159**: 13-21.

Tarchini, R., Biddle, P., Wineland, R., Tingey, S. and Rafalski, A. (2000). The complete sequence of 340 kb of DNA around the rice Adh1-adh2 region reveals interrupted colinearity with maize chromosome 4. Plant Cell. **12**: 381-391.

Tatham, A. S., Shewry, P. R. and Belton, P. S. (1990). Structural studies of cereal prolamines, including wheat gluten. In: Advances in cereal science and technology. (Ed. Y. Pomeranz). American Association of Cereal Chemists, USA.

Thompson, S., Bishop, D. H. L., Tatham, A. S., and Shewry, P. R. (1994). Exploring disulphide bond formation in a low molecular weight subunit of glutenin using a baculovirus expression system. In: Gluten Proteins 1993 (Detmold, Association of Cereal Research), pp. 345-355.

Tikhonov, A. P., SanMiguel, P. J., Nakajima, Y., Gorenstein, N. M., Bennetzen, J. L., and Avramova, Z. (1999). Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. Proc Natl Acad Sci USA. **96**: 7409-7414.

Tixier, M. H., Sourdille, P., Charmet, G., Gay, G., Jaby, C., Cadalen, T., Bernard, S., Nicolas, P., and Bernard, M. (1998). Detection of QTLs for crossability in wheat using a doubled haploid population. Theor Appl Genet. **97**: 1076-1082.

Turano, C., Coppari, S., Altieri, F., and Ferraro, A. (2002). Proteins of the PDI family: unpredicted non-ER locations and functions. J Cell Physiol. **193**: 154-163.

Uthayakumaran, S., Stoddard, F. L., Gras, P. W., and Békés, F. (2000). Optimized methods for incorporating glutenin subunits into wheat dough for extension and baking studies. Cereal Chem. **77**: 731-736.

Van Deynze, A. E., Sorrells, M. E., Park, W. D., Ayres, N. M., Fu, H., Cartinhour, S. W., Paul, E., and McCouch, S. R. (1998). Anchor probes for comparative mapping of grass genera. Theor Appl Genet. **97**: 356-369.

Van-Dijk, A. A., Van-Wijk, L. L., Van-Vliet, A., Haris, P., Van-Swieten, E., Tesser, G. I., and Robillard, G. T. (1997). Structure characterization of the central repetitive domain of high molecular weight gluten proteins. I. Model studies using cyclic and linear peptides. Protein Sci. **6**: 637-648.

Van Nguyen, P., Peter, F., and Soling, H. (1989). Four intracisternal calcium-binding glycoproteins from rat liver microsomes with high affinity for calcium. No indication for calsequestrin-like proteins in inositol 1,4,5-trisphosphate-sensitive calcium sequestering rat liver vesicles. J Biol Chem. **264**: 17494-17501.

Van Slageren, M. W. (1994). Wild wheats: a monograph of Aegilops L. and Amblyopyrum (Jaub. & Spach) Eig (Poaceae). Wageningen Agric Univ Pap. **1994 (7)**: 513 pp.

Vener, A. V., Rokka, A., Fulgosi, H., Andersson, B., and Herrmann, R. G. (1999). A cyclophilin-regulated PP2A-like protein phosphatase in thylakoid membranes of plant chloroplasts. Biochemistry. **38**: 14955-14965.

Vittozzi, L., and Silano, V. (1976). The phylogeny of protein alpha-amylase inhibitors from wheat seed and the speciation of polyploid wheats. Theor Appl Genet. **48**: 279-284.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. **23**: 4407-4414.

Vuori, K., Pihlajaniemi, T., Myllyla, R., and Kivirikko, K. I. (1992). Site-directed mutagenesis of human protein disulphide isomerase: effect on the assembly, activity and

endoplasmic reticulum retention of human prolyl 4-hydroxylase in *Spodoptera frugiperda* insect cells. EMBO J. **11**: 4213-4217.

Vuylsteke, M., Mank, R., Antonise, R., Bastiaans, E., Senior, M. L., Stuber, C. W., Melchinger, A. E., Lübberstedt, T., Xia, X. C., Stam, P., *et al.* (1999). Two high-density AFLP® linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers. Theor Appl Genet. **99**: 921-935.

Waldron, B. L., Moreno-Sevilla, B., Anderson, J. A., Stack, R. W., and Frohberg, R. C. (1999). RFLP mapping of QTL for fusarium head blight resistance in wheat. Crop Sci. **39**: 805-811.

Walker, K. W., Lyles, M. M., and Gilbert, H. F. (1996). Catalysis of oxidative protein folding by mutants of protein disulfide isomerase with a single active-site cysteine. Biochemistry. **35**: 1972-1980.

Wang, B. B., Hayenga, K. J., Payan, D. G. and Fisher, J. M. (1996). Identification of a nuclear-specific cyclophilin which interacts with the proteinase inhibitor eglin c. Biochem J. **314**: 313-319.

Ware, D., Jaiswal, P., Ni, J., Pan, X., Chang, K., Clark, K., Teytelman, L., Schmidt, S., Zhao, W., Cartinhour, S., McCouch, S. and Stein, L. (2002). Gramene: a resource for comparative grass genomics. Nucleic Acids Res. **30**: 103-105.

Wellner, N., Belton, P. S., and Tatham, A. S. (1996). Fourier trans-form IR spectroscopic study of hydration induced structure changes in the solid state of omega gliadins. Biochem J. **319**: 741-747.

Weng, Y., Tuleen, N. A. and Hart, G. E. (2000). Extended physical maps and a consensus physical map of the homoeologous group-6 chromosomes of wheat (*Triticum aestivum* L-em Thell.). Theor Appl Genet. **100**: 519-527.

Werner, W. E., Adalsteins, A. E., and Kasarda, D. D. (1992). Composition of high-molecular-weight glutenin subunit dimers formed by partial reduction of residue glutenin. Cereal Chem. **69**: 535-541.

Wetterau, J. R., Combs, K. A., McLean, L. R., Spinner, S. N., and Aggerbeck, L. P. (1991). Protein disulfide isomerase appears necessary to maintain the catalytically active structure of the microsomal triglyceride transfer protein. Biochemistry. **30**: 9728-9735.

Wilson, W. A., Harrington, S. E., Woodman, W. L., Lee, M., Sorrells, M. E., and McCouch, S. R. (1999). Inferences on the genome structure of progenitor maize through comparative analysis of rice, maize and the domesticated panicoids. Genetics. **153**: 453-473.

Winter, J., Klappa, P., Freedman, R. B., Lilie, H., and Rudolph, R. (2002). Catalytic activity and chaperone function of human protein-disulfide isomerase are required for the efficient refolding of proinsulin. J Biol Chem. **277**: 310-317.

Winter, P. and Kahl, G. (1995). Molecular marker technology for plant improvement. World J Microbiol Biotechnol. **11**: 438-448.

Wu, J., Matunis, M. J., Kraemer, D., Blobel, G. and Coutavas, E. (1995). Nup358, a cytoplasmically exposed nucleoporin with peptide repeats, ranGTP binding sites, zinc fingers, a cyclophilin A homologous domain and a leucine-rich region. J Biol Chem. **270**: 14209-14213.

Wu, J., Maehara, T., Shimokawa, T., Yamamoto, S., Harada, C., Takazaki, Y., Ono, N., Mukai, Y., Koike, K., Yazaki, J., Fujii, F., Shomura, A., Ando, T., Kono, I., Waki, K., Yamamoto, K., Yano, M., Matsumoto, T. and Sasaki, T. (2002). A comprehensive rice transcript map containing 6591 expressed sequence tag sites. Plant Cell. **14**: 525-535.

Xiao, R.Y., Wilkinson, B., Solovyov, A., Winther, J.R., Holmgren, A., Lundstrom-Ljung, J., Gilbert, H.F. (2004). The contributions of protein disulfide isomerase and its homologues to oxidative protein folding in the yeast endoplasmic reticulum. J Biol Chem. **279**: 49780-49786.

Xu, Z. J., Ueda, K., Masuda, K., Ono, M., and Inoue, M. (2002). Molecular characterization of a novel protein disulfide isomerase in carrot. Gene. **284**: 225-231.

Yamamori, M. and Quynh, N. T. (2000). Differential effects of Wx-A1, -B1 and -D1 protein deficiencies on apparent amylose content and starch pasting properties in common wheat. Theor Appl Genet. **100**: 32-38.

Yang, W.-M., Inouye, C. J., and Seto, E. (1995). Cyclophilin A and FKBP12 interact with YY1 and alter its transcriptional activity. J Biol Chem. **270**: 15187-15193.

Yao, Y., Zhou, Y., and Wang, C. (1997). Both the isomerase and chaperone activities of protein disulfide isomerase are required for the reactivation of reduced and denatured acidic phospholipase A2. EMBO J. **16**: 651-658.

Young, N. D. (1994). Constructing a genetic linkage map with DNA markers. In: DNA-based markers in plants. Phillips, R. L. and Vasil, I. K. (Eds.) Kluwer, London. pp. 39-57.

Yu, J., Hu, S., Wang, J., Wong, G. K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., *et al.* (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). Science. **296**: 79-92.

Zanetti, S., Winzeler, M., Keller, M., Keller, B. and Messmer, M. (2000). Genetic analysis of pre-harvest sprouting resistance in wheat and spelt. Crop Sci. **40**: 1406-1417.

Zanetti, S., Winzeler, M., Feuillet, C., Keller, B. and Messmer, M. (2001). Genetic analysis of bread-making quality in wheat and spelt. Plant Breed. **120**: 13-19.

APPENDICES

APPENDIX I

CHEMICALS AND REAGENTS

A. Enzymes

A number of enzymes were used in this project, many of which were supplied in the kits used for PCR, cloning and sequencing reactions. These enzymes included:

Platinum *Taq* polymerase and HotStarTaq for PCR amplifications which were provided by Invitrogen and Qiagen, respectively.

T4 DNA Ligase for ligation of fragments of the *PDI* genes into the pGEM-T-Easy vector, which was provided by Promega.

Polymerases for sequencing reactions were provided in the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit v.3 supplied by Applied Biosystems.

A number of restriction endonucleases used in the digestion of the pGEM-T-Easy and pBluescript SK+ vectors, Southern blotting experiments, RFLP analysis of the *PDI* and cyclophilin genes and digestion of the PCR-RFLP molecular markers for the *PDI* genes, which were supplied by Promega, MBI Fermentas and New England Biolabs.

The Klenow fragment of *E. coli* DNA polymerase I was used in the radiolabeling of probes used for Southern blot experiments with the Prime-a-Gene Labeling System or Ready-To-Go DNA Labeling Beads, supplied by Promega and Pharmacia, respectively.

An Alkaline Protease solution was used in the purification of plasmid DNA and was supplied in the Wizard *Plus* SV Minipreps DNA Purification System from Promega.

RNase A (10μg/μl) was prepared and purified as follows:
100mg of pancreatic RNase A raw extracts (Progen) was dissolved in 10mL of sterile 10mM Tris-Cl (pH 7.5) containing 15mM NaCl and heated in a boiling water bath for 5 minutes to inactivate any DNase present. The DNase-free RNase A was cooled to room temperature, dispensed in 20μl aliquots and stored at -20°C.

B. Buffers

A number of buffers were supplied with kits or enzymes for use in PCR reactions, DNA purifications, restriction digestions, ligation and radiolabelling reactions. In addition, the following buffers were prepared:

1. TE Buffer: 10mM Tris-Cl (pH 8.0), 1mM EDTA (pH 8.0).

2. TAE (50X): 242g Tris base, 57.1mL Glacial acetic acid, 100mL 0.5M EDTA (pH 8.0). These were dissolved and made up to 1000mL in sterile ddH_2O.

3. SM Buffer: 100mM NaCl, 50mM Tris-Cl (pH 7.5), 8mM $MgSO_4$, 0.01% gelatin. Chloroform was added to the buffer at the time of use at a final concentration of 1%.

4. Sequencing dilution buffer: 1.4mM $MgCl_2$, 60mM Tris-Cl (pH 7.5), 3M sodium acetate, pH 5.2.

5. SSC (20X): 3M NaCl, 0.3M sodium citrate.

6. Denaturation solution: 1.5M NaCl, 0.5M NaOH.

7. Neutralisation solution: 1.5M NaCl, 0.5M Tris-Cl (pH 7.2).

8. Prehybridisation Buffer I: 0.2% Ficoll, 0.2% bovine serum albumen (BSA), 0.2% polyvinyl pyrrolidone (PVP), 1% SDS, 1.2% HEPES, 3x SSC, 0.002% sheared herring sperm DNA (Sigma).

C. General Solutions and Stocks

1. 20% Maltose

   20g of maltose (Aldrich) was made up to 100mL in sterile, ddH_2O, sterilised with a 0.22μm filter and stored at 4°C.

2. X-Gal (20mg/mL)

   100mg of 5-Bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal) (Progen) was dissolved in dimethylformamide to a final volume of 5mL. The prepared X-gal was stored at -20°C in a light-safe tube.

3. 20% IPTG

   2g of isopropylthio-β-D-galactoside (IPTG) (Progen) was dissolved in ddH_2O to a final volume of 10mL. The prepared IPTG was sterilised with a 0.22μm filter and stored at 4°C.

4. Ethidium Bromide (10mg/mL)

   0.5g ethidium bromide (Sigma) was dissolved in 50mL of ddH_2O by stirring for 2 hours with a magnetic stirrer. The prepared ethidium bromide was stored in a light-safe tube at room temperature.

5. 10% SDS

   10g sodium dodecyl sulfate (SDS) (electrophoresis grade, BioRad) was dissolved in 90mL ddH_2O and heated at 68°C until the SDS was completely dissolved. This solution was adjusted to pH 7.2 with dilute acetic acid and made up to a final volume of 100mL with ddH_2O, sterilised with a 0.22μm filter and stored at room temperature.

6. 1M $MgSO_4$

101.5g of $MgSO_4$ (BDH) was made up to a final volume of 500mL with ddH_2O and stored at room temperature.

7. 2M Mg^{2+} stock

20.33g $MgCl_2$ • 6H2O and 24.65g $MgSO_4$ • 7H2O were made up to 100mL with ddH_2O and sterilised with a 0.22μm filter.

8. 0.5M EDTA

186.1g of disodium ethylenediaminetetra-acetate-$2H_2O$ (Ajax Chemicals) was added to 800mL of ddH_2O and the pH was adjusted to 8.0 by the addition of 10N NaOH. The solution was allowed to dissolve by stirring on a magnetic stirrer and made up to a final volume of 1000mL with ddH_2O. The EDTA was sterilised by autoclaving at 121°C for 20 minutes and stored at room temperature.

9. 3M Sodium Acetate

40.8g of sodium acetate-$3H_2O$ (Ajax Chemicals) was dissolved in ddH_2O, the pH was adjusted to 5.2 with glacial acetic acid and made up to 100mL final volume with ddH_2O. The prepared solution was sterilised by autoclaving at 121°C for 20 minutes and stored at room temperature.

10. Ampicillin (50mg/mL)

200mg of Ampicillin (Sigma) was dissolved in ddH_2O to a final volume of 4mL, sterilised with a 0.22μm filter and stored in a light-safe tube at -20°C.

D. Microbiological media

1. LB broth (Luria-Bertani broth)

1% w/v bacto-tryptone (Oxoid),
0.5% bacto-yeast extract (Oxoid),
1% w/v NaCl.

This media was sterilised by autoclaving at 121°C for 20 minutes. Where required, agar was added (20g/L) to make solid LB agar prior to autoclaving.

2. SOC media

2.0g Tryptone (Oxoid Media)

0.5g Yeast extract (Oxoid Media)

1ml 1M NaCl

0.25ml 1M KCl

1ml 2M Mg2+ stock, filter-sterilized.

1ml 2M glucose, filter-sterilized.

The tryptone, yeast extract, NaCl and KCl were added to 97ml ddH_2O and autoclaved at 121°C for 20 minutes. The 2M Mg^{2+} stock and 2M glucose were added and the final volume made to 100ml with sterile ddH_2O.

APPENDIX II

MAPS OF VECTORS USED IN THIS PROJECT

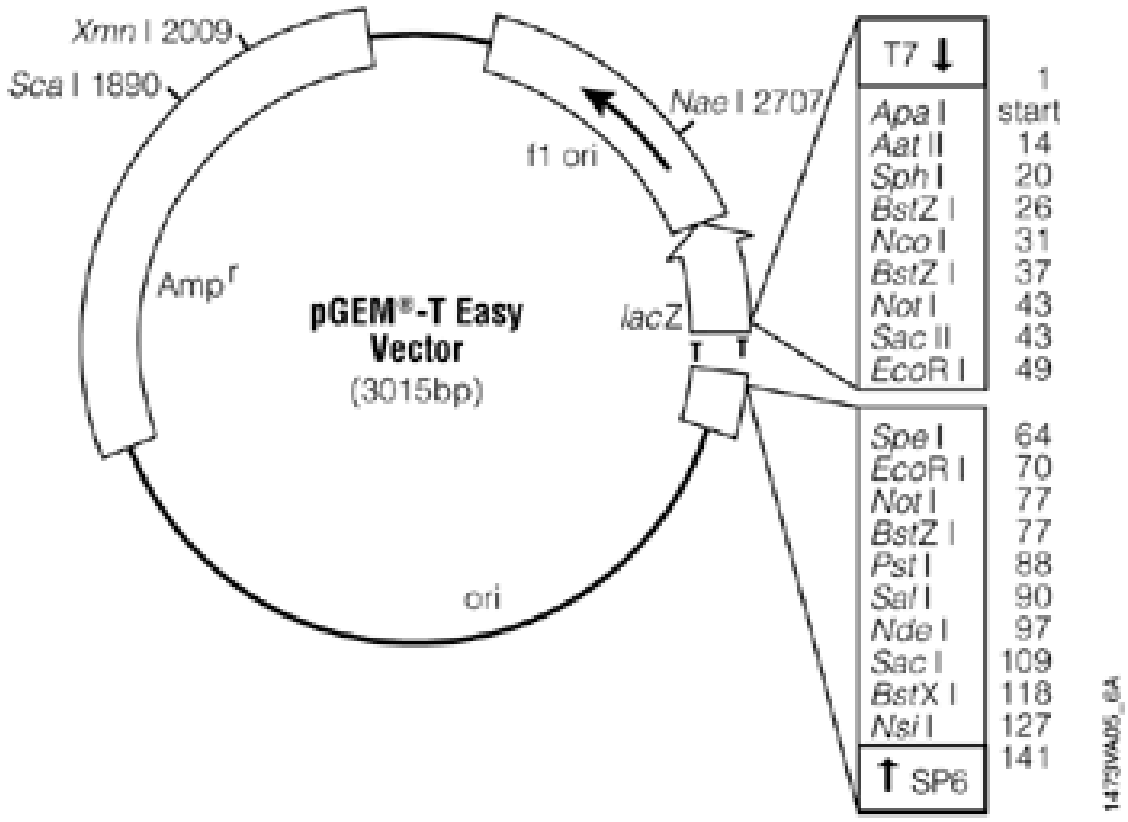


Figure II-1 Map of the pGEM-T-Easy cloning vector.

The pGEM-T-Easy cloning vector (Promega) was used for the cloning of *PDI* and cyclophilin gene fragments amplified by PCR. The presence of extended 3' thymidine termini facilitated the cloning due to the addition of 5' adenines to the PCR products generated with *Taq* DNA polymerase.

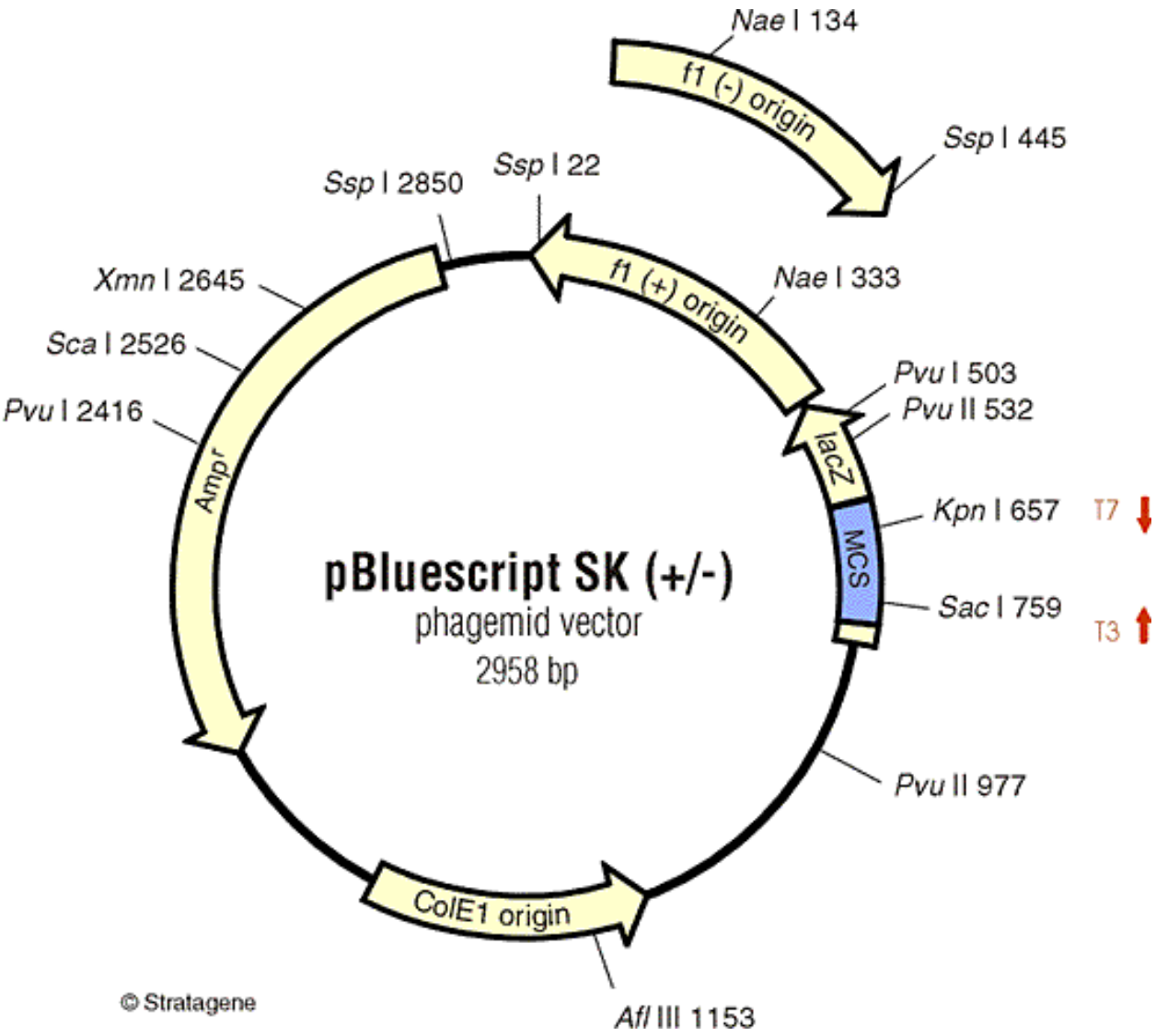
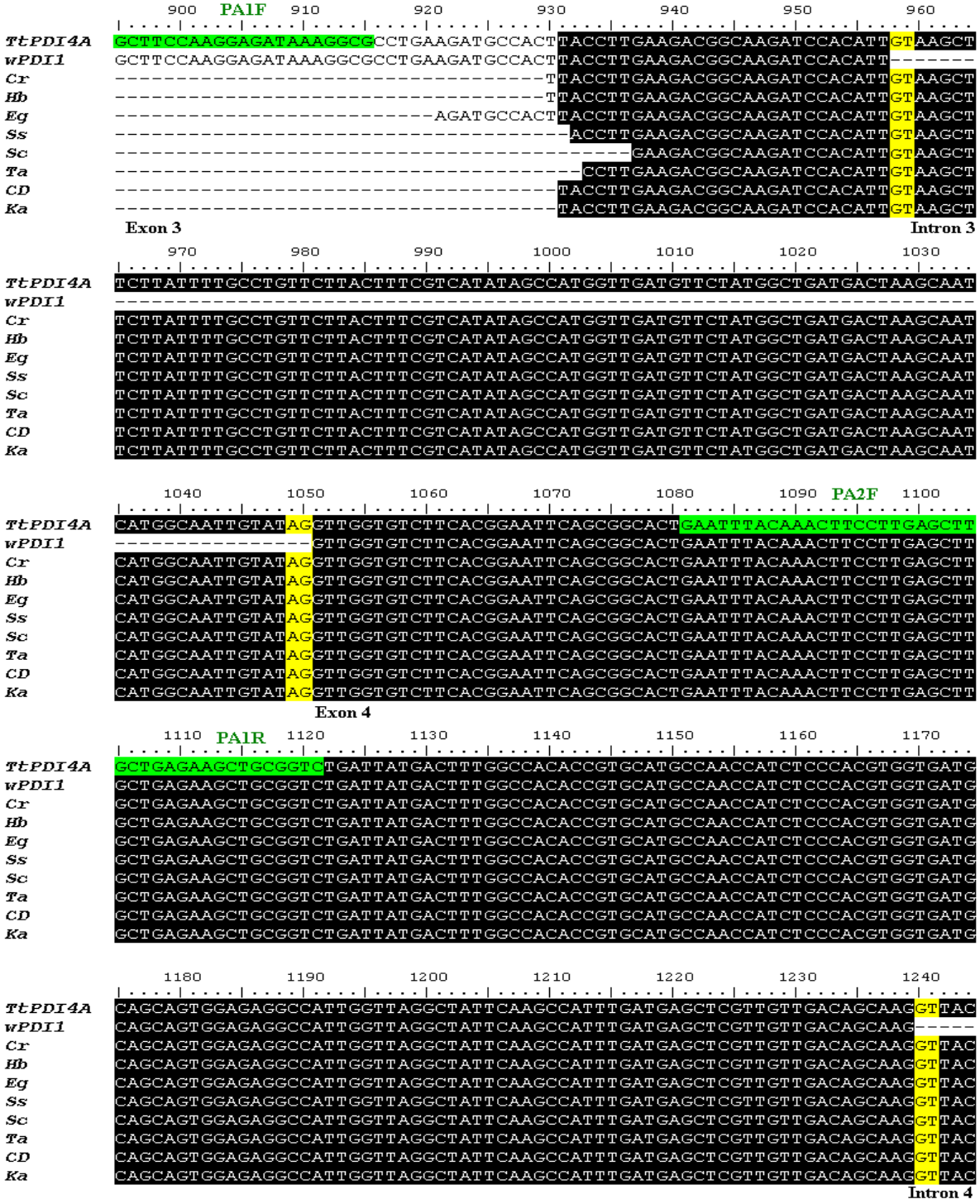


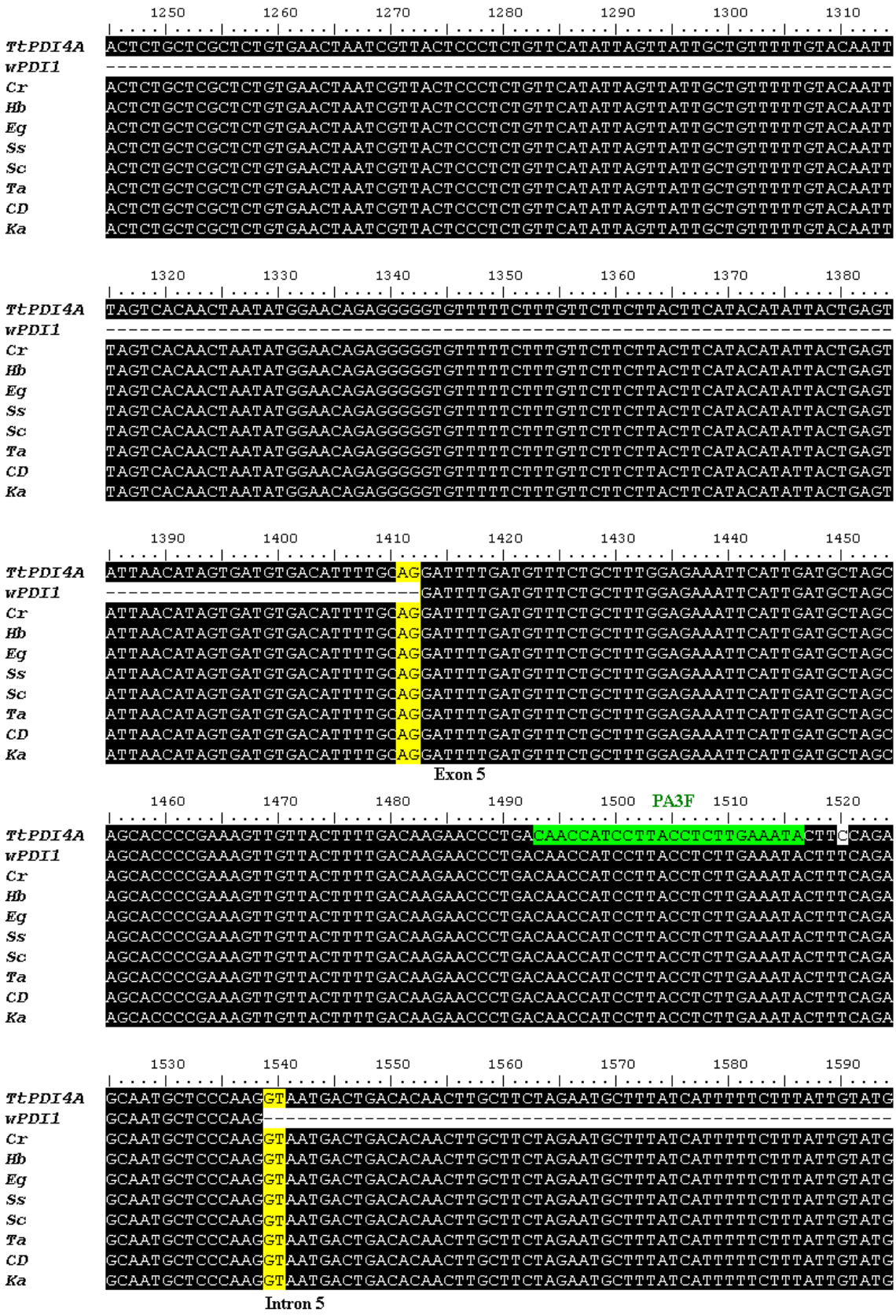
Figure II-2 Map of the pBluescript SK (+/-) cloning vector.

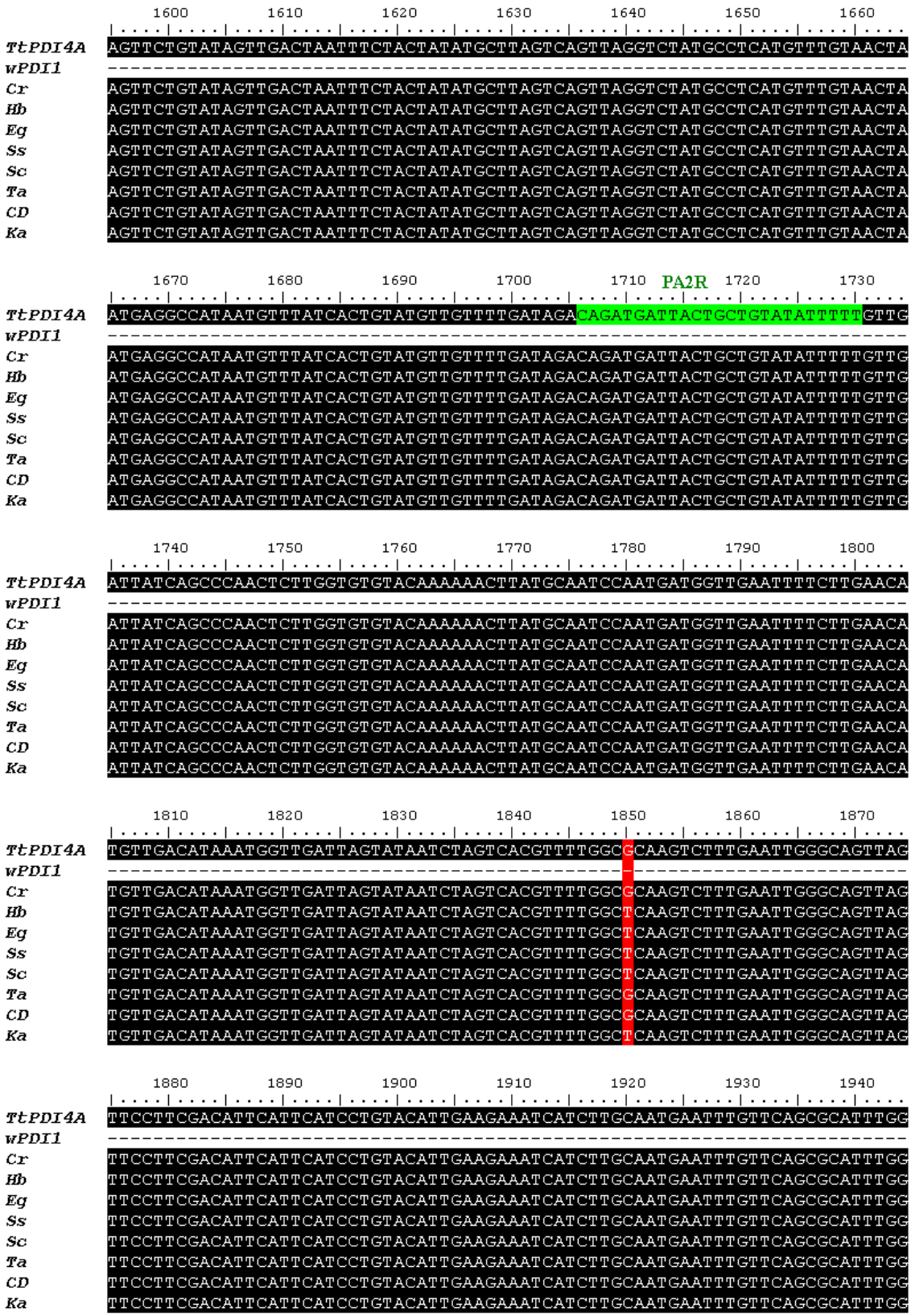
The pBluescript SK (+/-) vector containing putative *PDI* cDNA inserts was *in vivo* excised from the Lambda Zap II vector following screening of a *T. aestivum* 10-12 DPA endosperm library.

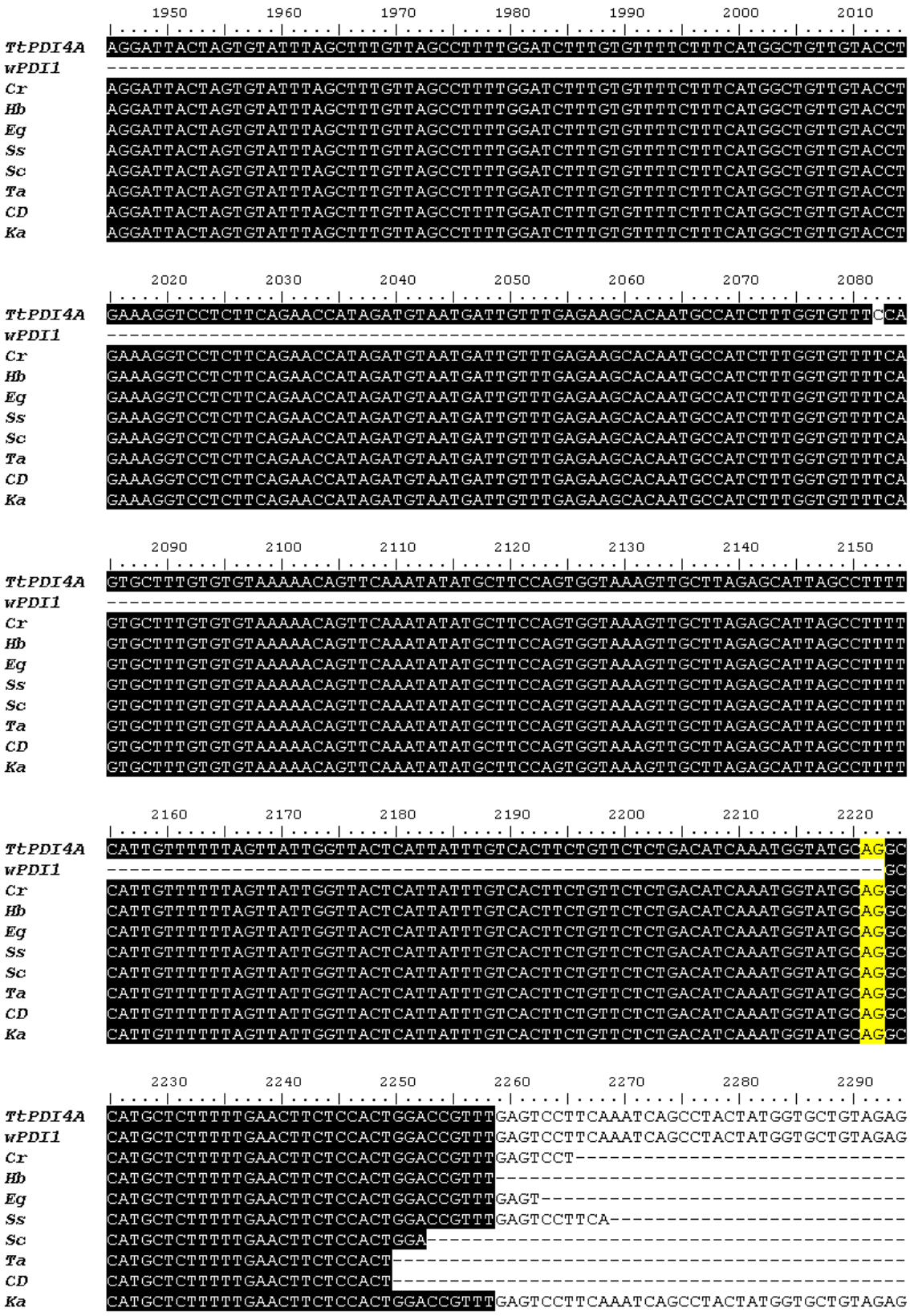
APPENDIX III

THE DNA SEQUENCES OF *PDI* GENES FROM EIGHT CULTIVARS OF *T. AESTIVUM*

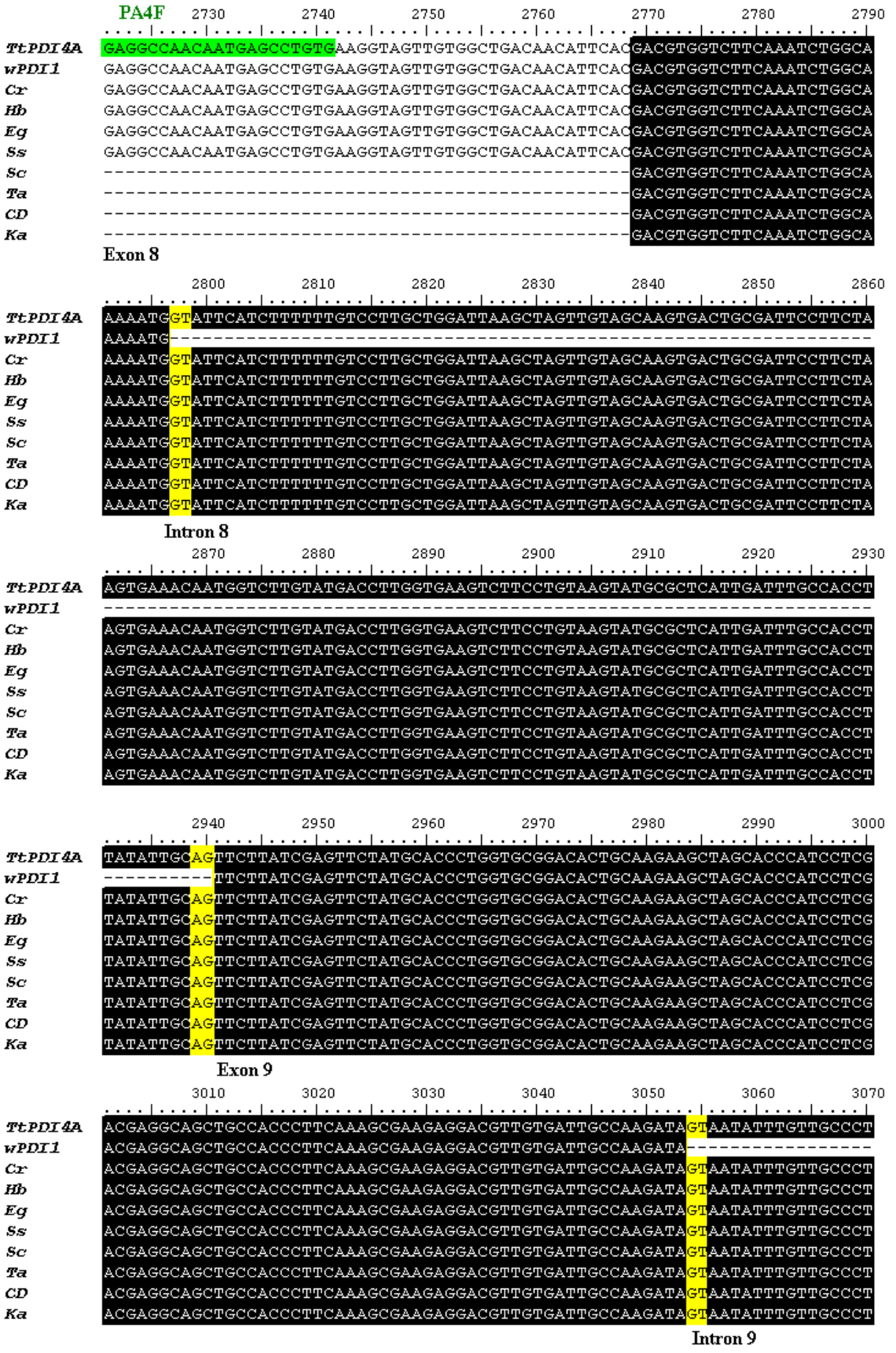



1250 1260 1270 1280 1290 1300 1310
TtPDI4A ACTCTGCTCGCTCTGTGAACTAATCGTTACTCCCTCTGTTCATATTAGTTATTGCTGTTTTTGTACAATT
wPDI1 ----------
Cr ACTCTGCTCGCTCTGTGAACTAATCGTTACTCCCTCTGTTCATATTAGTTATTGCTGTTTTTGTACAATT
Hb ACTCTGCTCGCTCTGTGAACTAATCGTTACTCCCTCTGTTCATATTAGTTATTGCTGTTTTTGTACAATT
Eg ACTCTGCTCGCTCTGTGAACTAATCGTTACTCCCTCTGTTCATATTAGTTATTGCTGTTTTTGTACAATT
Ss ACTCTGCTCGCTCTGTGAACTAATCGTTACTCCCTCTGTTCATATTAGTTATTGCTGTTTTTGTACAATT
Sc ACTCTGCTCGCTCTGTGAACTAATCGTTACTCCCTCTGTTCATATTAGTTATTGCTGTTTTTGTACAATT
Ta ACTCTGCTCGCTCTGTGAACTAATCGTTACTCCCTCTGTTCATATTAGTTATTGCTGTTTTTGTACAATT
CD ACTCTGCTCGCTCTGTGAACTAATCGTTACTCCCTCTGTTCATATTAGTTATTGCTGTTTTTGTACAATT
Ka ACTCTGCTCGCTCTGTGAACTAATCGTTACTCCCTCTGTTCATATTAGTTATTGCTGTTTTTGTACAATT
1320 1330 1340 1350 1360 1370 1380
TtPDI4A TAGTCACAACTAATATGGAACAGAGGGGGTGTTTTTCTTTGTTCTTCTTACTTCATACATATTACTGAGT
wPDI1 ----------
Cr TAGTCACAACTAATATGGAACAGAGGGGGTGTTTTTCTTTGTTCTTCTTACTTCATACATATTACTGAGT
Hb TAGTCACAACTAATATGGAACAGAGGGGGTGTTTTTCTTTGTTCTTCTTACTTCATACATATTACTGAGT
Eg TAGTCACAACTAATATGGAACAGAGGGGGTGTTTTTCTTTGTTCTTCTTACTTCATACATATTACTGAGT
Ss TAGTCACAACTAATATGGAACAGAGGGGGTGTTTTTCTTTGTTCTTCTTACTTCATACATATTACTGAGT
Sc TAGTCACAACTAATATGGAACAGAGGGGGTGTTTTTCTTTGTTCTTCTTACTTCATACATATTACTGAGT
Ta TAGTCACAACTAATATGGAACAGAGGGGGTGTTTTTCTTTGTTCTTCTTACTTCATACATATTACTGAGT
CD TAGTCACAACTAATATGGAACAGAGGGGGTGTTTTTCTTTGTTCTTCTTACTTCATACATATTACTGAGT
Ka TAGTCACAACTAATATGGAACAGAGGGGGTGTTTTTCTTTGTTCTTCTTACTTCATACATATTACTGAGT
1390 1400 1410 1420 1430 1440 1450
TtPDI4A ATTAACATAGTGATGTGACATTTTGCAGGATTTTGATGTTTCTGCTTTGGAGAAATTCATTGATGCTAGC
wPDI1 ----------GATTTTGATGTTTCTGCTTTGGAGAAATTCATTGATGCTAGC
Cr ATTAACATAGTGATGTGACATTTTGCAGGATTTTGATGTTTCTGCTTTGGAGAAATTCATTGATGCTAGC
Hb ATTAACATAGTGATGTGACATTTTGCAGGATTTTGATGTTTCTGCTTTGGAGAAATTCATTGATGCTAGC
Eg ATTAACATAGTGATGTGACATTTTGCAGGATTTTGATGTTTCTGCTTTGGAGAAATTCATTGATGCTAGC
Ss ATTAACATAGTGATGTGACATTTTGCAGGATTTTGATGTTTCTGCTTTGGAGAAATTCATTGATGCTAGC
Sc ATTAACATAGTGATGTGACATTTTGCAGGATTTTGATGTTTCTGCTTTGGAGAAATTCATTGATGCTAGC
Ta ATTAACATAGTGATGTGACATTTTGCAGGATTTTGATGTTTCTGCTTTGGAGAAATTCATTGATGCTAGC
CD ATTAACATAGTGATGTGACATTTTGCAGGATTTTGATGTTTCTGCTTTGGAGAAATTCATTGATGCTAGC
Ka ATTAACATAGTGATGTGACATTTTGCAGGATTTTGATGTTTCTGCTTTGGAGAAATTCATTGATGCTAGC
Exon 5
1460 1470 1480 1490 1500 PA3F 1510 1520
TtPDI4A AGCACCCCGAAAGTTGTTACTTTTGACAAGAACCCTGACAACCATCCTTACCTCTTGAAATACTTCCAGA
wPDI1 AGCACCCCGAAAGTTGTTACTTTTGACAAGAACCCTGACAACCATCCTTACCTCTTGAAATACTTTCAGA
Cr AGCACCCCGAAAGTTGTTACTTTTGACAAGAACCCTGACAACCATCCTTACCTCTTGAAATACTTTCAGA
Hb AGCACCCCGAAAGTTGTTACTTTTGACAAGAACCCTGACAACCATCCTTACCTCTTGAAATACTTTCAGA
Eg AGCACCCCGAAAGTTGTTACTTTTGACAAGAACCCTGACAACCATCCTTACCTCTTGAAATACTTTCAGA
Ss AGCACCCCGAAAGTTGTTACTTTTGACAAGAACCCTGACAACCATCCTTACCTCTTGAAATACTTTCAGA
Sc AGCACCCCGAAAGTTGTTACTTTTGACAAGAACCCTGACAACCATCCTTACCTCTTGAAATACTTTCAGA
Ta AGCACCCCGAAAGTTGTTACTTTTGACAAGAACCCTGACAACCATCCTTACCTCTTGAAATACTTTCAGA
CD AGCACCCCGAAAGTTGTTACTTTTGACAAGAACCCTGACAACCATCCTTACCTCTTGAAATACTTTCAGA
Ka AGCACCCCGAAAGTTGTTACTTTTGACAAGAACCCTGACAACCATCCTTACCTCTTGAAATACTTTCAGA
1530 1540 1550 1560 1570 1580 1590
TtPDI4A GCAATGCTCCCAAGGTAATGACTGACACAACTTGCTTCTAGAATGCTTTATCATTTTTCTTTATTGTATG
wPDI1 GCAATGCTCCCAAG----------
Cr GCAATGCTCCCAAGGTAATGACTGACACAACTTGCTTCTAGAATGCTTTATCATTTTTCTTTATTGTATG
Hb GCAATGCTCCCAAGGTAATGACTGACACAACTTGCTTCTAGAATGCTTTATCATTTTTCTTTATTGTATG
Eg GCAATGCTCCCAAGGTAATGACTGACACAACTTGCTTCTAGAATGCTTTATCATTTTTCTTTATTGTATG
Ss GCAATGCTCCCAAGGTAATGACTGACACAACTTGCTTCTAGAATGCTTTATCATTTTTCTTTATTGTATG
Sc GCAATGCTCCCAAGGTAATGACTGACACAACTTGCTTCTAGAATGCTTTATCATTTTTCTTTATTGTATG
Ta GCAATGCTCCCAAGGTAATGACTGACACAACTTGCTTCTAGAATGCTTTATCATTTTTCTTTATTGTATG
CD GCAATGCTCCCAAGGTAATGACTGACACAACTTGCTTCTAGAATGCTTTATCATTTTTCTTTATTGTATG
Ka GCAATGCTCCCAAGGTAATGACTGACACAACTTGCTTCTAGAATGCTTTATCATTTTTCTTTATTGTATG
Intron 5


1600 1610 1620 1630 1640 1650 1660
TtPDI4A AGTTCTGTATAGTTGACTAATTTCTACTATATGCTTAGTCAGTTAGGTCTATGCCTCATGTTTGTAACTA
wPDI1 ----------
Cr AGTTCTGTATAGTTGACTAATTTCTACTATATGCTTAGTCAGTTAGGTCTATGCCTCATGTTTGTAACTA
Hb AGTTCTGTATAGTTGACTAATTTCTACTATATGCTTAGTCAGTTAGGTCTATGCCTCATGTTTGTAACTA
Eg AGTTCTGTATAGTTGACTAATTTCTACTATATGCTTAGTCAGTTAGGTCTATGCCTCATGTTTGTAACTA
Ss AGTTCTGTATAGTTGACTAATTTCTACTATATGCTTAGTCAGTTAGGTCTATGCCTCATGTTTGTAACTA
Sc AGTTCTGTATAGTTGACTAATTTCTACTATATGCTTAGTCAGTTAGGTCTATGCCTCATGTTTGTAACTA
Ta AGTTCTGTATAGTTGACTAATTTCTACTATATGCTTAGTCAGTTAGGTCTATGCCTCATGTTTGTAACTA
CD AGTTCTGTATAGTTGACTAATTTCTACTATATGCTTAGTCAGTTAGGTCTATGCCTCATGTTTGTAACTA
Ka AGTTCTGTATAGTTGACTAATTTCTACTATATGCTTAGTCAGTTAGGTCTATGCCTCATGTTTGTAACTA
1670 1680 1690 1700 1710 PA2R 1720 1730
TtPDI4A ATGAGGCCATAATGTTTATCACTGTATGTTGTTTTGATAGACAGATGATTACTGCTGTATATTTTTGTTG
wPDI1 ----------
Cr ATGAGGCCATAATGTTTATCACTGTATGTTGTTTTGATAGACAGATGATTACTGCTGTATATTTTTGTTG
Hb ATGAGGCCATAATGTTTATCACTGTATGTTGTTTTGATAGACAGATGATTACTGCTGTATATTTTTGTTG
Eg ATGAGGCCATAATGTTTATCACTGTATGTTGTTTTGATAGACAGATGATTACTGCTGTATATTTTTGTTG
Ss ATGAGGCCATAATGTTTATCACTGTATGTTGTTTTGATAGACAGATGATTACTGCTGTATATTTTTGTTG
Sc ATGAGGCCATAATGTTTATCACTGTATGTTGTTTTGATAGACAGATGATTACTGCTGTATATTTTTGTTG
Ta ATGAGGCCATAATGTTTATCACTGTATGTTGTTTTGATAGACAGATGATTACTGCTGTATATTTTTGTTG
CD ATGAGGCCATAATGTTTATCACTGTATGTTGTTTTGATAGACAGATGATTACTGCTGTATATTTTTGTTG
Ka ATGAGGCCATAATGTTTATCACTGTATGTTGTTTTGATAGACAGATGATTACTGCTGTATATTTTTGTTG
1740 1750 1760 1770 1780 1790 1800
TtPDI4A ATTATCAGCCCAACTCTTGGTGTGTACAAAAAACTTATGCAATCCAATGATGGTTGAATTTTCTTGAACA
wPDI1 ----------
Cr ATTATCAGCCCAACTCTTGGTGTGTACAAAAAACTTATGCAATCCAATGATGGTTGAATTTTCTTGAACA
Hb ATTATCAGCCCAACTCTTGGTGTGTACAAAAAACTTATGCAATCCAATGATGGTTGAATTTTCTTGAACA
Eg ATTATCAGCCCAACTCTTGGTGTGTACAAAAAACTTATGCAATCCAATGATGGTTGAATTTTCTTGAACA
Ss ATTATCAGCCCAACTCTTGGTGTGTACAAAAAACTTATGCAATCCAATGATGGTTGAATTTTCTTGAACA
Sc ATTATCAGCCCAACTCTTGGTGTGTACAAAAAACTTATGCAATCCAATGATGGTTGAATTTTCTTGAACA
Ta ATTATCAGCCCAACTCTTGGTGTGTACAAAAAACTTATGCAATCCAATGATGGTTGAATTTTCTTGAACA
CD ATTATCAGCCCAACTCTTGGTGTGTACAAAAAACTTATGCAATCCAATGATGGTTGAATTTTCTTGAACA
Ka ATTATCAGCCCAACTCTTGGTGTGTACAAAAAACTTATGCAATCCAATGATGGTTGAATTTTCTTGAACA
1810 1820 1830 1840 1850 1860 1870
TtPDI4A TGTTGACATAAATGGTTGATTAGTATAATCTAGTCACGTTTTGGCGCAAGTCTTTGAATTGGGCAGTTAG
wPDI1 ----------
Cr TGTTGACATAAATGGTTGATTAGTATAATCTAGTCACGTTTTGGCGCAAGTCTTTGAATTGGGCAGTTAG
Hb TGTTGACATAAATGGTTGATTAGTATAATCTAGTCACGTTTTGGCTCAAGTCTTTGAATTGGGCAGTTAG
Eg TGTTGACATAAATGGTTGATTAGTATAATCTAGTCACGTTTTGGCTCAAGTCTTTGAATTGGGCAGTTAG
Ss TGTTGACATAAATGGTTGATTAGTATAATCTAGTCACGTTTTGGCTCAAGTCTTTGAATTGGGCAGTTAG
Sc TGTTGACATAAATGGTTGATTAGTATAATCTAGTCACGTTTTGGCTCAAGTCTTTGAATTGGGCAGTTAG
Ta TGTTGACATAAATGGTTGATTAGTATAATCTAGTCACGTTTTGGCGCAAGTCTTTGAATTGGGCAGTTAG
CD TGTTGACATAAATGGTTGATTAGTATAATCTAGTCACGTTTTGGCGCAAGTCTTTGAATTGGGCAGTTAG
Ka TGTTGACATAAATGGTTGATTAGTATAATCTAGTCACGTTTTGGCTCAAGTCTTTGAATTGGGCAGTTAG
1880 1890 1900 1910 1920 1930 1940
TtPDI4A TTCCTTCGACATTCATTCATCCTGTACATTGAAGAAATCATCTTGCAATGAATTTGTTCAGCGCATTTGG
wPDI1 ----------
Cr TTCCTTCGACATTCATTCATCCTGTACATTGAAGAAATCATCTTGCAATGAATTTGTTCAGCGCATTTGG
Hb TTCCTTCGACATTCATTCATCCTGTACATTGAAGAAATCATCTTGCAATGAATTTGTTCAGCGCATTTGG
Eg TTCCTTCGACATTCATTCATCCTGTACATTGAAGAAATCATCTTGCAATGAATTTGTTCAGCGCATTTGG
Ss TTCCTTCGACATTCATTCATCCTGTACATTGAAGAAATCATCTTGCAATGAATTTGTTCAGCGCATTTGG
Sc TTCCTTCGACATTCATTCATCCTGTACATTGAAGAAATCATCTTGCAATGAATTTGTTCAGCGCATTTGG
Ta TTCCTTCGACATTCATTCATCCTGTACATTGAAGAAATCATCTTGCAATGAATTTGTTCAGCGCATTTGG
CD TTCCTTCGACATTCATTCATCCTGTACATTGAAGAAATCATCTTGCAATGAATTTGTTCAGCGCATTTGG
Ka TTCCTTCGACATTCATTCATCCTGTACATTGAAGAAATCATCTTGCAATGAATTTGTTCAGCGCATTTGG


1950 1960 1970 1980 1990 2000 2010
TtPDI4A AGGATTACTAGTGTATTTAGCTTTGTTAGCCTTTTGGATCTTTGTGTTTTCTTTCATGGCTGTTGTACCT
wPDI1 ----------
Cr AGGATTACTAGTGTATTTAGCTTTGTTAGCCTTTTGGATCTTTGTGTTTTCTTTCATGGCTGTTGTACCT
Hb AGGATTACTAGTGTATTTAGCTTTGTTAGCCTTTTGGATCTTTGTGTTTTCTTTCATGGCTGTTGTACCT
Eg AGGATTACTAGTGTATTTAGCTTTGTTAGCCTTTTGGATCTTTGTGTTTTCTTTCATGGCTGTTGTACCT
Ss AGGATTACTAGTGTATTTAGCTTTGTTAGCCTTTTGGATCTTTGTGTTTTCTTTCATGGCTGTTGTACCT
Sc AGGATTACTAGTGTATTTAGCTTTGTTAGCCTTTTGGATCTTTGTGTTTTCTTTCATGGCTGTTGTACCT
Ta AGGATTACTAGTGTATTTAGCTTTGTTAGCCTTTTGGATCTTTGTGTTTTCTTTCATGGCTGTTGTACCT
CD AGGATTACTAGTGTATTTAGCTTTGTTAGCCTTTTGGATCTTTGTGTTTTCTTTCATGGCTGTTGTACCT
Ka AGGATTACTAGTGTATTTAGCTTTGTTAGCCTTTTGGATCTTTGTGTTTTCTTTCATGGCTGTTGTACCT
2020 2030 2040 2050 2060 2070 2080
TtPDI4A GAAAGGTCCTCTTCAGAACCATAGATGTAATGATTGTTTGAGAAGCACAATGCCATCTTTGGTGTTTCCA
wPDI1 ----------
Cr GAAAGGTCCTCTTCAGAACCATAGATGTAATGATTGTTTGAGAAGCACAATGCCATCTTTGGTGTTTTCA
Hb GAAAGGTCCTCTTCAGAACCATAGATGTAATGATTGTTTGAGAAGCACAATGCCATCTTTGGTGTTTTCA
Eg GAAAGGTCCTCTTCAGAACCATAGATGTAATGATTGTTTGAGAAGCACAATGCCATCTTTGGTGTTTTCA
Ss GAAAGGTCCTCTTCAGAACCATAGATGTAATGATTGTTTGAGAAGCACAATGCCATCTTTGGTGTTTTCA
Sc GAAAGGTCCTCTTCAGAACCATAGATGTAATGATTGTTTGAGAAGCACAATGCCATCTTTGGTGTTTTCA
Ta GAAAGGTCCTCTTCAGAACCATAGATGTAATGATTGTTTGAGAAGCACAATGCCATCTTTGGTGTTTTCA
CD GAAAGGTCCTCTTCAGAACCATAGATGTAATGATTGTTTGAGAAGCACAATGCCATCTTTGGTGTTTTCA
Ka GAAAGGTCCTCTTCAGAACCATAGATGTAATGATTGTTTGAGAAGCACAATGCCATCTTTGGTGTTTTCA
2090 2100 2110 2120 2130 2140 2150
TtPDI4A GTGCTTTGTGTGTAAAAACAGTTCAAATATATGCTTCCAGTGGTAAAGTTGCTTAGAGCATTAGCCTTTT
wPDI1 ----------
Cr GTGCTTTGTGTGTAAAAACAGTTCAAATATATGCTTCCAGTGGTAAAGTTGCTTAGAGCATTAGCCTTTT
Hb GTGCTTTGTGTGTAAAAACAGTTCAAATATATGCTTCCAGTGGTAAAGTTGCTTAGAGCATTAGCCTTTT
Eg GTGCTTTGTGTGTAAAAACAGTTCAAATATATGCTTCCAGTGGTAAAGTTGCTTAGAGCATTAGCCTTTT
Ss GTGCTTTGTGTGTAAAAACAGTTCAAATATATGCTTCCAGTGGTAAAGTTGCTTAGAGCATTAGCCTTTT
Sc GTGCTTTGTGTGTAAAAACAGTTCAAATATATGCTTCCAGTGGTAAAGTTGCTTAGAGCATTAGCCTTTT
Ta GTGCTTTGTGTGTAAAAACAGTTCAAATATATGCTTCCAGTGGTAAAGTTGCTTAGAGCATTAGCCTTTT
CD GTGCTTTGTGTGTAAAAACAGTTCAAATATATGCTTCCAGTGGTAAAGTTGCTTAGAGCATTAGCCTTTT
Ka GTGCTTTGTGTGTAAAAACAGTTCAAATATATGCTTCCAGTGGTAAAGTTGCTTAGAGCATTAGCCTTTT
2160 2170 2180 2190 2200 2210 2220
TtPDI4A CATTGTTTTTTAGTTATTGGTTACTCATTATTTGTCACTTCTGTTCTCTGACATCAAATGGTATGCAGGC
wPDI1 ----------GC
Cr CATTGTTTTTTAGTTATTGGTTACTCATTATTTGTCACTTCTGTTCTCTGACATCAAATGGTATGCAGGC
Hb CATTGTTTTTTAGTTATTGGTTACTCATTATTTGTCACTTCTGTTCTCTGACATCAAATGGTATGCAGGC
Eg CATTGTTTTTTAGTTATTGGTTACTCATTATTTGTCACTTCTGTTCTCTGACATCAAATGGTATGCAGGC
Ss CATTGTTTTTTAGTTATTGGTTACTCATTATTTGTCACTTCTGTTCTCTGACATCAAATGGTATGCAGGC
Sc CATTGTTTTTTAGTTATTGGTTACTCATTATTTGTCACTTCTGTTCTCTGACATCAAATGGTATGCAGGC
Ta CATTGTTTTTTAGTTATTGGTTACTCATTATTTGTCACTTCTGTTCTCTGACATCAAATGGTATGCAGGC
CD CATTGTTTTTTAGTTATTGGTTACTCATTATTTGTCACTTCTGTTCTCTGACATCAAATGGTATGCAGGC
Ka CATTGTTTTTTAGTTATTGGTTACTCATTATTTGTCACTTCTGTTCTCTGACATCAAATGGTATGCAGGC
2230 2240 2250 2260 2270 2280 2290
TtPDI4A CATGCTCTTTTTGAACTTCTCCACTGGACCGTTTGAGTCCTTCAAATCAGCCTACTATGGTGCTGTAGAG
wPDI1 CATGCTCTTTTTGAACTTCTCCACTGGACCGTTTGAGTCCTTCAAATCAGCCTACTATGGTGCTGTAGAG
Cr CATGCTCTTTTTGAACTTCTCCACTGGACCGTTTGAGTCCT----------
Hb CATGCTCTTTTTGAACTTCTCCACTGGACCGTTT----------
Eg CATGCTCTTTTTGAACTTCTCCACTGGACCGTTTGAGT----------
Ss CATGCTCTTTTTGAACTTCTCCACTGGACCGTTTGAGTCCTTCA----------
Sc CATGCTCTTTTTGAACTTCTCCACTGGA----------
Ta CATGCTCTTTTTGAACTTCTCCACT----------
CD CATGCTCTTTTTGAACTTCTCCACT----------
Ka CATGCTCTTTTTGAACTTCTCCACTGGACCGTTTGAGTCCTTCAAATCAGCCTACTATGGTGCTGTAGAG

Exon 6


PA4F
2730
2740
2750
2760
2770
2780
2790
TtPDI4A GAGGCCAACAATGAGCCTGTGAAGGTAGTTGTGGCTGACAACATTCACGACGTGGTCTTCAAATCTGGCA
wPDI1 GAGGCCAACAATGAGCCTGTGAAGGTAGTTGTGGCTGACAACATTCACGACGTGGTCTTCAAATCTGGCA
Cr GAGGCCAACAATGAGCCTGTGAAGGTAGTTGTGGCTGACAACATTCACGACGTGGTCTTCAAATCTGGCA
Hb GAGGCCAACAATGAGCCTGTGAAGGTAGTTGTGGCTGACAACATTCACGACGTGGTCTTCAAATCTGGCA
Eg GAGGCCAACAATGAGCCTGTGAAGGTAGTTGTGGCTGACAACATTCACGACGTGGTCTTCAAATCTGGCA
Ss GAGGCCAACAATGAGCCTGTGAAGGTAGTTGTGGCTGACAACATTCACGACGTGGTCTTCAAATCTGGCA
Sc ----------GACGTGGTCTTCAAATCTGGCA
Ta ----------GACGTGGTCTTCAAATCTGGCA
CD ----------GACGTGGTCTTCAAATCTGGCA
Ka ----------GACGTGGTCTTCAAATCTGGCA
Exon 8
2800
2810
2820
2830
2840
2850
2860
TtPDI4A AAAATGGTATTCATCTTTTTTGTCCTTGCTGGATTAAGCTAGTTGTAGCAAGTGACTGCGATTCCTTCTA
wPDI1 AAAATG----------
Cr AAAATGGTATTCATCTTTTTTGTCCTTGCTGGATTAAGCTAGTTGTAGCAAGTGACTGCGATTCCTTCTA
Hb AAAATGGTATTCATCTTTTTTGTCCTTGCTGGATTAAGCTAGTTGTAGCAAGTGACTGCGATTCCTTCTA
Eg AAAATGGTATTCATCTTTTTTGTCCTTGCTGGATTAAGCTAGTTGTAGCAAGTGACTGCGATTCCTTCTA
Ss AAAATGGTATTCATCTTTTTTGTCCTTGCTGGATTAAGCTAGTTGTAGCAAGTGACTGCGATTCCTTCTA
Sc AAAATGGTATTCATCTTTTTTGTCCTTGCTGGATTAAGCTAGTTGTAGCAAGTGACTGCGATTCCTTCTA
Ta AAAATGGTATTCATCTTTTTTGTCCTTGCTGGATTAAGCTAGTTGTAGCAAGTGACTGCGATTCCTTCTA
CD AAAATGGTATTCATCTTTTTTGTCCTTGCTGGATTAAGCTAGTTGTAGCAAGTGACTGCGATTCCTTCTA
Ka AAAATGGTATTCATCTTTTTTGTCCTTGCTGGATTAAGCTAGTTGTAGCAAGTGACTGCGATTCCTTCTA
Intron 8
2870
2880
2890
2900
2910
2920
2930
TtPDI4A AGTGAAACAATGGTCTTGTATGACCTTGGTGAAGTCTTCCTGTAAGTATGCGCTCATTGATTTGCCACCT
wPDI1 ----------
Cr AGTGAAACAATGGTCTTGTATGACCTTGGTGAAGTCTTCCTGTAAGTATGCGCTCATTGATTTGCCACCT
Hb AGTGAAACAATGGTCTTGTATGACCTTGGTGAAGTCTTCCTGTAAGTATGCGCTCATTGATTTGCCACCT
Eg AGTGAAACAATGGTCTTGTATGACCTTGGTGAAGTCTTCCTGTAAGTATGCGCTCATTGATTTGCCACCT
Ss AGTGAAACAATGGTCTTGTATGACCTTGGTGAAGTCTTCCTGTAAGTATGCGCTCATTGATTTGCCACCT
Sc AGTGAAACAATGGTCTTGTATGACCTTGGTGAAGTCTTCCTGTAAGTATGCGCTCATTGATTTGCCACCT
Ta AGTGAAACAATGGTCTTGTATGACCTTGGTGAAGTCTTCCTGTAAGTATGCGCTCATTGATTTGCCACCT
CD AGTGAAACAATGGTCTTGTATGACCTTGGTGAAGTCTTCCTGTAAGTATGCGCTCATTGATTTGCCACCT
Ka AGTGAAACAATGGTCTTGTATGACCTTGGTGAAGTCTTCCTGTAAGTATGCGCTCATTGATTTGCCACCT
2940
2950
2960
2970
2980
2990
3000
TtPDI4A TATATTGCAGTTCTTATCGAGTTCTATGCACCCTGGTGCGGACACTGCAAGAAGCTAGCACCCATCCTCG
wPDI1 ----------TTCTTATCGAGTTCTATGCACCCTGGTGCGGACACTGCAAGAAGCTAGCACCCATCCTCG
Cr TATATTGCAGTTCTTATCGAGTTCTATGCACCCTGGTGCGGACACTGCAAGAAGCTAGCACCCATCCTCG
Hb TATATTGCAGTTCTTATCGAGTTCTATGCACCCTGGTGCGGACACTGCAAGAAGCTAGCACCCATCCTCG
Eg TATATTGCAGTTCTTATCGAGTTCTATGCACCCTGGTGCGGACACTGCAAGAAGCTAGCACCCATCCTCG
Ss TATATTGCAGTTCTTATCGAGTTCTATGCACCCTGGTGCGGACACTGCAAGAAGCTAGCACCCATCCTCG
Sc TATATTGCAGTTCTTATCGAGTTCTATGCACCCTGGTGCGGACACTGCAAGAAGCTAGCACCCATCCTCG
Ta TATATTGCAGTTCTTATCGAGTTCTATGCACCCTGGTGCGGACACTGCAAGAAGCTAGCACCCATCCTCG
CD TATATTGCAGTTCTTATCGAGTTCTATGCACCCTGGTGCGGACACTGCAAGAAGCTAGCACCCATCCTCG
Ka TATATTGCAGTTCTTATCGAGTTCTATGCACCCTGGTGCGGACACTGCAAGAAGCTAGCACCCATCCTCG
Exon 9
3010
3020
3030
3040
3050
3060
3070
TtPDI4A ACGAGGCAGCTGCCACCCTTCAAAGCGAAGAGGACGTTGTGATTGCCAAGATAGTAATATTTGTTGCCCT
wPDI1 ACGAGGCAGCTGCCACCCTTCAAAGCGAAGAGGACGTTGTGATTGCCAAGATA----------
Cr ACGAGGCAGCTGCCACCCTTCAAAGCGAAGAGGACGTTGTGATTGCCAAGATAGTAATATTTGTTGCCCT
Hb ACGAGGCAGCTGCCACCCTTCAAAGCGAAGAGGACGTTGTGATTGCCAAGATAGTAATATTTGTTGCCCT
Eg ACGAGGCAGCTGCCACCCTTCAAAGCGAAGAGGACGTTGTGATTGCCAAGATAGTAATATTTGTTGCCCT
Ss ACGAGGCAGCTGCCACCCTTCAAAGCGAAGAGGACGTTGTGATTGCCAAGATAGTAATATTTGTTGCCCT
Sc ACGAGGCAGCTGCCACCCTTCAAAGCGAAGAGGACGTTGTGATTGCCAAGATAGTAATATTTGTTGCCCT
Ta ACGAGGCAGCTGCCACCCTTCAAAGCGAAGAGGACGTTGTGATTGCCAAGATAGTAATATTTGTTGCCCT
CD ACGAGGCAGCTGCCACCCTTCAAAGCGAAGAGGACGTTGTGATTGCCAAGATAGTAATATTTGTTGCCCT
Ka ACGAGGCAGCTGCCACCCTTCAAAGCGAAGAGGACGTTGTGATTGCCAAGATAGTAATATTTGTTGCCCT
Intron 9

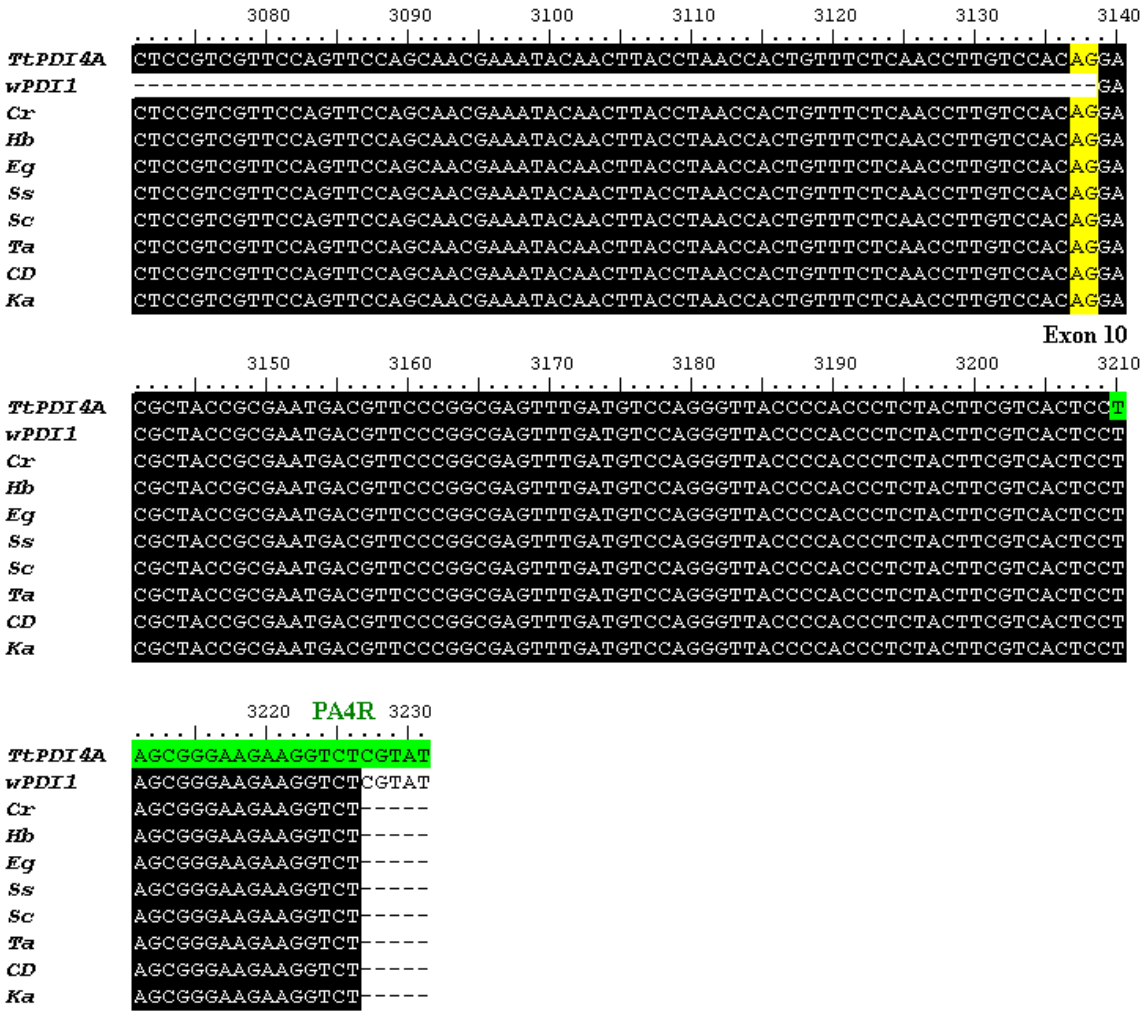
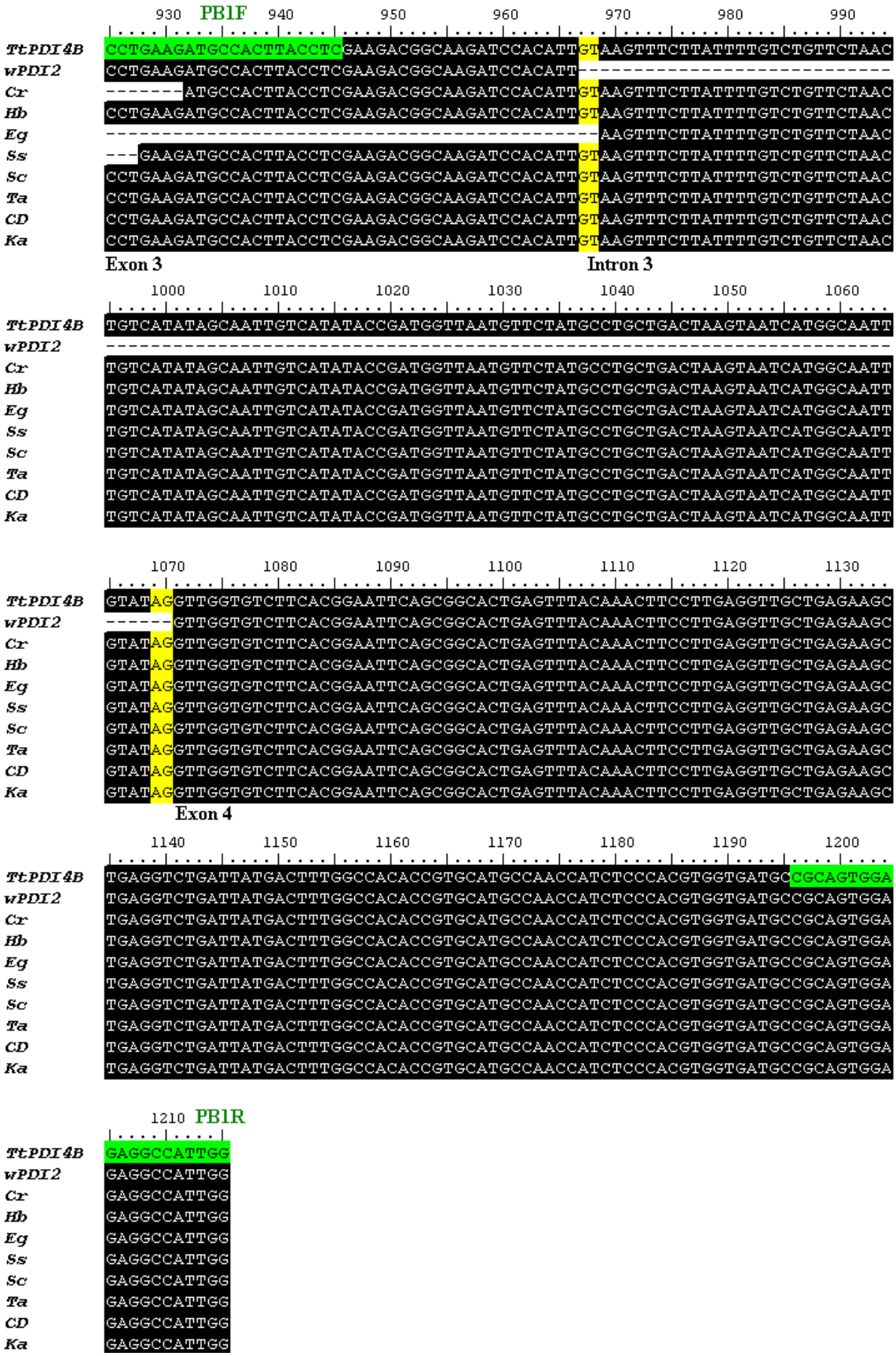
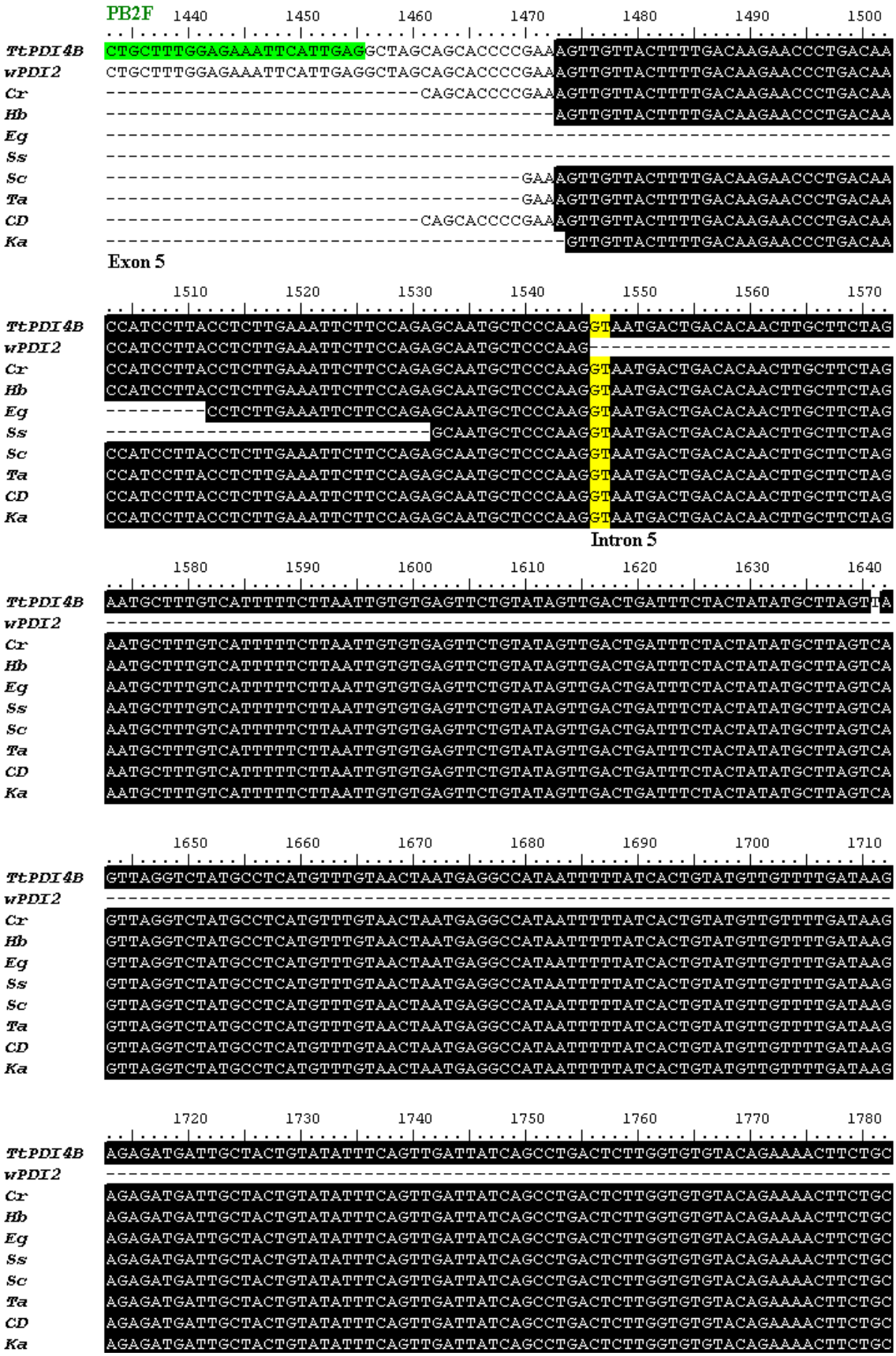
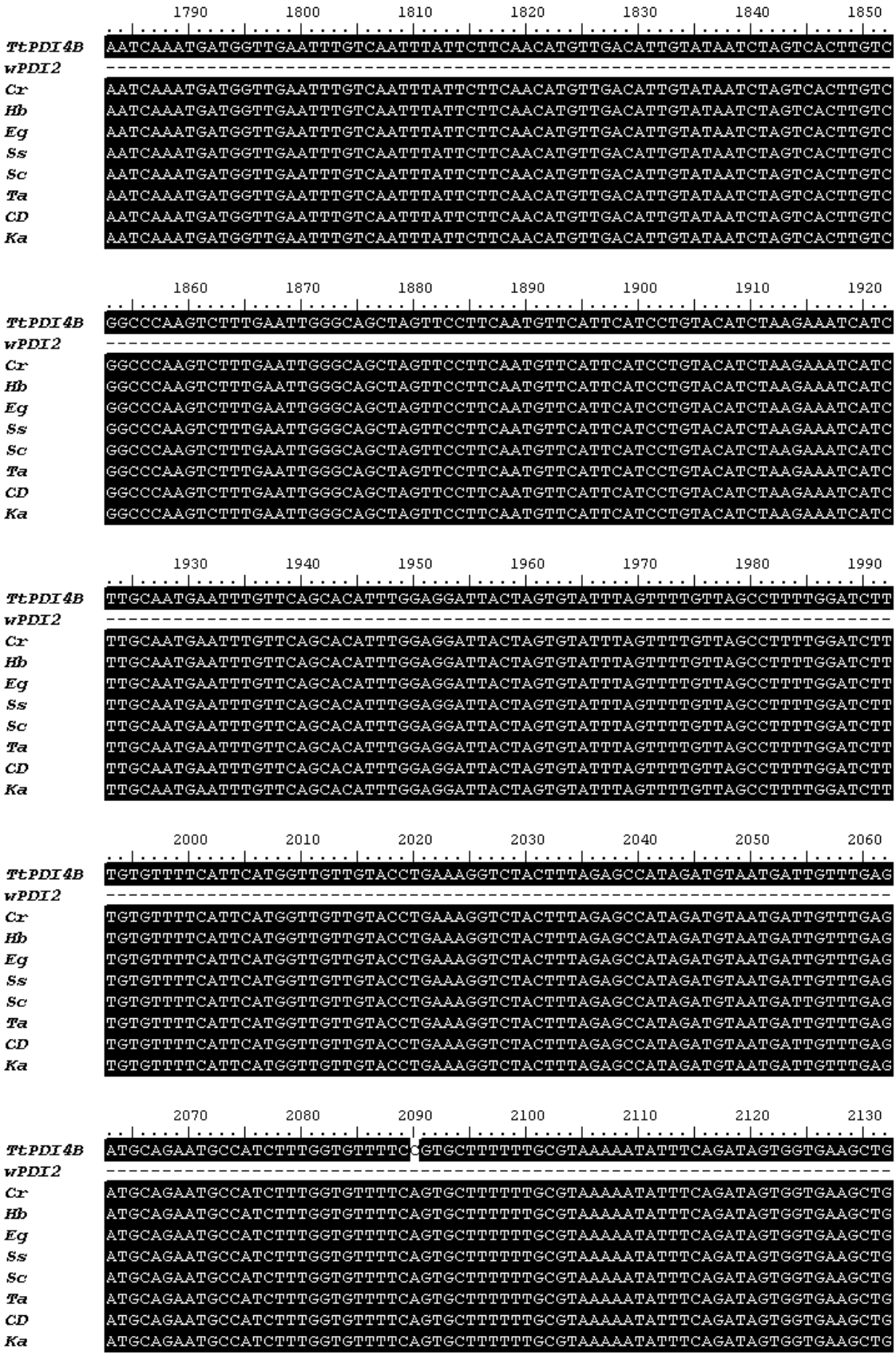
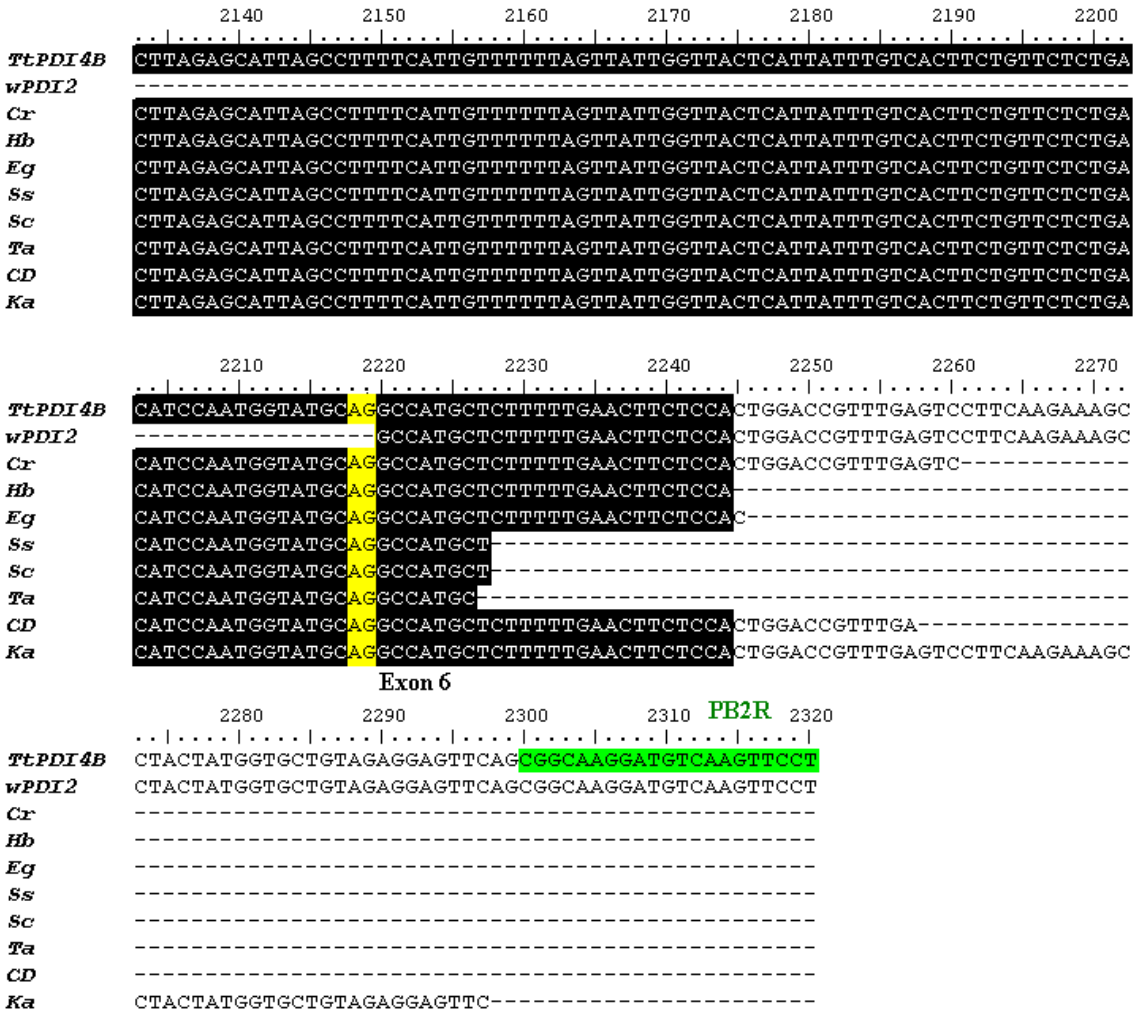


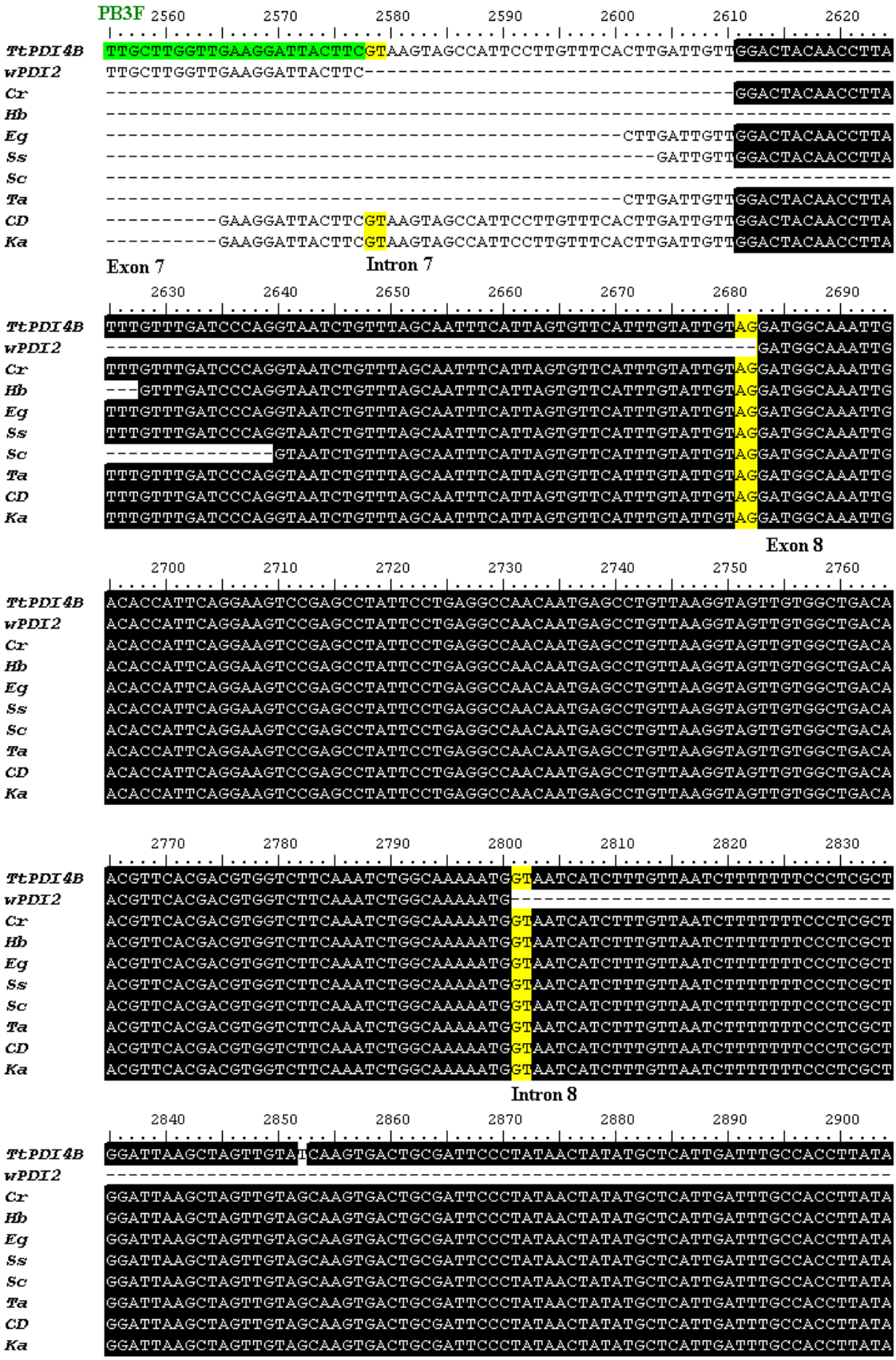
Figure IIIA The partial sequence of the *PDI* genes in the A genome of wheat


930 PB1F 940 950 960 970 980 990
TtPDI4B CCTGAAGATGCCACTTACCTCGAAGACGGCAAGATCCACATTGTAAGTTTCTTATTTTGTCTGTTCTAAC
wPDI2 CCTGAAGATGCCACTTACCTCGAAGACGGCAAGATCCACATT----------
Cr -------ATGCCACTTACCTCGAAGACGGCAAGATCCACATTGTAAGTTTCTTATTTTGTCTGTTCTAAC
Hb CCTGAAGATGCCACTTACCTCGAAGACGGCAAGATCCACATTGTAAGTTTCTTATTTTGTCTGTTCTAAC
Eg ----------AAGTTTCTTATTTTGTCTGTTCTAAC
Ss ---GAAGATGCCACTTACCTCGAAGACGGCAAGATCCACATTGTAAGTTTCTTATTTTGTCTGTTCTAAC
Sc CCTGAAGATGCCACTTACCTCGAAGACGGCAAGATCCACATTGTAAGTTTCTTATTTTGTCTGTTCTAAC
Ta CCTGAAGATGCCACTTACCTCGAAGACGGCAAGATCCACATTGTAAGTTTCTTATTTTGTCTGTTCTAAC
CD CCTGAAGATGCCACTTACCTCGAAGACGGCAAGATCCACATTGTAAGTTTCTTATTTTGTCTGTTCTAAC
Ka CCTGAAGATGCCACTTACCTCGAAGACGGCAAGATCCACATTGTAAGTTTCTTATTTTGTCTGTTCTAAC
Exon 3
Intron 3
1000 1010 1020 1030 1040 1050 1060
TtPDI4B TGTCATATAGCAATTGTCATATACCGATGGTTAATGTTCTATGCCTGCTGACTAAGTAATCATGGCAATT
wPDI2 ----------
Cr TGTCATATAGCAATTGTCATATACCGATGGTTAATGTTCTATGCCTGCTGACTAAGTAATCATGGCAATT
Hb TGTCATATAGCAATTGTCATATACCGATGGTTAATGTTCTATGCCTGCTGACTAAGTAATCATGGCAATT
Eg TGTCATATAGCAATTGTCATATACCGATGGTTAATGTTCTATGCCTGCTGACTAAGTAATCATGGCAATT
Ss TGTCATATAGCAATTGTCATATACCGATGGTTAATGTTCTATGCCTGCTGACTAAGTAATCATGGCAATT
Sc TGTCATATAGCAATTGTCATATACCGATGGTTAATGTTCTATGCCTGCTGACTAAGTAATCATGGCAATT
Ta TGTCATATAGCAATTGTCATATACCGATGGTTAATGTTCTATGCCTGCTGACTAAGTAATCATGGCAATT
CD TGTCATATAGCAATTGTCATATACCGATGGTTAATGTTCTATGCCTGCTGACTAAGTAATCATGGCAATT
Ka TGTCATATAGCAATTGTCATATACCGATGGTTAATGTTCTATGCCTGCTGACTAAGTAATCATGGCAATT
1070 1080 1090 1100 1110 1120 1130
TtPDI4B GTATAGGTTGGTGTCTTCACGGAATTCAGCGGCACTGAGTTTACAAACTTCCTTGAGGTTGCTGAGAAGC
wPDI2 ------GTTGGTGTCTTCACGGAATTCAGCGGCACTGAGTTTACAAACTTCCTTGAGGTTGCTGAGAAGC
Cr GTATAGGTTGGTGTCTTCACGGAATTCAGCGGCACTGAGTTTACAAACTTCCTTGAGGTTGCTGAGAAGC
Hb GTATAGGTTGGTGTCTTCACGGAATTCAGCGGCACTGAGTTTACAAACTTCCTTGAGGTTGCTGAGAAGC
Eg GTATAGGTTGGTGTCTTCACGGAATTCAGCGGCACTGAGTTTACAAACTTCCTTGAGGTTGCTGAGAAGC
Ss GTATAGGTTGGTGTCTTCACGGAATTCAGCGGCACTGAGTTTACAAACTTCCTTGAGGTTGCTGAGAAGC
Sc GTATAGGTTGGTGTCTTCACGGAATTCAGCGGCACTGAGTTTACAAACTTCCTTGAGGTTGCTGAGAAGC
Ta GTATAGGTTGGTGTCTTCACGGAATTCAGCGGCACTGAGTTTACAAACTTCCTTGAGGTTGCTGAGAAGC
CD GTATAGGTTGGTGTCTTCACGGAATTCAGCGGCACTGAGTTTACAAACTTCCTTGAGGTTGCTGAGAAGC
Ka GTATAGGTTGGTGTCTTCACGGAATTCAGCGGCACTGAGTTTACAAACTTCCTTGAGGTTGCTGAGAAGC
Exon 4
1140 1150 1160 1170 1180 1190 1200
TtPDI4B TGAGGTCTGATTATGACTTTGGCCACACCGTGCATGCCAACCATCTCCCACGTGGTGATGCCGCAGTGGA
wPDI2 TGAGGTCTGATTATGACTTTGGCCACACCGTGCATGCCAACCATCTCCCACGTGGTGATGCCGCAGTGGA
Cr TGAGGTCTGATTATGACTTTGGCCACACCGTGCATGCCAACCATCTCCCACGTGGTGATGCCGCAGTGGA
Hb TGAGGTCTGATTATGACTTTGGCCACACCGTGCATGCCAACCATCTCCCACGTGGTGATGCCGCAGTGGA
Eg TGAGGTCTGATTATGACTTTGGCCACACCGTGCATGCCAACCATCTCCCACGTGGTGATGCCGCAGTGGA
Ss TGAGGTCTGATTATGACTTTGGCCACACCGTGCATGCCAACCATCTCCCACGTGGTGATGCCGCAGTGGA
Sc TGAGGTCTGATTATGACTTTGGCCACACCGTGCATGCCAACCATCTCCCACGTGGTGATGCCGCAGTGGA
Ta TGAGGTCTGATTATGACTTTGGCCACACCGTGCATGCCAACCATCTCCCACGTGGTGATGCCGCAGTGGA
CD TGAGGTCTGATTATGACTTTGGCCACACCGTGCATGCCAACCATCTCCCACGTGGTGATGCCGCAGTGGA
Ka TGAGGTCTGATTATGACTTTGGCCACACCGTGCATGCCAACCATCTCCCACGTGGTGATGCCGCAGTGGA
1210 PB1R
TtPDI4B GAGGCCATTGG
wPDI2 GAGGCCATTGG
Cr GAGGCCATTGG
Hb GAGGCCATTGG
Eg GAGGCCATTGG
Ss GAGGCCATTGG
Sc GAGGCCATTGG
Ta GAGGCCATTGG
CD GAGGCCATTGG
Ka GAGGCCATTGG


PB2F
1440 1450 1460 1470 1480 1490 1500
TtPDI4B CTGCTTTGGAGAAATTCATTGAGGCTAGCAGCACCCCGAAAGTTGTTACTTTTGACAAGAACCCTGACAA
wPDI2 CTGCTTTGGAGAAATTCATTGAGGCTAGCAGCACCCCGAAAGTTGTTACTTTTGACAAGAACCCTGACAA
Cr ----------CAGCACCCCGAAAGTTGTTACTTTTGACAAGAACCCTGACAA
Hb ----------AGTTGTTACTTTTGACAAGAACCCTGACAA
Eg ----------
Ss ----------
Sc ----------GAAAGTTGTTACTTTTGACAAGAACCCTGACAA
Ta ----------GAAAGTTGTTACTTTTGACAAGAACCCTGACAA
CD ----------CAGCACCCCGAAAGTTGTTACTTTTGACAAGAACCCTGACAA
Ka ----------GTTGTTACTTTTGACAAGAACCCTGACAA
Exon 5
1510 1520 1530 1540 1550 1560 1570
TtPDI4B CCATCCTTACCTCTTGAAATTCTTCCAGAGCAATGCTCCCAAGGTAATGACTGACACAACTTGCTTCTAG
wPDI2 CCATCCTTACCTCTTGAAATTCTTCCAGAGCAATGCTCCCAAG----------
Cr CCATCCTTACCTCTTGAAATTCTTCCAGAGCAATGCTCCCAAGGTAATGACTGACACAACTTGCTTCTAG
Hb CCATCCTTACCTCTTGAAATTCTTCCAGAGCAATGCTCCCAAGGTAATGACTGACACAACTTGCTTCTAG
Eg ---------CCTCTTGAAATTCTTCCAGAGCAATGCTCCCAAGGTAATGACTGACACAACTTGCTTCTAG
Ss ----------GCAATGCTCCCAAGGTAATGACTGACACAACTTGCTTCTAG
Sc CCATCCTTACCTCTTGAAATTCTTCCAGAGCAATGCTCCCAAGGTAATGACTGACACAACTTGCTTCTAG
Ta CCATCCTTACCTCTTGAAATTCTTCCAGAGCAATGCTCCCAAGGTAATGACTGACACAACTTGCTTCTAG
CD CCATCCTTACCTCTTGAAATTCTTCCAGAGCAATGCTCCCAAGGTAATGACTGACACAACTTGCTTCTAG
Ka CCATCCTTACCTCTTGAAATTCTTCCAGAGCAATGCTCCCAAGGTAATGACTGACACAACTTGCTTCTAG
Intron 5
1580 1590 1600 1610 1620 1630 1640
TtPDI4B AATGCTTTGTCATTTTTCTTAATTGTGTGAGTTCTGTATAGTTGACTGATTTCTACTATATGCTTAGTTA
wPDI2 ----------
Cr AATGCTTTGTCATTTTTCTTAATTGTGTGAGTTCTGTATAGTTGACTGATTTCTACTATATGCTTAGTCA
Hb AATGCTTTGTCATTTTTCTTAATTGTGTGAGTTCTGTATAGTTGACTGATTTCTACTATATGCTTAGTCA
Eg AATGCTTTGTCATTTTTCTTAATTGTGTGAGTTCTGTATAGTTGACTGATTTCTACTATATGCTTAGTCA
Ss AATGCTTTGTCATTTTTCTTAATTGTGTGAGTTCTGTATAGTTGACTGATTTCTACTATATGCTTAGTCA
Sc AATGCTTTGTCATTTTTCTTAATTGTGTGAGTTCTGTATAGTTGACTGATTTCTACTATATGCTTAGTCA
Ta AATGCTTTGTCATTTTTCTTAATTGTGTGAGTTCTGTATAGTTGACTGATTTCTACTATATGCTTAGTCA
CD AATGCTTTGTCATTTTTCTTAATTGTGTGAGTTCTGTATAGTTGACTGATTTCTACTATATGCTTAGTCA
Ka AATGCTTTGTCATTTTTCTTAATTGTGTGAGTTCTGTATAGTTGACTGATTTCTACTATATGCTTAGTCA
1650 1660 1670 1680 1690 1700 1710
TtPDI4B GTTAGGTCTATGCCTCATGTTTGTAACTAATGAGGCCATAATTTTTATCACTGTATGTTGTTTTGATAAG
wPDI2 ----------
Cr GTTAGGTCTATGCCTCATGTTTGTAACTAATGAGGCCATAATTTTTATCACTGTATGTTGTTTTGATAAG
Hb GTTAGGTCTATGCCTCATGTTTGTAACTAATGAGGCCATAATTTTTATCACTGTATGTTGTTTTGATAAG
Eg GTTAGGTCTATGCCTCATGTTTGTAACTAATGAGGCCATAATTTTTATCACTGTATGTTGTTTTGATAAG
Ss GTTAGGTCTATGCCTCATGTTTGTAACTAATGAGGCCATAATTTTTATCACTGTATGTTGTTTTGATAAG
Sc GTTAGGTCTATGCCTCATGTTTGTAACTAATGAGGCCATAATTTTTATCACTGTATGTTGTTTTGATAAG
Ta GTTAGGTCTATGCCTCATGTTTGTAACTAATGAGGCCATAATTTTTATCACTGTATGTTGTTTTGATAAG
CD GTTAGGTCTATGCCTCATGTTTGTAACTAATGAGGCCATAATTTTTATCACTGTATGTTGTTTTGATAAG
Ka GTTAGGTCTATGCCTCATGTTTGTAACTAATGAGGCCATAATTTTTATCACTGTATGTTGTTTTGATAAG
1720 1730 1740 1750 1760 1770 1780
TtPDI4B AGAGATGATTGCTACTGTATATTTCAGTTGATTATCAGCCTGACTCTTGGTGTGTACAGAAAACTTCTGC
wPDI2 ----------
Cr AGAGATGATTGCTACTGTATATTTCAGTTGATTATCAGCCTGACTCTTGGTGTGTACAGAAAACTTCTGC
Hb AGAGATGATTGCTACTGTATATTTCAGTTGATTATCAGCCTGACTCTTGGTGTGTACAGAAAACTTCTGC
Eg AGAGATGATTGCTACTGTATATTTCAGTTGATTATCAGCCTGACTCTTGGTGTGTACAGAAAACTTCTGC
Ss AGAGATGATTGCTACTGTATATTTCAGTTGATTATCAGCCTGACTCTTGGTGTGTACAGAAAACTTCTGC
Sc AGAGATGATTGCTACTGTATATTTCAGTTGATTATCAGCCTGACTCTTGGTGTGTACAGAAAACTTCTGC
Ta AGAGATGATTGCTACTGTATATTTCAGTTGATTATCAGCCTGACTCTTGGTGTGTACAGAAAACTTCTGC
CD AGAGATGATTGCTACTGTATATTTCAGTTGATTATCAGCCTGACTCTTGGTGTGTACAGAAAACTTCTGC
Ka AGAGATGATTGCTACTGTATATTTCAGTTGATTATCAGCCTGACTCTTGGTGTGTACAGAAAACTTCTGC


1790 1800 1810 1820 1830 1840 1850
TtPDI4B AATCAAATGATGGTTGAATTTGTCAATTTATTCTTCAACATGTTGACATTGTATAATCTAGTCACTTGTC
wPDI2 ---
Cx AATCAAATGATGGTTGAATTTGTCAATTTATTCTTCAACATGTTGACATTGTATAATCTAGTCACTTGTC
Hb AATCAAATGATGGTTGAATTTGTCAATTTATTCTTCAACATGTTGACATTGTATAATCTAGTCACTTGTC
Eg AATCAAATGATGGTTGAATTTGTCAATTTATTCTTCAACATGTTGACATTGTATAATCTAGTCACTTGTC
Ss AATCAAATGATGGTTGAATTTGTCAATTTATTCTTCAACATGTTGACATTGTATAATCTAGTCACTTGTC
Sc AATCAAATGATGGTTGAATTTGTCAATTTATTCTTCAACATGTTGACATTGTATAATCTAGTCACTTGTC
Ta AATCAAATGATGGTTGAATTTGTCAATTTATTCTTCAACATGTTGACATTGTATAATCTAGTCACTTGTC
CD AATCAAATGATGGTTGAATTTGTCAATTTATTCTTCAACATGTTGACATTGTATAATCTAGTCACTTGTC
Ka AATCAAATGATGGTTGAATTTGTCAATTTATTCTTCAACATGTTGACATTGTATAATCTAGTCACTTGTC
1860 1870 1880 1890 1900 1910 1920
TtPDI4B GGCCCAAGTCTTTGAATTGGGCAGCTAGTTCCTTCAATGTTCATTCATCCTGTACATCTAAGAAATCATC
wPDI2 ---
Cx GGCCCAAGTCTTTGAATTGGGCAGCTAGTTCCTTCAATGTTCATTCATCCTGTACATCTAAGAAATCATC
Hb GGCCCAAGTCTTTGAATTGGGCAGCTAGTTCCTTCAATGTTCATTCATCCTGTACATCTAAGAAATCATC
Eg GGCCCAAGTCTTTGAATTGGGCAGCTAGTTCCTTCAATGTTCATTCATCCTGTACATCTAAGAAATCATC
Ss GGCCCAAGTCTTTGAATTGGGCAGCTAGTTCCTTCAATGTTCATTCATCCTGTACATCTAAGAAATCATC
Sc GGCCCAAGTCTTTGAATTGGGCAGCTAGTTCCTTCAATGTTCATTCATCCTGTACATCTAAGAAATCATC
Ta GGCCCAAGTCTTTGAATTGGGCAGCTAGTTCCTTCAATGTTCATTCATCCTGTACATCTAAGAAATCATC
CD GGCCCAAGTCTTTGAATTGGGCAGCTAGTTCCTTCAATGTTCATTCATCCTGTACATCTAAGAAATCATC
Ka GGCCCAAGTCTTTGAATTGGGCAGCTAGTTCCTTCAATGTTCATTCATCCTGTACATCTAAGAAATCATC
1930 1940 1950 1960 1970 1980 1990
TtPDI4B TTGCAATGAATTTGTTCAGCACATTTGGAGGATTACTAGTGTATTTAGTTTTGTTAGCCTTTTGGATCTT
wPDI2 ---
Cx TTGCAATGAATTTGTTCAGCACATTTGGAGGATTACTAGTGTATTTAGTTTTGTTAGCCTTTTGGATCTT
Hb TTGCAATGAATTTGTTCAGCACATTTGGAGGATTACTAGTGTATTTAGTTTTGTTAGCCTTTTGGATCTT
Eg TTGCAATGAATTTGTTCAGCACATTTGGAGGATTACTAGTGTATTTAGTTTTGTTAGCCTTTTGGATCTT
Ss TTGCAATGAATTTGTTCAGCACATTTGGAGGATTACTAGTGTATTTAGTTTTGTTAGCCTTTTGGATCTT
Sc TTGCAATGAATTTGTTCAGCACATTTGGAGGATTACTAGTGTATTTAGTTTTGTTAGCCTTTTGGATCTT
Ta TTGCAATGAATTTGTTCAGCACATTTGGAGGATTACTAGTGTATTTAGTTTTGTTAGCCTTTTGGATCTT
CD TTGCAATGAATTTGTTCAGCACATTTGGAGGATTACTAGTGTATTTAGTTTTGTTAGCCTTTTGGATCTT
Ka TTGCAATGAATTTGTTCAGCACATTTGGAGGATTACTAGTGTATTTAGTTTTGTTAGCCTTTTGGATCTT
2000 2010 2020 2030 2040 2050 2060
TtPDI4B TGTGTTTTCATTCATGGTTGTTGTACCTGAAAGGTCTACTTTAGAGCCATAGATGTAATGATTGTTTGAG
wPDI2 ---
Cx TGTGTTTTCATTCATGGTTGTTGTACCTGAAAGGTCTACTTTAGAGCCATAGATGTAATGATTGTTTGAG
Hb TGTGTTTTCATTCATGGTTGTTGTACCTGAAAGGTCTACTTTAGAGCCATAGATGTAATGATTGTTTGAG
Eg TGTGTTTTCATTCATGGTTGTTGTACCTGAAAGGTCTACTTTAGAGCCATAGATGTAATGATTGTTTGAG
Ss TGTGTTTTCATTCATGGTTGTTGTACCTGAAAGGTCTACTTTAGAGCCATAGATGTAATGATTGTTTGAG
Sc TGTGTTTTCATTCATGGTTGTTGTACCTGAAAGGTCTACTTTAGAGCCATAGATGTAATGATTGTTTGAG
Ta TGTGTTTTCATTCATGGTTGTTGTACCTGAAAGGTCTACTTTAGAGCCATAGATGTAATGATTGTTTGAG
CD TGTGTTTTCATTCATGGTTGTTGTACCTGAAAGGTCTACTTTAGAGCCATAGATGTAATGATTGTTTGAG
Ka TGTGTTTTCATTCATGGTTGTTGTACCTGAAAGGTCTACTTTAGAGCCATAGATGTAATGATTGTTTGAG
2070 2080 2090 2100 2110 2120 2130
TtPDI4B ATGCAGAATGCCATCTTTGGTGTTTTCGGTGCTTTTTTGCGTAAAAATATTTCAGATAGTGGTGAAGCTG
wPDI2 ---
Cx ATGCAGAATGCCATCTTTGGTGTTTTCAGTGCTTTTTTGCGTAAAAATATTTCAGATAGTGGTGAAGCTG
Hb ATGCAGAATGCCATCTTTGGTGTTTTCAGTGCTTTTTTGCGTAAAAATATTTCAGATAGTGGTGAAGCTG
Eg ATGCAGAATGCCATCTTTGGTGTTTTCAGTGCTTTTTTGCGTAAAAATATTTCAGATAGTGGTGAAGCTG
Ss ATGCAGAATGCCATCTTTGGTGTTTTCAGTGCTTTTTTGCGTAAAAATATTTCAGATAGTGGTGAAGCTG
Sc ATGCAGAATGCCATCTTTGGTGTTTTCAGTGCTTTTTTGCGTAAAAATATTTCAGATAGTGGTGAAGCTG
Ta ATGCAGAATGCCATCTTTGGTGTTTTCAGTGCTTTTTTGCGTAAAAATATTTCAGATAGTGGTGAAGCTG
CD ATGCAGAATGCCATCTTTGGTGTTTTCAGTGCTTTTTTGCGTAAAAATATTTCAGATAGTGGTGAAGCTG
Ka ATGCAGAATGCCATCTTTGGTGTTTTCAGTGCTTTTTTGCGTAAAAATATTTCAGATAGTGGTGAAGCTG


2140 2150 2160 2170 2180 2190 2200
TtPDI4B CTTAGAGCATTAGCCTTTTCATTGTTTTTTAGTTATTGGTTACTCATTATTTGTCACTTCTGTTCTCTGA
wPDI2 ----------
Cr CTTAGAGCATTAGCCTTTTCATTGTTTTTTAGTTATTGGTTACTCATTATTTGTCACTTCTGTTCTCTGA
Hb CTTAGAGCATTAGCCTTTTCATTGTTTTTTAGTTATTGGTTACTCATTATTTGTCACTTCTGTTCTCTGA
Eg CTTAGAGCATTAGCCTTTTCATTGTTTTTTAGTTATTGGTTACTCATTATTTGTCACTTCTGTTCTCTGA
Ss CTTAGAGCATTAGCCTTTTCATTGTTTTTTAGTTATTGGTTACTCATTATTTGTCACTTCTGTTCTCTGA
Sc CTTAGAGCATTAGCCTTTTCATTGTTTTTTAGTTATTGGTTACTCATTATTTGTCACTTCTGTTCTCTGA
Ta CTTAGAGCATTAGCCTTTTCATTGTTTTTTAGTTATTGGTTACTCATTATTTGTCACTTCTGTTCTCTGA
CD CTTAGAGCATTAGCCTTTTCATTGTTTTTTAGTTATTGGTTACTCATTATTTGTCACTTCTGTTCTCTGA
Ka CTTAGAGCATTAGCCTTTTCATTGTTTTTTAGTTATTGGTTACTCATTATTTGTCACTTCTGTTCTCTGA
2210 2220 2230 2240 2250 2260 2270
TtPDI4B CATCCAATGGTATGCAGGCCATGCTCTTTTTGAACTTCTCCACTGGACCGTTTGAGTCCTTCAAGAAAGC
wPDI2 ----------GCCATGCTCTTTTTGAACTTCTCCACTGGACCGTTTGAGTCCTTCAAGAAAGC
Cr CATCCAATGGTATGCAGGCCATGCTCTTTTTGAACTTCTCCACTGGACCGTTTGAGTC----------
Hb CATCCAATGGTATGCAGGCCATGCTCTTTTTGAACTTCTCCA----------
Eg CATCCAATGGTATGCAGGCCATGCTCTTTTTGAACTTCTCCAC----------
Ss CATCCAATGGTATGCAGGCCATGCT----------
Sc CATCCAATGGTATGCAGGCCATGCT----------
Ta CATCCAATGGTATGCAGGCCATGC----------
CD CATCCAATGGTATGCAGGCCATGCTCTTTTTGAACTTCTCCACTGGACCGTTTGA----------
Ka CATCCAATGGTATGCAGGCCATGCTCTTTTTGAACTTCTCCACTGGACCGTTTGAGTCCTTCAAGAAAGC
Exon 6
2280 2290 2300 2310 PB2R 2320
TtPDI4B CTACTATGGTGCTGTAGAGGAGTTCAGCGGCAAGGATGTCAAGTTCCT
wPDI2 CTACTATGGTGCTGTAGAGGAGTTCAGCGGCAAGGATGTCAAGTTCCT
Cr ----------
Hb ----------
Eg ----------
Ss ----------
Sc ----------
Ta ----------
CD ----------
Ka CTACTATGGTGCTGTAGAGGAGTTC----------


PB3F
Exon 7
Intron 7
Exon 8
Intron 8
TtPDI4B
wPDI2
Cx
Hb
Eg
Ss
Sc
Ta
CD
Ka

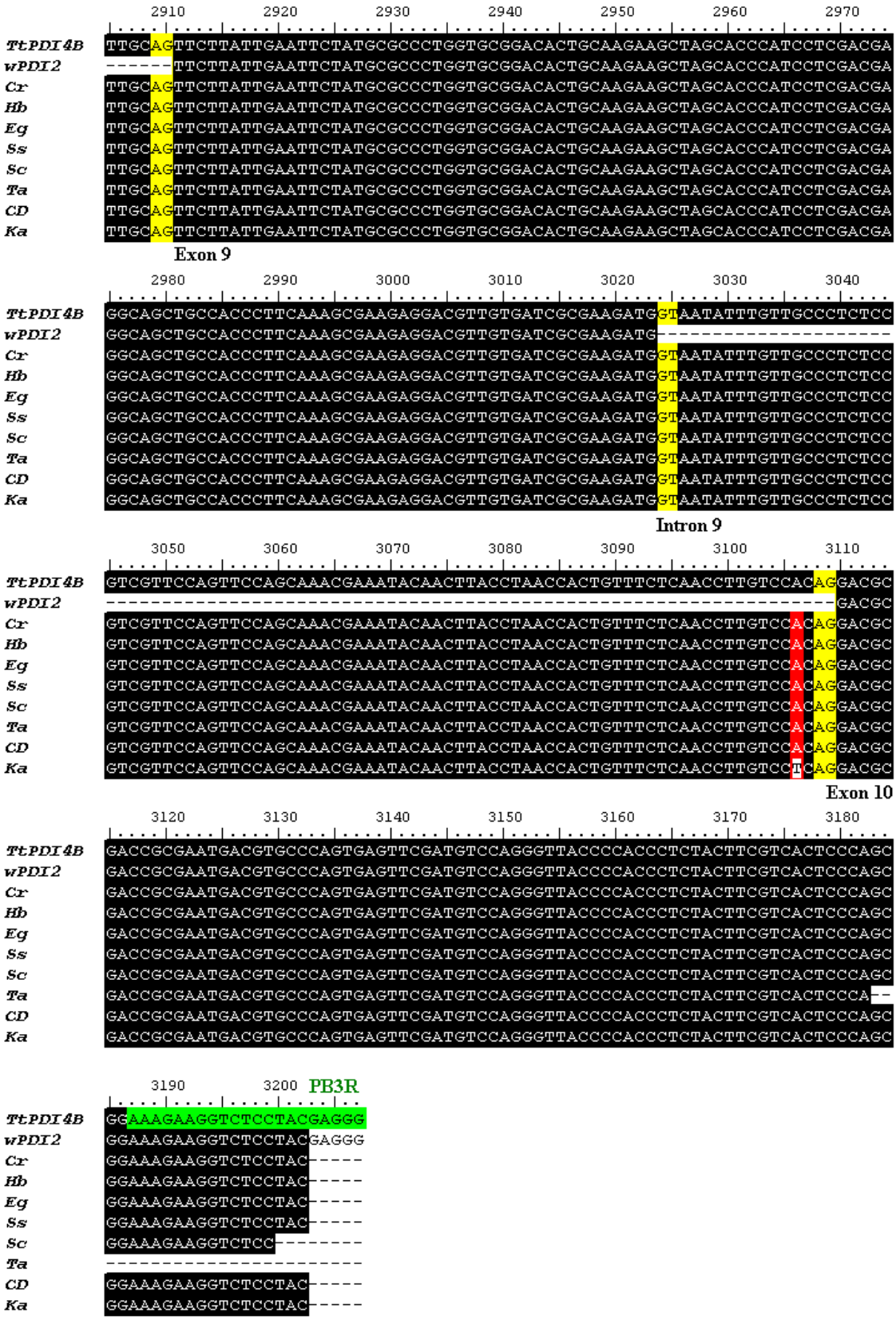
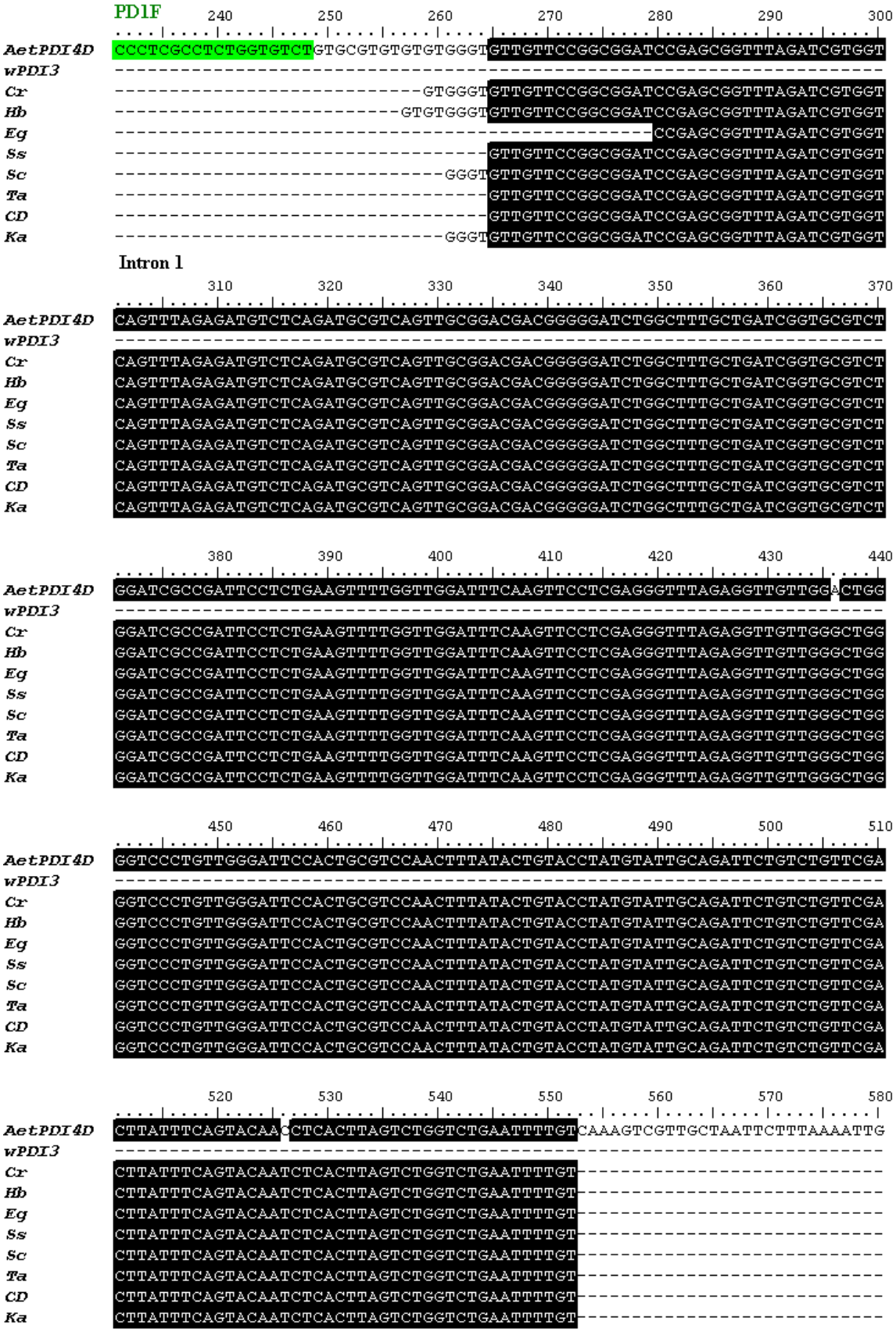
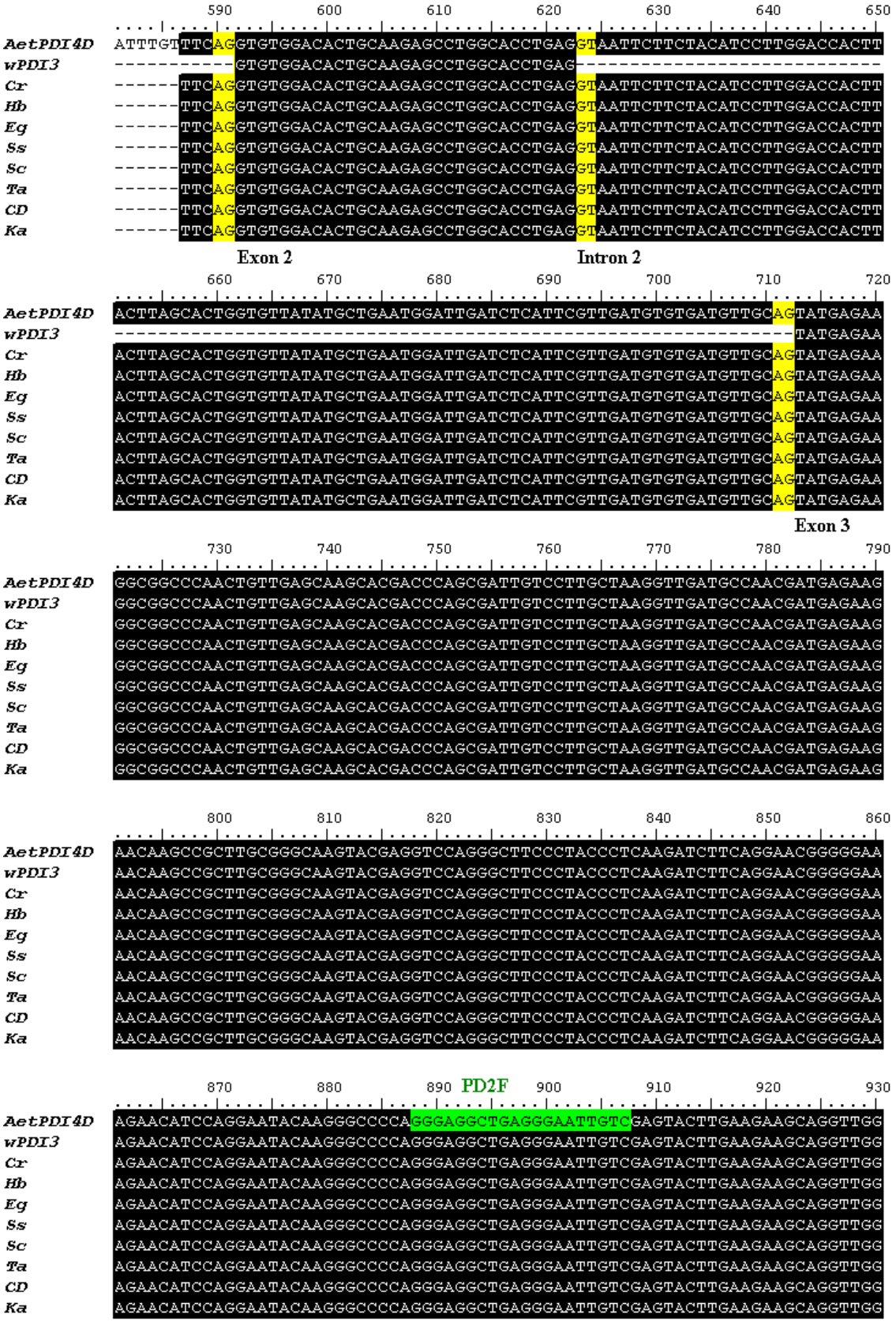
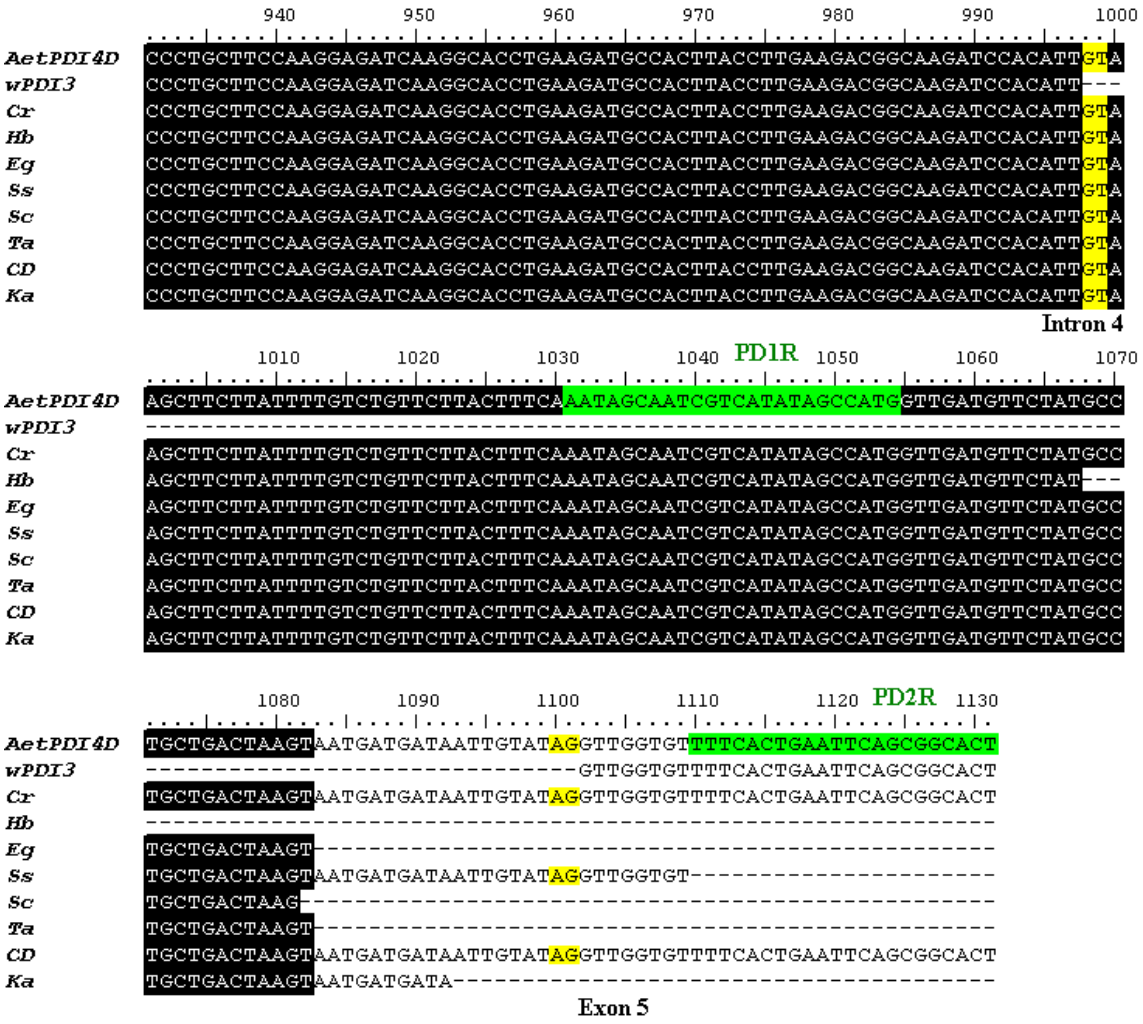
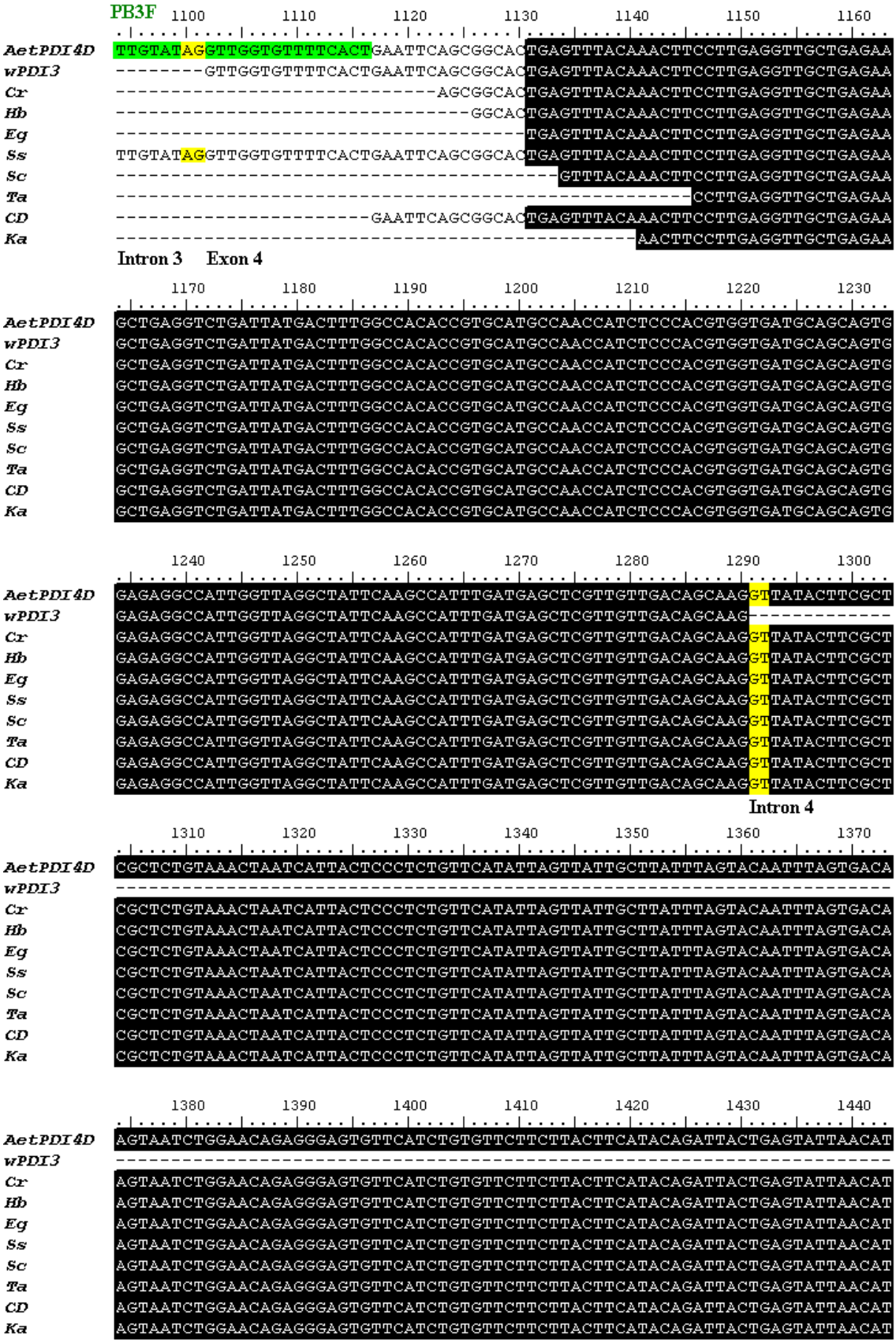


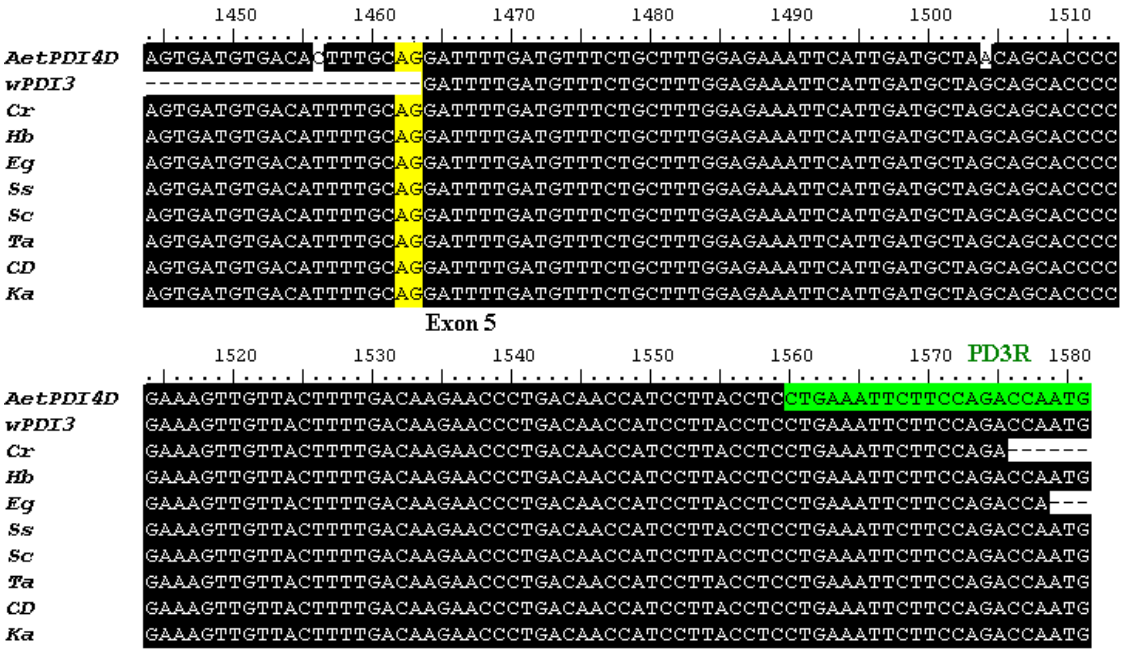
Figure IIIB The partial sequence of the *PDI* genes in the B genome of wheat

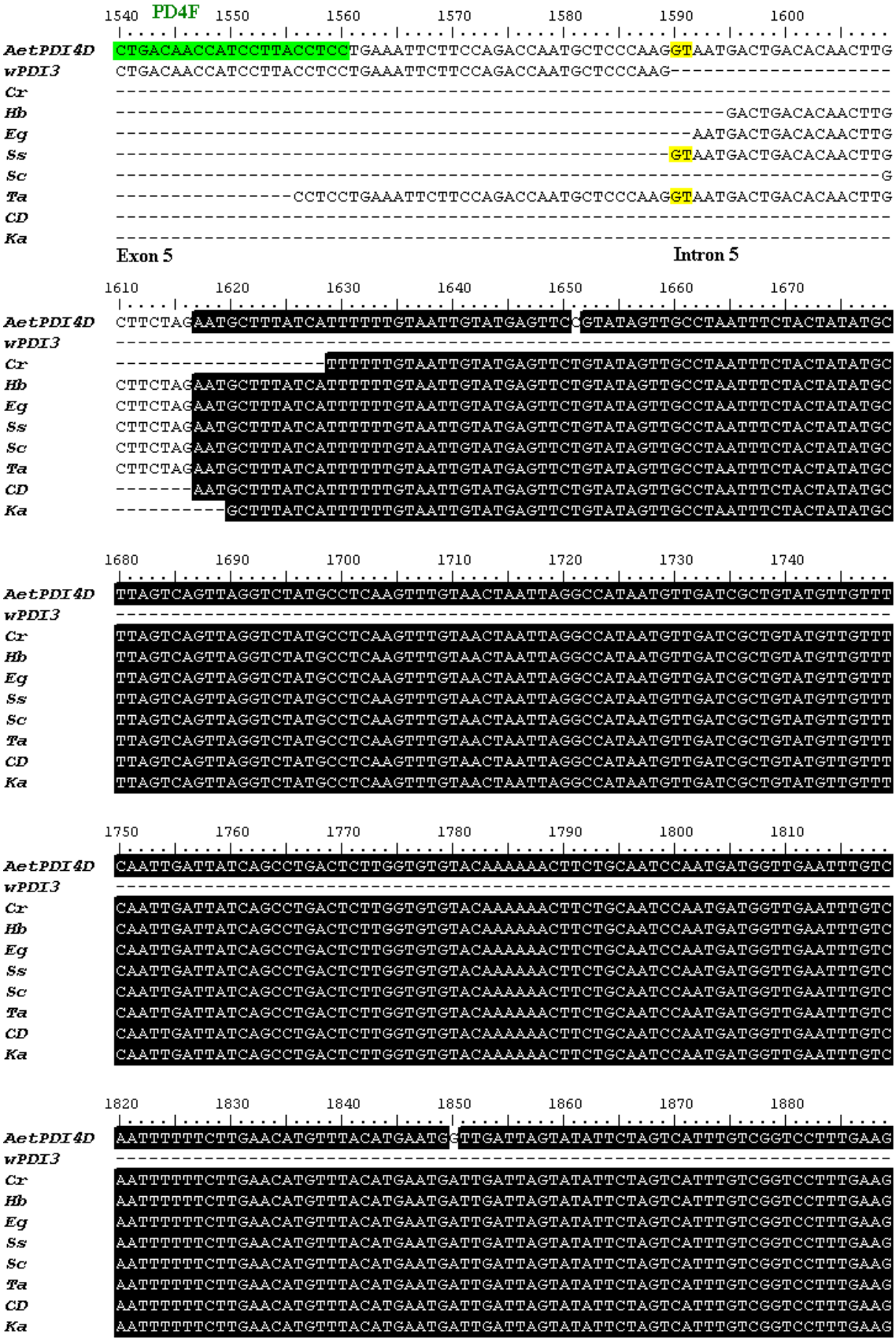

PD1F
Intron 1
AetPDI4D
wPDI3
Cx
Hb
Eg
Ss
Sc
Ta
CD
Ka

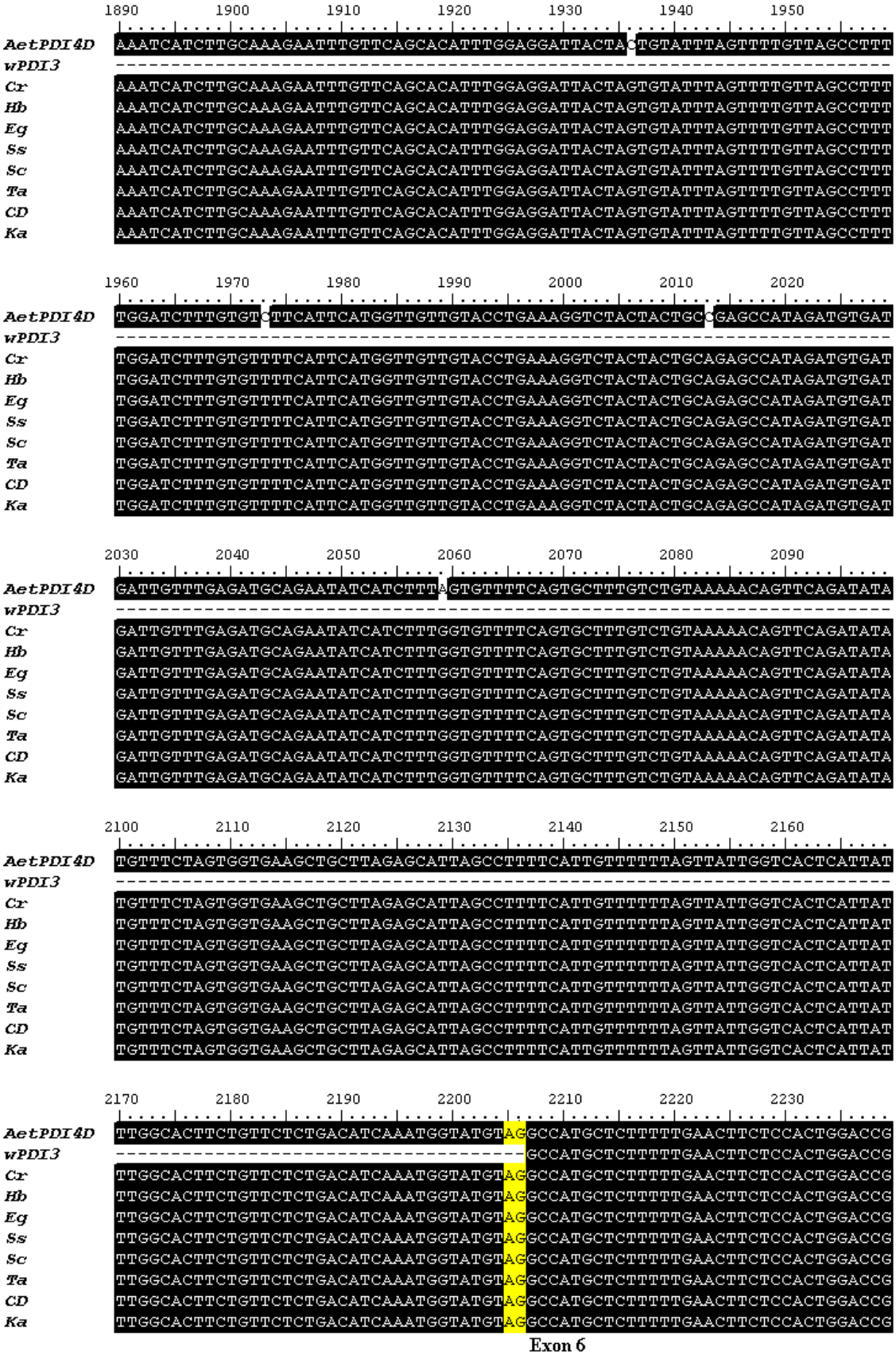

590 600 610 620 630 640 650
AetPDI4D ATTTGTTTCAGGTGTGGACACTGCAAGAGCCTGGCACCTGAGGTAATTCTTCTACATCCTTGGACCACTT
wPDI3 ----------GTGTGGACACTGCAAGAGCCTGGCACCTGAG----------
Cr ------TTCAGGTGTGGACACTGCAAGAGCCTGGCACCTGAGGTAATTCTTCTACATCCTTGGACCACTT
Hb ------TTCAGGTGTGGACACTGCAAGAGCCTGGCACCTGAGGTAATTCTTCTACATCCTTGGACCACTT
Eg ------TTCAGGTGTGGACACTGCAAGAGCCTGGCACCTGAGGTAATTCTTCTACATCCTTGGACCACTT
Ss ------TTCAGGTGTGGACACTGCAAGAGCCTGGCACCTGAGGTAATTCTTCTACATCCTTGGACCACTT
Sc ------TTCAGGTGTGGACACTGCAAGAGCCTGGCACCTGAGGTAATTCTTCTACATCCTTGGACCACTT
Ta ------TTCAGGTGTGGACACTGCAAGAGCCTGGCACCTGAGGTAATTCTTCTACATCCTTGGACCACTT
CD ------TTCAGGTGTGGACACTGCAAGAGCCTGGCACCTGAGGTAATTCTTCTACATCCTTGGACCACTT
Ka ------TTCAGGTGTGGACACTGCAAGAGCCTGGCACCTGAGGTAATTCTTCTACATCCTTGGACCACTT
Exon 2
Intron 2
660 670 680 690 700 710 720
AetPDI4D ACTTAGCACTGGTGTTATATGCTGAATGGATTGATCTCATTCGTTGATGTGTGATGTTGCAGTATGAGAA
wPDI3 ----------TATGAGAA
Cr ACTTAGCACTGGTGTTATATGCTGAATGGATTGATCTCATTCGTTGATGTGTGATGTTGCAGTATGAGAA
Hb ACTTAGCACTGGTGTTATATGCTGAATGGATTGATCTCATTCGTTGATGTGTGATGTTGCAGTATGAGAA
Eg ACTTAGCACTGGTGTTATATGCTGAATGGATTGATCTCATTCGTTGATGTGTGATGTTGCAGTATGAGAA
Ss ACTTAGCACTGGTGTTATATGCTGAATGGATTGATCTCATTCGTTGATGTGTGATGTTGCAGTATGAGAA
Sc ACTTAGCACTGGTGTTATATGCTGAATGGATTGATCTCATTCGTTGATGTGTGATGTTGCAGTATGAGAA
Ta ACTTAGCACTGGTGTTATATGCTGAATGGATTGATCTCATTCGTTGATGTGTGATGTTGCAGTATGAGAA
CD ACTTAGCACTGGTGTTATATGCTGAATGGATTGATCTCATTCGTTGATGTGTGATGTTGCAGTATGAGAA
Ka ACTTAGCACTGGTGTTATATGCTGAATGGATTGATCTCATTCGTTGATGTGTGATGTTGCAGTATGAGAA
Exon 3
730 740 750 760 770 780 790
AetPDI4D GGCGGCCCAACTGTTGAGCAAGCACGACCCAGCGATTGTCCTTGCTAAGGTTGATGCCAACGATGAGAAG
wPDI3 GGCGGCCCAACTGTTGAGCAAGCACGACCCAGCGATTGTCCTTGCTAAGGTTGATGCCAACGATGAGAAG
Cr GGCGGCCCAACTGTTGAGCAAGCACGACCCAGCGATTGTCCTTGCTAAGGTTGATGCCAACGATGAGAAG
Hb GGCGGCCCAACTGTTGAGCAAGCACGACCCAGCGATTGTCCTTGCTAAGGTTGATGCCAACGATGAGAAG
Eg GGCGGCCCAACTGTTGAGCAAGCACGACCCAGCGATTGTCCTTGCTAAGGTTGATGCCAACGATGAGAAG
Ss GGCGGCCCAACTGTTGAGCAAGCACGACCCAGCGATTGTCCTTGCTAAGGTTGATGCCAACGATGAGAAG
Sc GGCGGCCCAACTGTTGAGCAAGCACGACCCAGCGATTGTCCTTGCTAAGGTTGATGCCAACGATGAGAAG
Ta GGCGGCCCAACTGTTGAGCAAGCACGACCCAGCGATTGTCCTTGCTAAGGTTGATGCCAACGATGAGAAG
CD GGCGGCCCAACTGTTGAGCAAGCACGACCCAGCGATTGTCCTTGCTAAGGTTGATGCCAACGATGAGAAG
Ka GGCGGCCCAACTGTTGAGCAAGCACGACCCAGCGATTGTCCTTGCTAAGGTTGATGCCAACGATGAGAAG
800 810 820 830 840 850 860
AetPDI4D AACAAGCCGCTTGCGGGCAAGTACGAGGTCCAGGGCTTCCCTACCCTCAAGATCTTCAGGAACGGGGGAA
wPDI3 AACAAGCCGCTTGCGGGCAAGTACGAGGTCCAGGGCTTCCCTACCCTCAAGATCTTCAGGAACGGGGGAA
Cr AACAAGCCGCTTGCGGGCAAGTACGAGGTCCAGGGCTTCCCTACCCTCAAGATCTTCAGGAACGGGGGAA
Hb AACAAGCCGCTTGCGGGCAAGTACGAGGTCCAGGGCTTCCCTACCCTCAAGATCTTCAGGAACGGGGGAA
Eg AACAAGCCGCTTGCGGGCAAGTACGAGGTCCAGGGCTTCCCTACCCTCAAGATCTTCAGGAACGGGGGAA
Ss AACAAGCCGCTTGCGGGCAAGTACGAGGTCCAGGGCTTCCCTACCCTCAAGATCTTCAGGAACGGGGGAA
Sc AACAAGCCGCTTGCGGGCAAGTACGAGGTCCAGGGCTTCCCTACCCTCAAGATCTTCAGGAACGGGGGAA
Ta AACAAGCCGCTTGCGGGCAAGTACGAGGTCCAGGGCTTCCCTACCCTCAAGATCTTCAGGAACGGGGGAA
CD AACAAGCCGCTTGCGGGCAAGTACGAGGTCCAGGGCTTCCCTACCCTCAAGATCTTCAGGAACGGGGGAA
Ka AACAAGCCGCTTGCGGGCAAGTACGAGGTCCAGGGCTTCCCTACCCTCAAGATCTTCAGGAACGGGGGAA
870 880 890 PD2F 900 910 920 930
AetPDI4D AGAACATCCAGGAATACAAGGGCCCCAGGGAGGCTGAGGGAATTGTCGAGTACTTGAAGAAGCAGGTTGG
wPDI3 AGAACATCCAGGAATACAAGGGCCCCAGGGAGGCTGAGGGAATTGTCGAGTACTTGAAGAAGCAGGTTGG
Cr AGAACATCCAGGAATACAAGGGCCCCAGGGAGGCTGAGGGAATTGTCGAGTACTTGAAGAAGCAGGTTGG
Hb AGAACATCCAGGAATACAAGGGCCCCAGGGAGGCTGAGGGAATTGTCGAGTACTTGAAGAAGCAGGTTGG
Eg AGAACATCCAGGAATACAAGGGCCCCAGGGAGGCTGAGGGAATTGTCGAGTACTTGAAGAAGCAGGTTGG
Ss AGAACATCCAGGAATACAAGGGCCCCAGGGAGGCTGAGGGAATTGTCGAGTACTTGAAGAAGCAGGTTGG
Sc AGAACATCCAGGAATACAAGGGCCCCAGGGAGGCTGAGGGAATTGTCGAGTACTTGAAGAAGCAGGTTGG
Ta AGAACATCCAGGAATACAAGGGCCCCAGGGAGGCTGAGGGAATTGTCGAGTACTTGAAGAAGCAGGTTGG
CD AGAACATCCAGGAATACAAGGGCCCCAGGGAGGCTGAGGGAATTGTCGAGTACTTGAAGAAGCAGGTTGG
Ka AGAACATCCAGGAATACAAGGGCCCCAGGGAGGCTGAGGGAATTGTCGAGTACTTGAAGAAGCAGGTTGG


940 950 960 970 980 990 1000
AetPDI4D CCCTGCTTCCAAGGAGATCAAGGCACCTGAAGATGCCACTTACCTTGAAGACGGCAAGATCCACATTGTA
wPDI3 CCCTGCTTCCAAGGAGATCAAGGCACCTGAAGATGCCACTTACCTTGAAGACGGCAAGATCCACATT---
Cr CCCTGCTTCCAAGGAGATCAAGGCACCTGAAGATGCCACTTACCTTGAAGACGGCAAGATCCACATTGTA
Hb CCCTGCTTCCAAGGAGATCAAGGCACCTGAAGATGCCACTTACCTTGAAGACGGCAAGATCCACATTGTA
Eg CCCTGCTTCCAAGGAGATCAAGGCACCTGAAGATGCCACTTACCTTGAAGACGGCAAGATCCACATTGTA
Ss CCCTGCTTCCAAGGAGATCAAGGCACCTGAAGATGCCACTTACCTTGAAGACGGCAAGATCCACATTGTA
Sc CCCTGCTTCCAAGGAGATCAAGGCACCTGAAGATGCCACTTACCTTGAAGACGGCAAGATCCACATTGTA
Ta CCCTGCTTCCAAGGAGATCAAGGCACCTGAAGATGCCACTTACCTTGAAGACGGCAAGATCCACATTGTA
CD CCCTGCTTCCAAGGAGATCAAGGCACCTGAAGATGCCACTTACCTTGAAGACGGCAAGATCCACATTGTA
Ka CCCTGCTTCCAAGGAGATCAAGGCACCTGAAGATGCCACTTACCTTGAAGACGGCAAGATCCACATTGTA
Intron 4
1010 1020 1030 1040 PD1R 1050 1060 1070
AetPDI4D AGCTTCTTATTTTGTCTGTTCTTACTTTCAAATAGCAATCGTCATATAGCCATGGTTGATGTTCTATGCC
wPDI3 ----------
Cr AGCTTCTTATTTTGTCTGTTCTTACTTTCAAATAGCAATCGTCATATAGCCATGGTTGATGTTCTATGCC
Hb AGCTTCTTATTTTGTCTGTTCTTACTTTCAAATAGCAATCGTCATATAGCCATGGTTGATGTTCTAT---
Eg AGCTTCTTATTTTGTCTGTTCTTACTTTCAAATAGCAATCGTCATATAGCCATGGTTGATGTTCTATGCC
Ss AGCTTCTTATTTTGTCTGTTCTTACTTTCAAATAGCAATCGTCATATAGCCATGGTTGATGTTCTATGCC
Sc AGCTTCTTATTTTGTCTGTTCTTACTTTCAAATAGCAATCGTCATATAGCCATGGTTGATGTTCTATGCC
Ta AGCTTCTTATTTTGTCTGTTCTTACTTTCAAATAGCAATCGTCATATAGCCATGGTTGATGTTCTATGCC
CD AGCTTCTTATTTTGTCTGTTCTTACTTTCAAATAGCAATCGTCATATAGCCATGGTTGATGTTCTATGCC
Ka AGCTTCTTATTTTGTCTGTTCTTACTTTCAAATAGCAATCGTCATATAGCCATGGTTGATGTTCTATGCC
1080 1090 1100 1110 1120 PD2R 1130
AetPDI4D TGCTGACTAAGTAATGATGATAATTGTATAGGTTGGTGTTTTCACTGAATTCAGCGGCACT
wPDI3 ----------GTTGGTGTTTTCACTGAATTCAGCGGCACT
Cr TGCTGACTAAGTAATGATGATAATTGTATAGGTTGGTGTTTTCACTGAATTCAGCGGCACT
Hb ----------
Eg TGCTGACTAAGT----------
Ss TGCTGACTAAGTAATGATGATAATTGTATAGGTTGGTGT----------
Sc TGCTGACTAAG----------
Ta TGCTGACTAAGT----------
CD TGCTGACTAAGTAATGATGATAATTGTATAGGTTGGTGTTTTCACTGAATTCAGCGGCACT
Ka TGCTGACTAAGTAATGATGATA----------
Exon 5

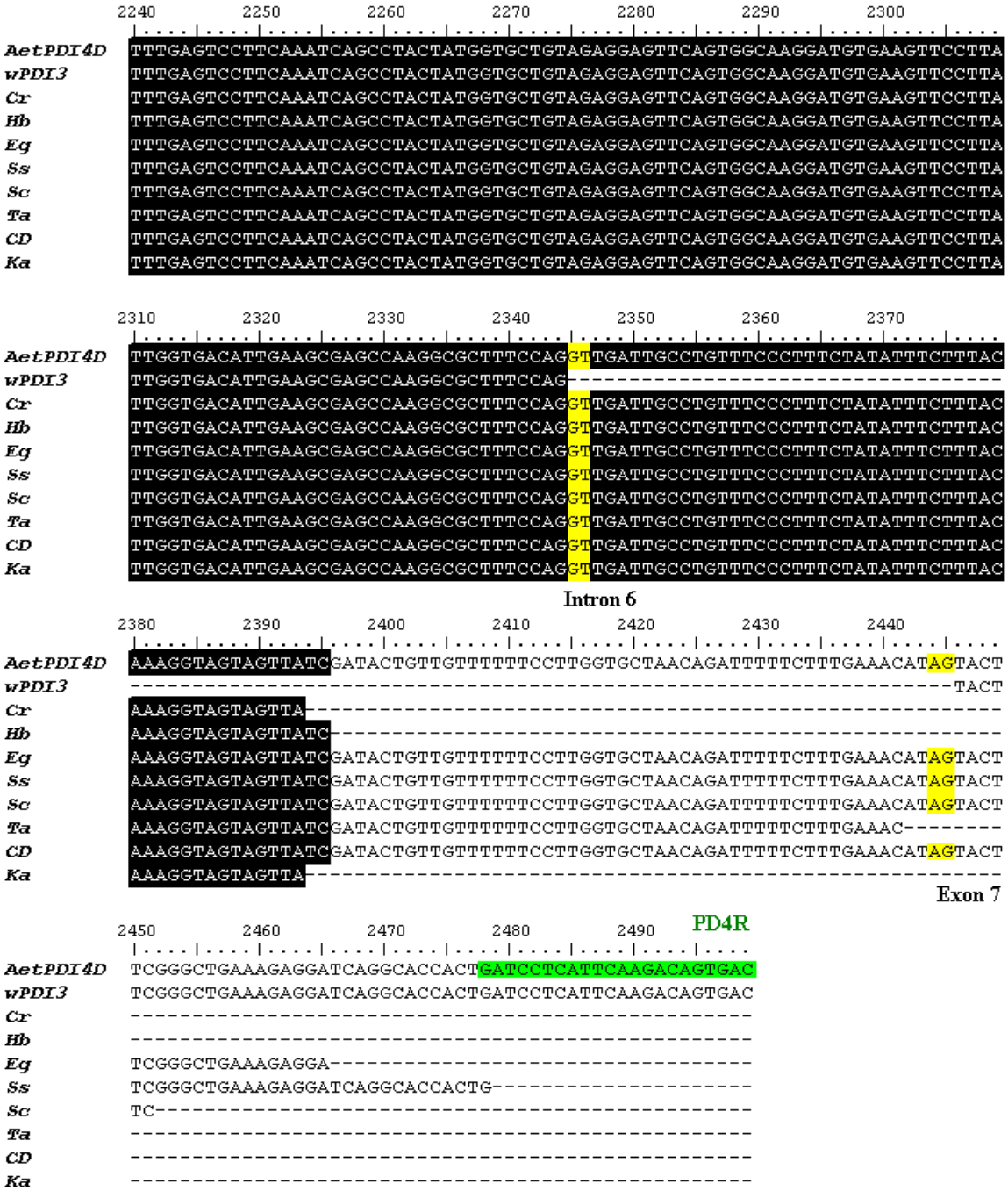

PB3F
1100 1110 1120 1130 1140 1150 1160
AetPDI4D TTGTATAGGTTGGTGTTTTCACTGAATTCAGCGGCACTGAGTTTACAAACTTCCTTGAGGTTGCTGAGAA
wPDI3 --------GTTGGTGTTTTCACTGAATTCAGCGGCACTGAGTTTACAAACTTCCTTGAGGTTGCTGAGAA
Cx ----------AGCGGCACTGAGTTTACAAACTTCCTTGAGGTTGCTGAGAA
Hb ----------GGCACTGAGTTTACAAACTTCCTTGAGGTTGCTGAGAA
Eg ----------TGAGTTTACAAACTTCCTTGAGGTTGCTGAGAA
Ss TTGTATAGGTTGGTGTTTTCACTGAATTCAGCGGCACTGAGTTTACAAACTTCCTTGAGGTTGCTGAGAA
Sc ----------GTTTACAAACTTCCTTGAGGTTGCTGAGAA
Ta ----------CCTTGAGGTTGCTGAGAA
CD ----------GAATTCAGCGGCACTGAGTTTACAAACTTCCTTGAGGTTGCTGAGAA
Ka ----------AACTTCCTTGAGGTTGCTGAGAA
Intron 3 Exon 4
1170 1180 1190 1200 1210 1220 1230
AetPDI4D GCTGAGGTCTGATTATGACTTTGGCCACACCGTGCATGCCAACCATCTCCCACGTGGTGATGCAGCAGTG
wPDI3 GCTGAGGTCTGATTATGACTTTGGCCACACCGTGCATGCCAACCATCTCCCACGTGGTGATGCAGCAGTG
Cx GCTGAGGTCTGATTATGACTTTGGCCACACCGTGCATGCCAACCATCTCCCACGTGGTGATGCAGCAGTG
Hb GCTGAGGTCTGATTATGACTTTGGCCACACCGTGCATGCCAACCATCTCCCACGTGGTGATGCAGCAGTG
Eg GCTGAGGTCTGATTATGACTTTGGCCACACCGTGCATGCCAACCATCTCCCACGTGGTGATGCAGCAGTG
Ss GCTGAGGTCTGATTATGACTTTGGCCACACCGTGCATGCCAACCATCTCCCACGTGGTGATGCAGCAGTG
Sc GCTGAGGTCTGATTATGACTTTGGCCACACCGTGCATGCCAACCATCTCCCACGTGGTGATGCAGCAGTG
Ta GCTGAGGTCTGATTATGACTTTGGCCACACCGTGCATGCCAACCATCTCCCACGTGGTGATGCAGCAGTG
CD GCTGAGGTCTGATTATGACTTTGGCCACACCGTGCATGCCAACCATCTCCCACGTGGTGATGCAGCAGTG
Ka GCTGAGGTCTGATTATGACTTTGGCCACACCGTGCATGCCAACCATCTCCCACGTGGTGATGCAGCAGTG
1240 1250 1260 1270 1280 1290 1300
AetPDI4D GAGAGGCCATTGGTTAGGCTATTCAAGCCATTTGATGAGCTCGTTGTTGACAGCAAGGTTATACTTCGCT
wPDI3 GAGAGGCCATTGGTTAGGCTATTCAAGCCATTTGATGAGCTCGTTGTTGACAGCAAG----------
Cx GAGAGGCCATTGGTTAGGCTATTCAAGCCATTTGATGAGCTCGTTGTTGACAGCAAGGTTATACTTCGCT
Hb GAGAGGCCATTGGTTAGGCTATTCAAGCCATTTGATGAGCTCGTTGTTGACAGCAAGGTTATACTTCGCT
Eg GAGAGGCCATTGGTTAGGCTATTCAAGCCATTTGATGAGCTCGTTGTTGACAGCAAGGTTATACTTCGCT
Ss GAGAGGCCATTGGTTAGGCTATTCAAGCCATTTGATGAGCTCGTTGTTGACAGCAAGGTTATACTTCGCT
Sc GAGAGGCCATTGGTTAGGCTATTCAAGCCATTTGATGAGCTCGTTGTTGACAGCAAGGTTATACTTCGCT
Ta GAGAGGCCATTGGTTAGGCTATTCAAGCCATTTGATGAGCTCGTTGTTGACAGCAAGGTTATACTTCGCT
CD GAGAGGCCATTGGTTAGGCTATTCAAGCCATTTGATGAGCTCGTTGTTGACAGCAAGGTTATACTTCGCT
Ka GAGAGGCCATTGGTTAGGCTATTCAAGCCATTTGATGAGCTCGTTGTTGACAGCAAGGTTATACTTCGCT
Intron 4
1310 1320 1330 1340 1350 1360 1370
AetPDI4D CGCTCTGTAAACTAATCATTACTCCCTCTGTTCATATTAGTTATTGCTTATTTAGTACAATTTAGTGACA
wPDI3 ----------
Cx CGCTCTGTAAACTAATCATTACTCCCTCTGTTCATATTAGTTATTGCTTATTTAGTACAATTTAGTGACA
Hb CGCTCTGTAAACTAATCATTACTCCCTCTGTTCATATTAGTTATTGCTTATTTAGTACAATTTAGTGACA
Eg CGCTCTGTAAACTAATCATTACTCCCTCTGTTCATATTAGTTATTGCTTATTTAGTACAATTTAGTGACA
Ss CGCTCTGTAAACTAATCATTACTCCCTCTGTTCATATTAGTTATTGCTTATTTAGTACAATTTAGTGACA
Sc CGCTCTGTAAACTAATCATTACTCCCTCTGTTCATATTAGTTATTGCTTATTTAGTACAATTTAGTGACA
Ta CGCTCTGTAAACTAATCATTACTCCCTCTGTTCATATTAGTTATTGCTTATTTAGTACAATTTAGTGACA
CD CGCTCTGTAAACTAATCATTACTCCCTCTGTTCATATTAGTTATTGCTTATTTAGTACAATTTAGTGACA
Ka CGCTCTGTAAACTAATCATTACTCCCTCTGTTCATATTAGTTATTGCTTATTTAGTACAATTTAGTGACA
1380 1390 1400 1410 1420 1430 1440
AetPDI4D AGTAATCTGGAACAGAGGGAGTGTTCATCTGTGTTCTTCTTACTTCATACAGATTACTGAGTATTAACAT
wPDI3 ----------
Cx AGTAATCTGGAACAGAGGGAGTGTTCATCTGTGTTCTTCTTACTTCATACAGATTACTGAGTATTAACAT
Hb AGTAATCTGGAACAGAGGGAGTGTTCATCTGTGTTCTTCTTACTTCATACAGATTACTGAGTATTAACAT
Eg AGTAATCTGGAACAGAGGGAGTGTTCATCTGTGTTCTTCTTACTTCATACAGATTACTGAGTATTAACAT
Ss AGTAATCTGGAACAGAGGGAGTGTTCATCTGTGTTCTTCTTACTTCATACAGATTACTGAGTATTAACAT
Sc AGTAATCTGGAACAGAGGGAGTGTTCATCTGTGTTCTTCTTACTTCATACAGATTACTGAGTATTAACAT
Ta AGTAATCTGGAACAGAGGGAGTGTTCATCTGTGTTCTTCTTACTTCATACAGATTACTGAGTATTAACAT
CD AGTAATCTGGAACAGAGGGAGTGTTCATCTGTGTTCTTCTTACTTCATACAGATTACTGAGTATTAACAT
Ka AGTAATCTGGAACAGAGGGAGTGTTCATCTGTGTTCTTCTTACTTCATACAGATTACTGAGTATTAACAT

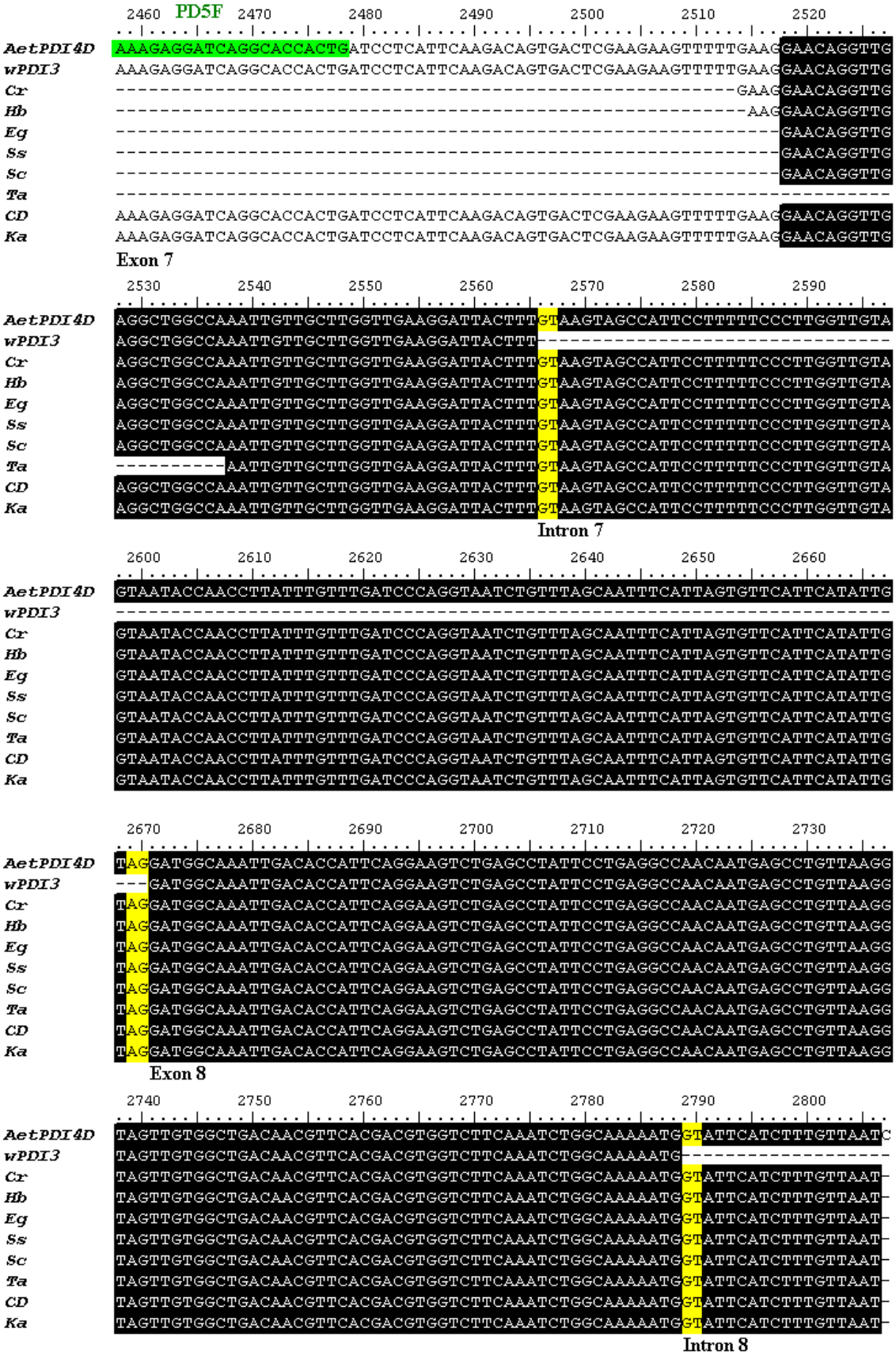

1450
1460
1470
1480
1490
1500
1510
AetPDI4D
wPDI3
Cr
Hb
Eg
Ss
Sc
Ta
CD
Ka
Exon 5
1520
1530
1540
1550
1560
1570
PD3R
1580

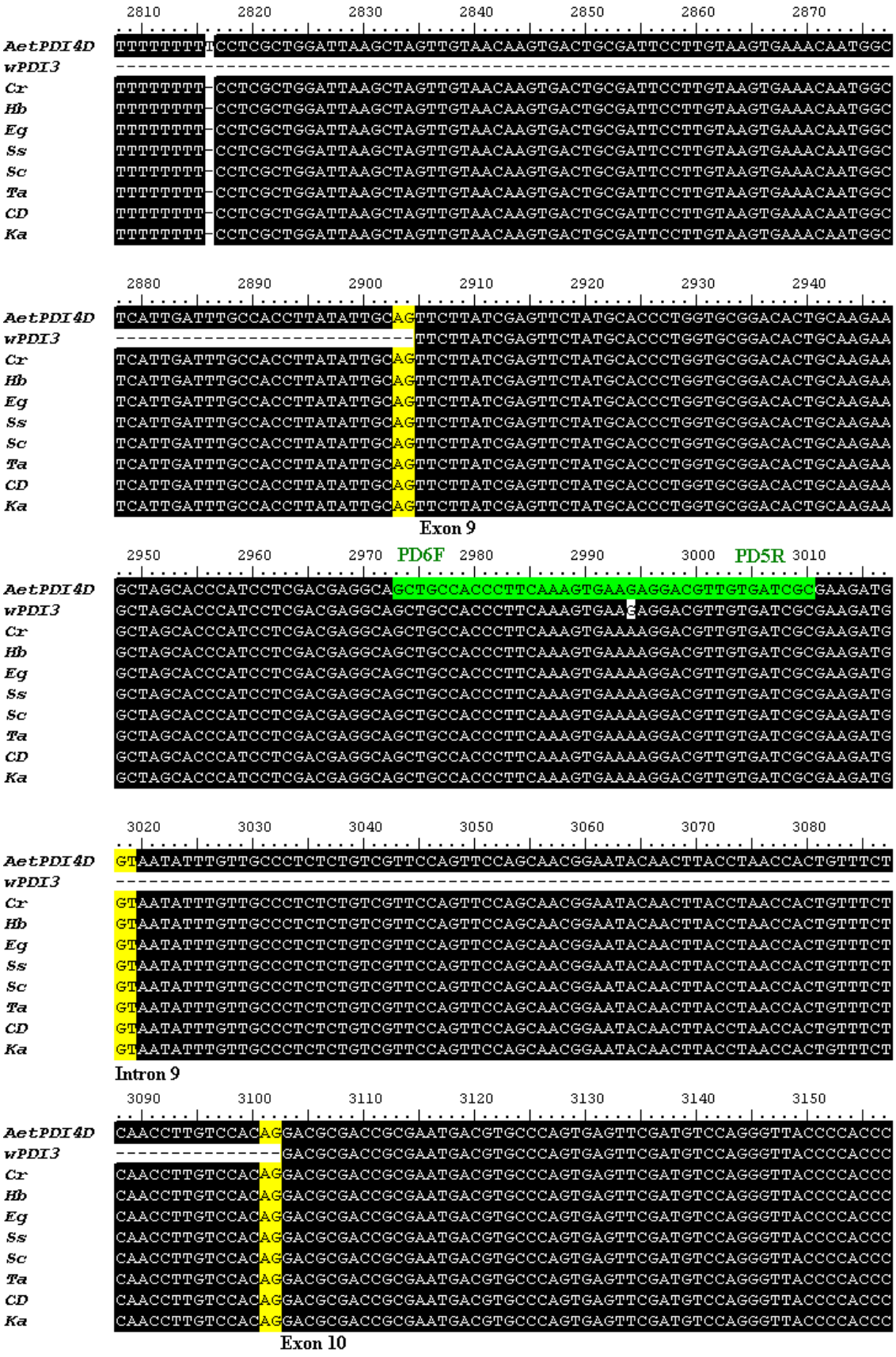

PD4F
Exon 5
Intron 5
AetPDI4D
wPDI3
Cx
Hb
Eg
Ss
Sc
Ta
CD
Ka


1890 1900 1910 1920 1930 1940 1950
AetPDI4D AAATCATCTTGCAAAGAATTTGTTCAGCACATTTGGAGGATTACTACTGTATTTAGTTTTGTTAGCCTTT
wPDI3 ----------
Cr AAATCATCTTGCAAAGAATTTGTTCAGCACATTTGGAGGATTACTAGTGTATTTAGTTTTGTTAGCCTTT
Hb AAATCATCTTGCAAAGAATTTGTTCAGCACATTTGGAGGATTACTAGTGTATTTAGTTTTGTTAGCCTTT
Eg AAATCATCTTGCAAAGAATTTGTTCAGCACATTTGGAGGATTACTAGTGTATTTAGTTTTGTTAGCCTTT
Ss AAATCATCTTGCAAAGAATTTGTTCAGCACATTTGGAGGATTACTAGTGTATTTAGTTTTGTTAGCCTTT
Sc AAATCATCTTGCAAAGAATTTGTTCAGCACATTTGGAGGATTACTAGTGTATTTAGTTTTGTTAGCCTTT
Ta AAATCATCTTGCAAAGAATTTGTTCAGCACATTTGGAGGATTACTAGTGTATTTAGTTTTGTTAGCCTTT
CD AAATCATCTTGCAAAGAATTTGTTCAGCACATTTGGAGGATTACTAGTGTATTTAGTTTTGTTAGCCTTT
Ka AAATCATCTTGCAAAGAATTTGTTCAGCACATTTGGAGGATTACTAGTGTATTTAGTTTTGTTAGCCTTT
1960 1970 1980 1990 2000 2010 2020
AetPDI4D TGGATCTTTGTGTCTTCATTCATGGTTGTTGTACCTGAAAGGTCTACTACTGCCGAGCCATAGATGTGAT
wPDI3 ----------
Cr TGGATCTTTGTGTTTTCATTCATGGTTGTTGTACCTGAAAGGTCTACTACTGCAGAGCCATAGATGTGAT
Hb TGGATCTTTGTGTTTTCATTCATGGTTGTTGTACCTGAAAGGTCTACTACTGCAGAGCCATAGATGTGAT
Eg TGGATCTTTGTGTTTTCATTCATGGTTGTTGTACCTGAAAGGTCTACTACTGCAGAGCCATAGATGTGAT
Ss TGGATCTTTGTGTTTTCATTCATGGTTGTTGTACCTGAAAGGTCTACTACTGCAGAGCCATAGATGTGAT
Sc TGGATCTTTGTGTTTTCATTCATGGTTGTTGTACCTGAAAGGTCTACTACTGCAGAGCCATAGATGTGAT
Ta TGGATCTTTGTGTTTTCATTCATGGTTGTTGTACCTGAAAGGTCTACTACTGCAGAGCCATAGATGTGAT
CD TGGATCTTTGTGTTTTCATTCATGGTTGTTGTACCTGAAAGGTCTACTACTGCAGAGCCATAGATGTGAT
Ka TGGATCTTTGTGTTTTCATTCATGGTTGTTGTACCTGAAAGGTCTACTACTGCAGAGCCATAGATGTGAT
2030 2040 2050 2060 2070 2080 2090
AetPDI4D GATTGTTTGAGATGCAGAATATCATCTTTAGTGTTTTCAGTGCTTTGTCTGTAAAAACAGTTCAGATATA
wPDI3 ----------
Cr GATTGTTTGAGATGCAGAATATCATCTTTGGTGTTTTCAGTGCTTTGTCTGTAAAAACAGTTCAGATATA
Hb GATTGTTTGAGATGCAGAATATCATCTTTGGTGTTTTCAGTGCTTTGTCTGTAAAAACAGTTCAGATATA
Eg GATTGTTTGAGATGCAGAATATCATCTTTGGTGTTTTCAGTGCTTTGTCTGTAAAAACAGTTCAGATATA
Ss GATTGTTTGAGATGCAGAATATCATCTTTGGTGTTTTCAGTGCTTTGTCTGTAAAAACAGTTCAGATATA
Sc GATTGTTTGAGATGCAGAATATCATCTTTGGTGTTTTCAGTGCTTTGTCTGTAAAAACAGTTCAGATATA
Ta GATTGTTTGAGATGCAGAATATCATCTTTGGTGTTTTCAGTGCTTTGTCTGTAAAAACAGTTCAGATATA
CD GATTGTTTGAGATGCAGAATATCATCTTTGGTGTTTTCAGTGCTTTGTCTGTAAAAACAGTTCAGATATA
Ka GATTGTTTGAGATGCAGAATATCATCTTTGGTGTTTTCAGTGCTTTGTCTGTAAAAACAGTTCAGATATA
2100 2110 2120 2130 2140 2150 2160
AetPDI4D TGTTTCTAGTGGTGAAGCTGCTTAGAGCATTAGCCTTTTCATTGTTTTTTAGTTATTGGTCACTCATTAT
wPDI3 ----------
Cr TGTTTCTAGTGGTGAAGCTGCTTAGAGCATTAGCCTTTTCATTGTTTTTTAGTTATTGGTCACTCATTAT
Hb TGTTTCTAGTGGTGAAGCTGCTTAGAGCATTAGCCTTTTCATTGTTTTTTAGTTATTGGTCACTCATTAT
Eg TGTTTCTAGTGGTGAAGCTGCTTAGAGCATTAGCCTTTTCATTGTTTTTTAGTTATTGGTCACTCATTAT
Ss TGTTTCTAGTGGTGAAGCTGCTTAGAGCATTAGCCTTTTCATTGTTTTTTAGTTATTGGTCACTCATTAT
Sc TGTTTCTAGTGGTGAAGCTGCTTAGAGCATTAGCCTTTTCATTGTTTTTTAGTTATTGGTCACTCATTAT
Ta TGTTTCTAGTGGTGAAGCTGCTTAGAGCATTAGCCTTTTCATTGTTTTTTAGTTATTGGTCACTCATTAT
CD TGTTTCTAGTGGTGAAGCTGCTTAGAGCATTAGCCTTTTCATTGTTTTTTAGTTATTGGTCACTCATTAT
Ka TGTTTCTAGTGGTGAAGCTGCTTAGAGCATTAGCCTTTTCATTGTTTTTTAGTTATTGGTCACTCATTAT
2170 2180 2190 2200 2210 2220 2230
AetPDI4D TTGGCACTTCTGTTCTCTGACATCAAATGGTATGTAGGCCATGCTCTTTTTTGAACTTCTCCACTGGACCG
wPDI3 ----------GCCATGCTCTTTTTTGAACTTCTCCACTGGACCG
Cr TTGGCACTTCTGTTCTCTGACATCAAATGGTATGTAGGCCATGCTCTTTTTTGAACTTCTCCACTGGACCG
Hb TTGGCACTTCTGTTCTCTGACATCAAATGGTATGTAGGCCATGCTCTTTTTTGAACTTCTCCACTGGACCG
Eg TTGGCACTTCTGTTCTCTGACATCAAATGGTATGTAGGCCATGCTCTTTTTTGAACTTCTCCACTGGACCG
Ss TTGGCACTTCTGTTCTCTGACATCAAATGGTATGTAGGCCATGCTCTTTTTTGAACTTCTCCACTGGACCG
Sc TTGGCACTTCTGTTCTCTGACATCAAATGGTATGTAGGCCATGCTCTTTTTTGAACTTCTCCACTGGACCG
Ta TTGGCACTTCTGTTCTCTGACATCAAATGGTATGTAGGCCATGCTCTTTTTTGAACTTCTCCACTGGACCG
CD TTGGCACTTCTGTTCTCTGACATCAAATGGTATGTAGGCCATGCTCTTTTTTGAACTTCTCCACTGGACCG
Ka TTGGCACTTCTGTTCTCTGACATCAAATGGTATGTAGGCCATGCTCTTTTTTGAACTTCTCCACTGGACCG

Exon 6


2240 2250 2260 2270 2280 2290 2300
AetPDI4D TTTGAGTCCTTCAAATCAGCCTACTATGGTGCTGTAGAGGAGTTCAGTGGCAAGGATGTGAAGTTCCTTA
wPDI3 TTTGAGTCCTTCAAATCAGCCTACTATGGTGCTGTAGAGGAGTTCAGTGGCAAGGATGTGAAGTTCCTTA
Cx TTTGAGTCCTTCAAATCAGCCTACTATGGTGCTGTAGAGGAGTTCAGTGGCAAGGATGTGAAGTTCCTTA
Hb TTTGAGTCCTTCAAATCAGCCTACTATGGTGCTGTAGAGGAGTTCAGTGGCAAGGATGTGAAGTTCCTTA
Eg TTTGAGTCCTTCAAATCAGCCTACTATGGTGCTGTAGAGGAGTTCAGTGGCAAGGATGTGAAGTTCCTTA
Ss TTTGAGTCCTTCAAATCAGCCTACTATGGTGCTGTAGAGGAGTTCAGTGGCAAGGATGTGAAGTTCCTTA
Sc TTTGAGTCCTTCAAATCAGCCTACTATGGTGCTGTAGAGGAGTTCAGTGGCAAGGATGTGAAGTTCCTTA
Ta TTTGAGTCCTTCAAATCAGCCTACTATGGTGCTGTAGAGGAGTTCAGTGGCAAGGATGTGAAGTTCCTTA
CD TTTGAGTCCTTCAAATCAGCCTACTATGGTGCTGTAGAGGAGTTCAGTGGCAAGGATGTGAAGTTCCTTA
Ka TTTGAGTCCTTCAAATCAGCCTACTATGGTGCTGTAGAGGAGTTCAGTGGCAAGGATGTGAAGTTCCTTA
2310 2320 2330 2340 2350 2360 2370
AetPDI4D TTGGTGACATTGAAGCGAGCCAAGGCGCTTTCCAGGTTGATTGCCTGTTTCCCTTTCTATATTTCTTTAC
wPDI3 TTGGTGACATTGAAGCGAGCCAAGGCGCTTTCCAG----------
Cx TTGGTGACATTGAAGCGAGCCAAGGCGCTTTCCAGGTTGATTGCCTGTTTCCCTTTCTATATTTCTTTAC
Hb TTGGTGACATTGAAGCGAGCCAAGGCGCTTTCCAGGTTGATTGCCTGTTTCCCTTTCTATATTTCTTTAC
Eg TTGGTGACATTGAAGCGAGCCAAGGCGCTTTCCAGGTTGATTGCCTGTTTCCCTTTCTATATTTCTTTAC
Ss TTGGTGACATTGAAGCGAGCCAAGGCGCTTTCCAGGTTGATTGCCTGTTTCCCTTTCTATATTTCTTTAC
Sc TTGGTGACATTGAAGCGAGCCAAGGCGCTTTCCAGGTTGATTGCCTGTTTCCCTTTCTATATTTCTTTAC
Ta TTGGTGACATTGAAGCGAGCCAAGGCGCTTTCCAGGTTGATTGCCTGTTTCCCTTTCTATATTTCTTTAC
CD TTGGTGACATTGAAGCGAGCCAAGGCGCTTTCCAGGTTGATTGCCTGTTTCCCTTTCTATATTTCTTTAC
Ka TTGGTGACATTGAAGCGAGCCAAGGCGCTTTCCAGGTTGATTGCCTGTTTCCCTTTCTATATTTCTTTAC
Intron 6
2380 2390 2400 2410 2420 2430 2440
AetPDI4D AAAGGTAGTAGTTATCGATACTGTTGTTTTTTCCTTGGTGCTAACAGATTTTTCTTTGAAACATAGTACT
wPDI3 ----------TACT
Cx AAAGGTAGTAGTTA----------
Hb AAAGGTAGTAGTTATC----------
Eg AAAGGTAGTAGTTATCGATACTGTTGTTTTTTCCTTGGTGCTAACAGATTTTTCTTTGAAACATAGTACT
Ss AAAGGTAGTAGTTATCGATACTGTTGTTTTTTCCTTGGTGCTAACAGATTTTTCTTTGAAACATAGTACT
Sc AAAGGTAGTAGTTATCGATACTGTTGTTTTTTCCTTGGTGCTAACAGATTTTTCTTTGAAACATAGTACT
Ta AAAGGTAGTAGTTATCGATACTGTTGTTTTTTCCTTGGTGCTAACAGATTTTTCTTTGAAAC--------
CD AAAGGTAGTAGTTATCGATACTGTTGTTTTTTCCTTGGTGCTAACAGATTTTTCTTTGAAACATAGTACT
Ka AAAGGTAGTAGTTA----------
Exon 7
PD4R
2450 2460 2470 2480 2490
AetPDI4D TCGGGCTGAAAGAGGATCAGGCACCACTGATCCTCATTCAAGACAGTGAC
wPDI3 TCGGGCTGAAAGAGGATCAGGCACCACTGATCCTCATTCAAGACAGTGAC
Cx ----------
Hb ----------
Eg TCGGGCTGAAAGAGGA----------
Ss TCGGGCTGAAAGAGGATCAGGCACCACTG----------
Sc TC----------
Ta ----------
CD ----------
Ka ----------


PD5F
AetPDI4D
wPDI3
Cr
Hb
Eg
Ss
Sc
Ta
CD
Ka
Exon 7
Intron 7
Exon 8
Intron 8


2810 2820 2830 2840 2850 2860 2870
AetPDI4D TTTTTTTTTCCTCGCTGGATTAAGCTAGTTGTAACAAGTGACTGCGATTCCTTGTAAGTGAAACAATGGC
wPDI3 ----------
Cr TTTTTTTT-CCTCGCTGGATTAAGCTAGTTGTAACAAGTGACTGCGATTCCTTGTAAGTGAAACAATGGC
Hb TTTTTTTT-CCTCGCTGGATTAAGCTAGTTGTAACAAGTGACTGCGATTCCTTGTAAGTGAAACAATGGC
Eg TTTTTTTT-CCTCGCTGGATTAAGCTAGTTGTAACAAGTGACTGCGATTCCTTGTAAGTGAAACAATGGC
Ss TTTTTTTT-CCTCGCTGGATTAAGCTAGTTGTAACAAGTGACTGCGATTCCTTGTAAGTGAAACAATGGC
Sc TTTTTTTT-CCTCGCTGGATTAAGCTAGTTGTAACAAGTGACTGCGATTCCTTGTAAGTGAAACAATGGC
Ta TTTTTTTT-CCTCGCTGGATTAAGCTAGTTGTAACAAGTGACTGCGATTCCTTGTAAGTGAAACAATGGC
CD TTTTTTTT-CCTCGCTGGATTAAGCTAGTTGTAACAAGTGACTGCGATTCCTTGTAAGTGAAACAATGGC
Ka TTTTTTTT-CCTCGCTGGATTAAGCTAGTTGTAACAAGTGACTGCGATTCCTTGTAAGTGAAACAATGGC
2880 2890 2900 2910 2920 2930 2940
AetPDI4D TCATTGATTTGCCACCTTATATTGCAGTTCTTATCGAGTTCTATGCACCCTGGTGCGGACACTGCAAGAA
wPDI3 ----------TTCTTATCGAGTTCTATGCACCCTGGTGCGGACACTGCAAGAA
Cr TCATTGATTTGCCACCTTATATTGCAGTTCTTATCGAGTTCTATGCACCCTGGTGCGGACACTGCAAGAA
Hb TCATTGATTTGCCACCTTATATTGCAGTTCTTATCGAGTTCTATGCACCCTGGTGCGGACACTGCAAGAA
Eg TCATTGATTTGCCACCTTATATTGCAGTTCTTATCGAGTTCTATGCACCCTGGTGCGGACACTGCAAGAA
Ss TCATTGATTTGCCACCTTATATTGCAGTTCTTATCGAGTTCTATGCACCCTGGTGCGGACACTGCAAGAA
Sc TCATTGATTTGCCACCTTATATTGCAGTTCTTATCGAGTTCTATGCACCCTGGTGCGGACACTGCAAGAA
Ta TCATTGATTTGCCACCTTATATTGCAGTTCTTATCGAGTTCTATGCACCCTGGTGCGGACACTGCAAGAA
CD TCATTGATTTGCCACCTTATATTGCAGTTCTTATCGAGTTCTATGCACCCTGGTGCGGACACTGCAAGAA
Ka TCATTGATTTGCCACCTTATATTGCAGTTCTTATCGAGTTCTATGCACCCTGGTGCGGACACTGCAAGAA
Exon 9
2950 2960 2970 PD6F 2980 2990 3000 PD5R 3010
AetPDI4D GCTAGCACCCATCCTCGACGAGGCAGCTGCCACCCTTCAAAGTGAAGAGGACGTTGTGATCGCGAAGATG
wPDI3 GCTAGCACCCATCCTCGACGAGGCAGCTGCCACCCTTCAAAGTGAAGAGGACGTTGTGATCGCGAAGATG
Cr GCTAGCACCCATCCTCGACGAGGCAGCTGCCACCCTTCAAAGTGAAAAGGACGTTGTGATCGCGAAGATG
Hb GCTAGCACCCATCCTCGACGAGGCAGCTGCCACCCTTCAAAGTGAAAAGGACGTTGTGATCGCGAAGATG
Eg GCTAGCACCCATCCTCGACGAGGCAGCTGCCACCCTTCAAAGTGAAAAGGACGTTGTGATCGCGAAGATG
Ss GCTAGCACCCATCCTCGACGAGGCAGCTGCCACCCTTCAAAGTGAAAAGGACGTTGTGATCGCGAAGATG
Sc GCTAGCACCCATCCTCGACGAGGCAGCTGCCACCCTTCAAAGTGAAAAGGACGTTGTGATCGCGAAGATG
Ta GCTAGCACCCATCCTCGACGAGGCAGCTGCCACCCTTCAAAGTGAAAAGGACGTTGTGATCGCGAAGATG
CD GCTAGCACCCATCCTCGACGAGGCAGCTGCCACCCTTCAAAGTGAAAAGGACGTTGTGATCGCGAAGATG
Ka GCTAGCACCCATCCTCGACGAGGCAGCTGCCACCCTTCAAAGTGAAAAGGACGTTGTGATCGCGAAGATG
3020 3030 3040 3050 3060 3070 3080
AetPDI4D GTAATATTTGTTGCCCTCTCTGTCGTTCCAGTTCCAGCAACGGAATACAACTTACCTAACCACTGTTTCT
wPDI3 ----------
Cr GTAATATTTGTTGCCCTCTCTGTCGTTCCAGTTCCAGCAACGGAATACAACTTACCTAACCACTGTTTCT
Hb GTAATATTTGTTGCCCTCTCTGTCGTTCCAGTTCCAGCAACGGAATACAACTTACCTAACCACTGTTTCT
Eg GTAATATTTGTTGCCCTCTCTGTCGTTCCAGTTCCAGCAACGGAATACAACTTACCTAACCACTGTTTCT
Ss GTAATATTTGTTGCCCTCTCTGTCGTTCCAGTTCCAGCAACGGAATACAACTTACCTAACCACTGTTTCT
Sc GTAATATTTGTTGCCCTCTCTGTCGTTCCAGTTCCAGCAACGGAATACAACTTACCTAACCACTGTTTCT
Ta GTAATATTTGTTGCCCTCTCTGTCGTTCCAGTTCCAGCAACGGAATACAACTTACCTAACCACTGTTTCT
CD GTAATATTTGTTGCCCTCTCTGTCGTTCCAGTTCCAGCAACGGAATACAACTTACCTAACCACTGTTTCT
Ka GTAATATTTGTTGCCCTCTCTGTCGTTCCAGTTCCAGCAACGGAATACAACTTACCTAACCACTGTTTCT
Intron 9
3090 3100 3110 3120 3130 3140 3150
AetPDI4D CAACCTTGTCCACAGGACGCGACCGCGAATGACGTGCCCAGTGAGTTCGATGTCCAGGGTTACCCCACCC
wPDI3 ----------GACGCGACCGCGAATGACGTGCCCAGTGAGTTCGATGTCCAGGGTTACCCCACCC
Cr CAACCTTGTCCACAGGACGCGACCGCGAATGACGTGCCCAGTGAGTTCGATGTCCAGGGTTACCCCACCC
Hb CAACCTTGTCCACAGGACGCGACCGCGAATGACGTGCCCAGTGAGTTCGATGTCCAGGGTTACCCCACCC
Eg CAACCTTGTCCACAGGACGCGACCGCGAATGACGTGCCCAGTGAGTTCGATGTCCAGGGTTACCCCACCC
Ss CAACCTTGTCCACAGGACGCGACCGCGAATGACGTGCCCAGTGAGTTCGATGTCCAGGGTTACCCCACCC
Sc CAACCTTGTCCACAGGACGCGACCGCGAATGACGTGCCCAGTGAGTTCGATGTCCAGGGTTACCCCACCC
Ta CAACCTTGTCCACAGGACGCGACCGCGAATGACGTGCCCAGTGAGTTCGATGTCCAGGGTTACCCCACCC
CD CAACCTTGTCCACAGGACGCGACCGCGAATGACGTGCCCAGTGAGTTCGATGTCCAGGGTTACCCCACCC
Ka CAACCTTGTCCACAGGACGCGACCGCGAATGACGTGCCCAGTGAGTTCGATGTCCAGGGTTACCCCACCC
Exon 10

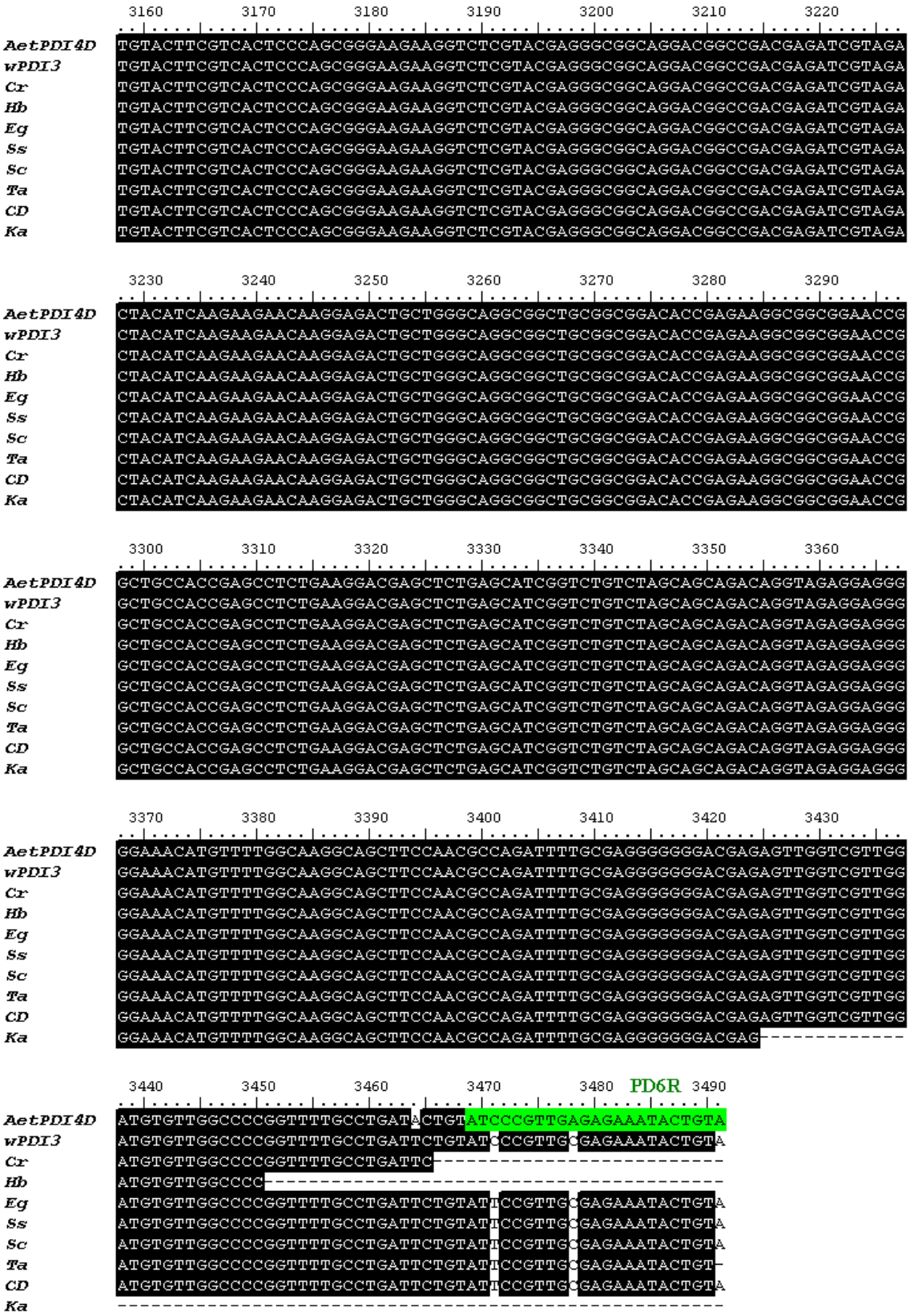


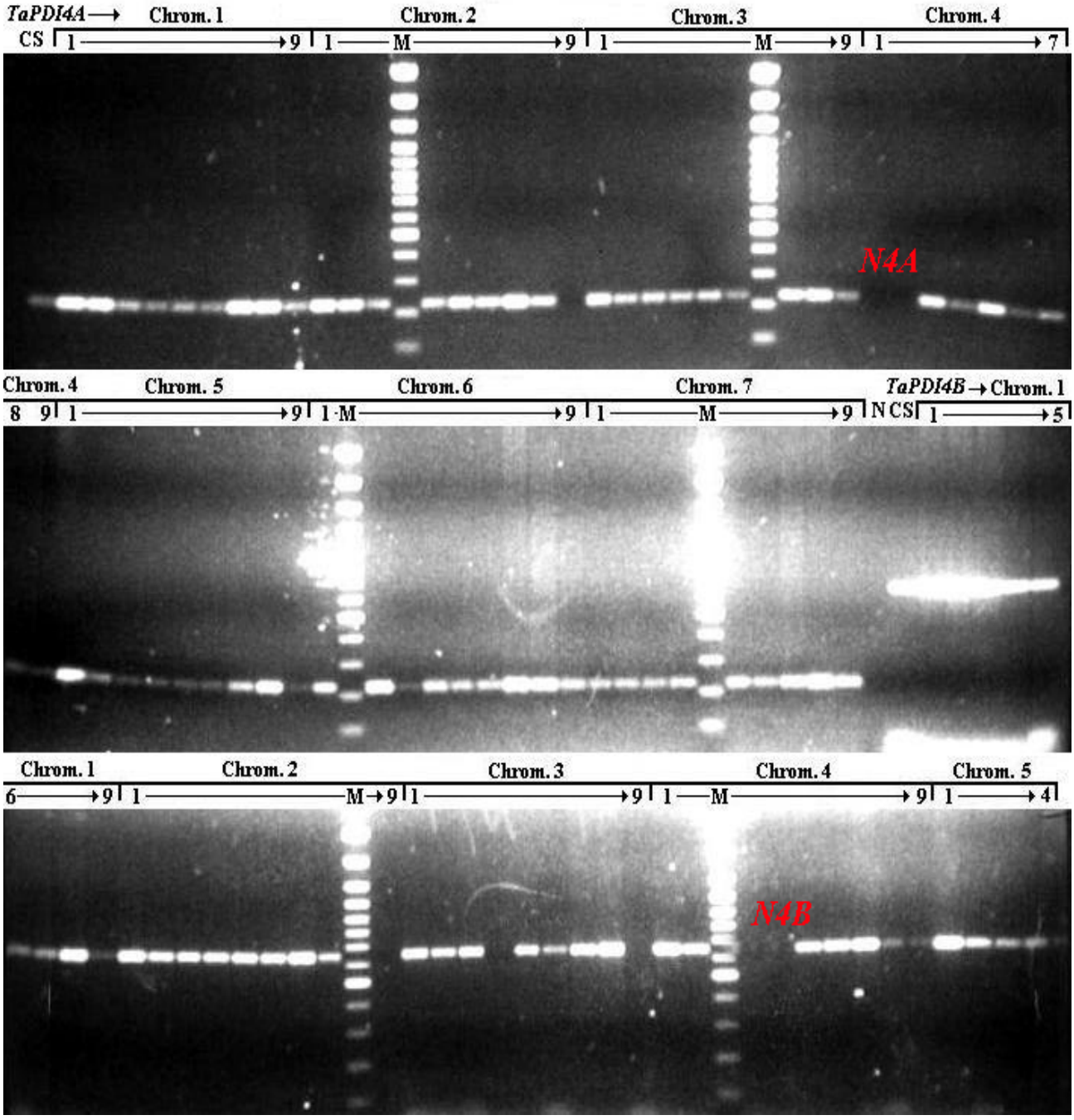
Figure IIID The partial sequence of the *PDI* genes in the D genome of wheat

Legend for Figures IIIA, IIIB and IIID

The DNA sequences of the eight cultivars of *T. aestivum* are shown aligned with the corresponding *PDI* cDNA and progenitor sequences (A genome = *wPDI1* and *TtPDI4A*; B genome = *wPDI2* and *TtPDI4B*; D genome = *wPDI3* and *AetPDI4D*). The numbering system relates to the nucleotide number in the progenitor sequence. Mapping cultivars: Cr: Cranbrook, Hb: Halberd, Eg: Egret, Ss: Sunstar, Sc: Sunco, Ta: Tasman, CD: CD87, Ka: Katepwa. Allele-specific primers (Table 2.6, Figure 2.4) used in the amplification and sequencing of these gene sections are highlighted in green with the name of the primer. Intron and Exons are shown and the ubiquitous GT/AG dinucleotide at the intron/exon boundaries are highlighted in yellow. The intercultivar polymorphisms identified between the *PDI* genes in the A and B genome are highlighted in red. The positions of these alignments in relation to the intron/exon structure of the *PDI* genes are shown in Figure 4.3.

APPENDIX IV

RESULTS OF THE NULLISOMIC/TETRASOMIC AND DITELOSOMIC ANALYSIS OF THE *PDI* AND *CYCLOPHILIN* GENES



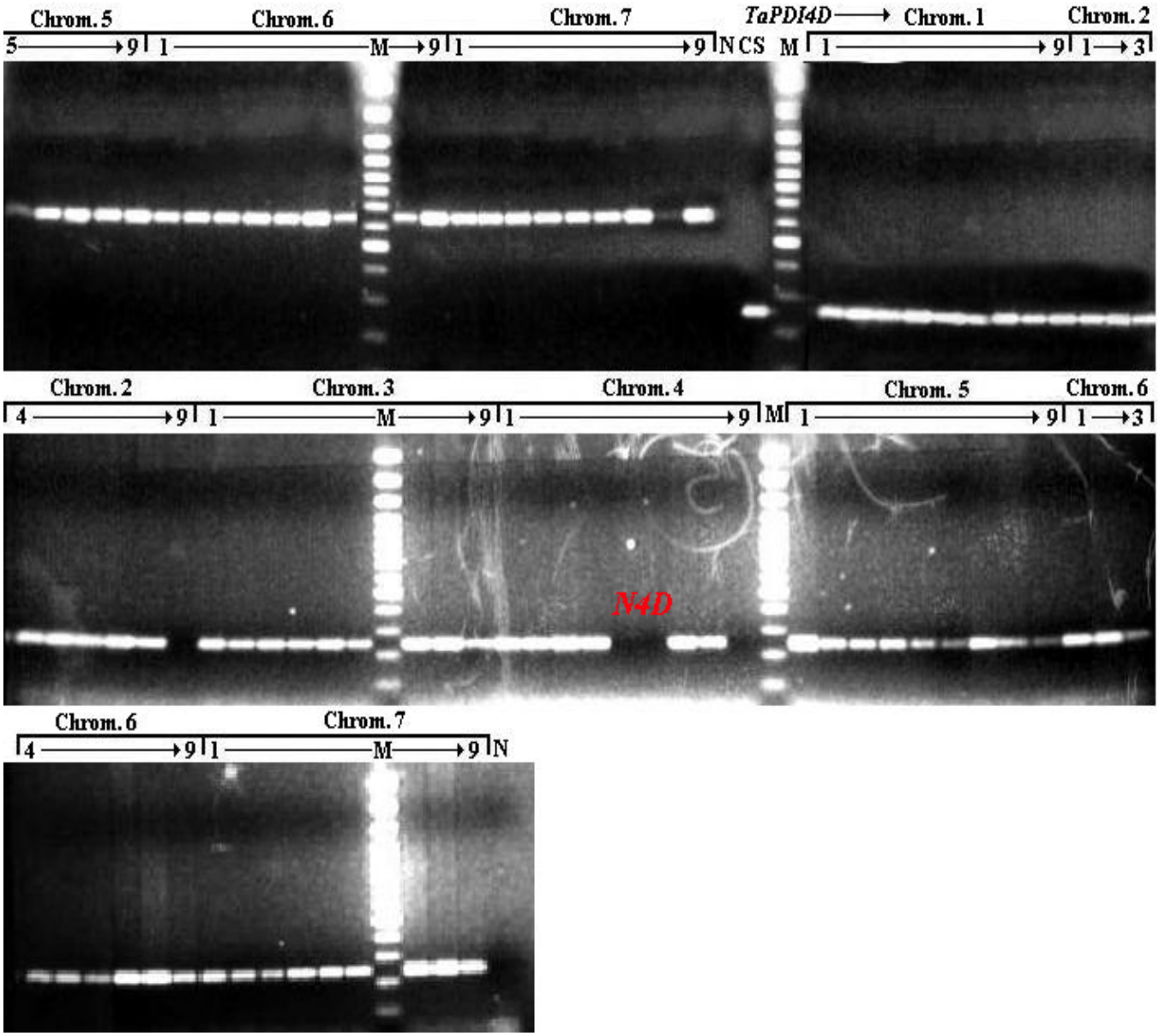
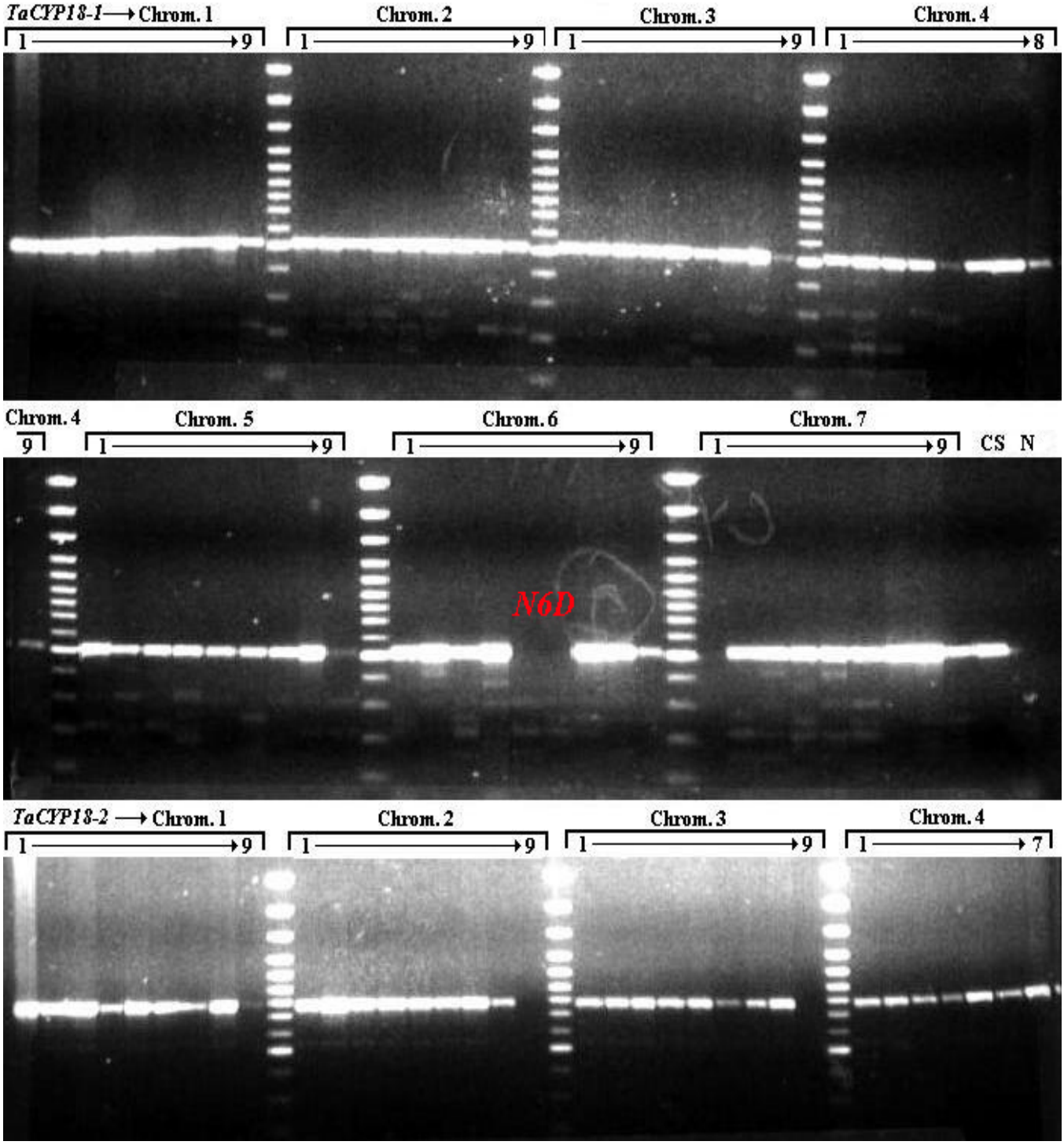


Figure IV-1 The N/T and ditelosomic analysis of the *PDI* genes of *T. aestivum*


TaCYP18-1 → Chrom. 1
1 → 9
Chrom. 2
1 → 9
Chrom. 3
1 → 9
Chrom. 4
1 → 8
Chrom. 4
9
Chrom. 5
1 → 9
Chrom. 6
1 → 9
Chrom. 7
1 → 9
CS N
N6D
TaCYP18-2 → Chrom. 1
1 → 9
Chrom. 2
1 → 9
Chrom. 3
1 → 9
Chrom. 4
1 → 7

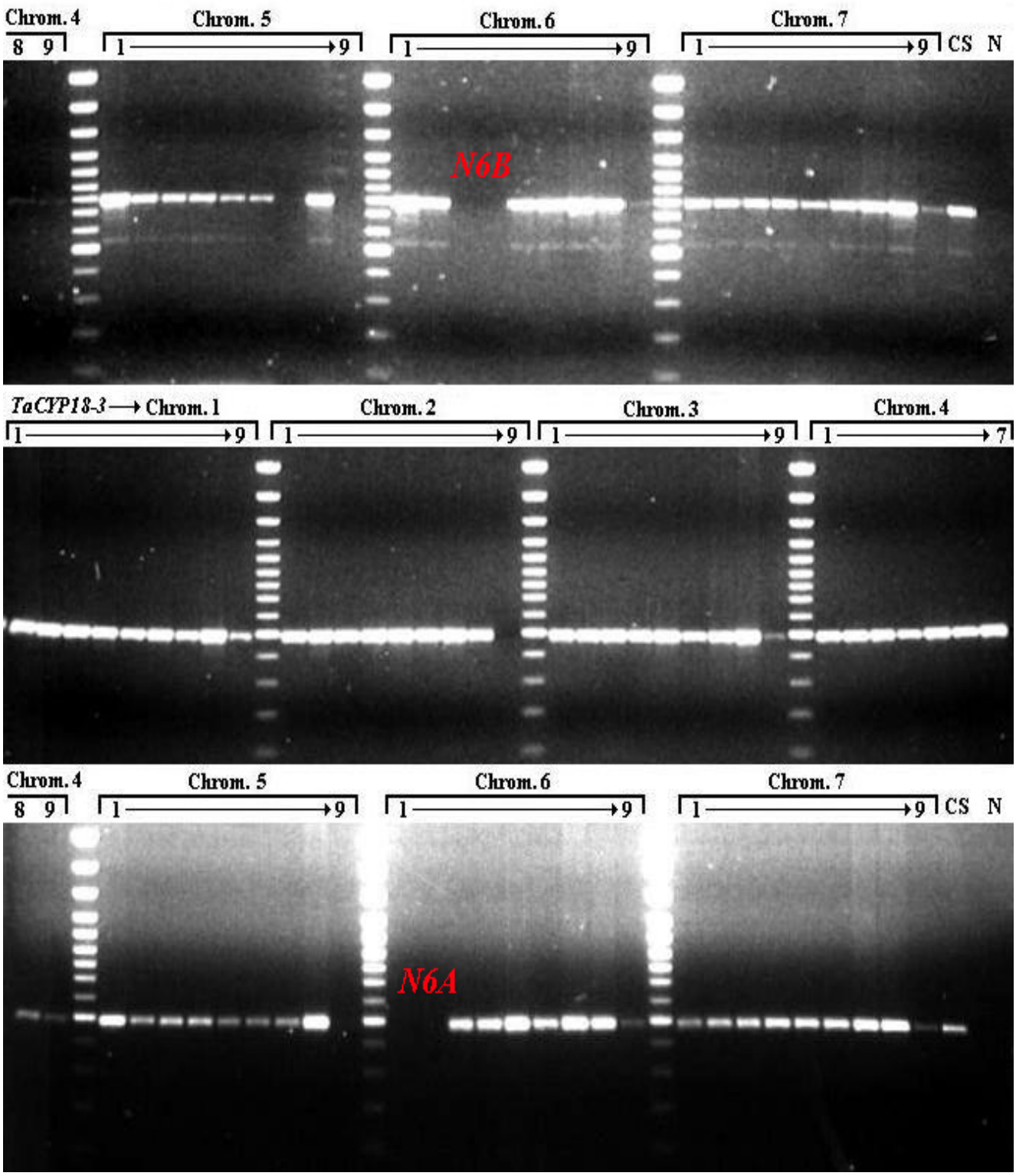


Figure IV-2 The N/T and ditelosomic analysis of the *cyclophilin* genes of *T. aestivum*

Legend for Figures IV-1 and IV-2

The partial genes were amplified with the *PDI* allele-specific (AS) primers PA1F/PA1R (*TaPDI4A*), PB3F/PB3R (*TaPDI4B*) and PD2F/PD2R (*TAPDI4D*) (Figure 2.4, Table 2.6) or the cyclophilin AS-primers C1F1/C1R1 (*TaCYP18-1*), C2F1/C2R1 (*TaCYP18-2*) and C3F1/C3R1 (*TaCYP18-3*) (Figure 2.5, Table 2.7). Template DNA:

Chrom.	Lane 1	Lane 2	Lane 3	Lane 4	Lane 5	Lane 6	Lane 7	Lane 8	Lane 9
1	1BBD	1BDD	1AAD	1ADD	1AAB	1ABB	DT1AL	DT1BL	DT1DL
2	2BBD	2BDD	2AAD	2ADD	2AAB	2ABB	DT2AL	DT2BL	DT2DL
3	3BBD	3BDD	3AAD	3ADD	3AAB	3ABB	DT3AL	DT3BL	DT3DL
4	4BBD	4BDD	4AAD	4ADD	4AAB	4ABB	DT4AL	DT4BL	DT4DL
5	5BBD	5BDD	5AAD	5ADD	5AAB	5ABB	DT5AL	DT5BL	DT5DL
6	6BBD	6BDD	6AAD	6ADD	6AAB	6ABB	DT6AL	DT6BL	DT6DL
7	7BBD	7BDD	7AAD	7ADD	7AAB	7ABB	DT7AL	DT7BL	DT7DL

CS: *T. aestivum* cv. Chinese Spring gDNA; N: no-template negative control; M: 100bp ladder molecular weight marker.

APPENDIX V

SOUTHERN BLOTS DATA FOR THE PUTATIVE WHEAT ORTHOLOGUES AT THE *esp2* LOCUS OF RICE

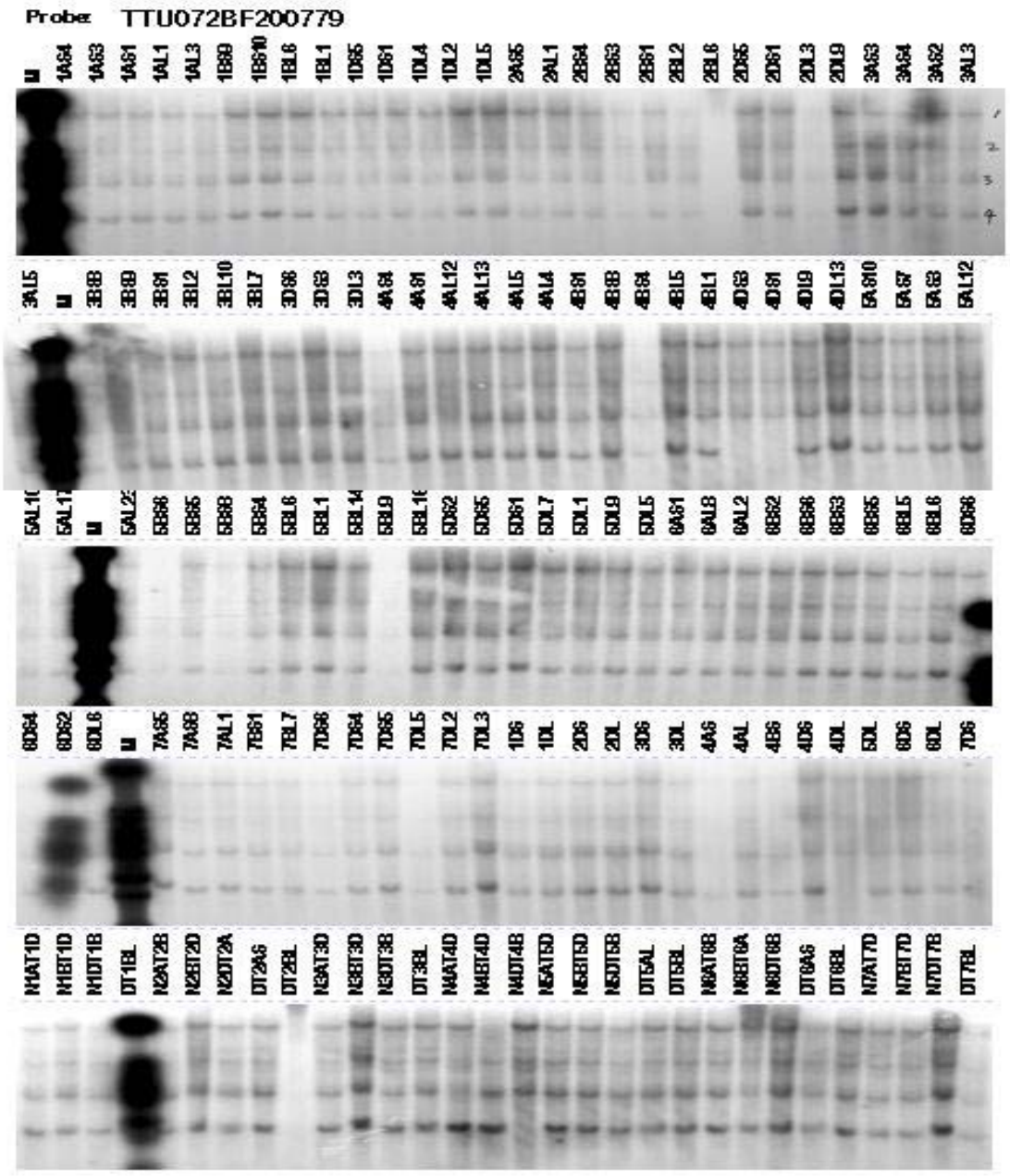


Figure V-1 The mapping data generated with probe BF200779 (Probe 1).

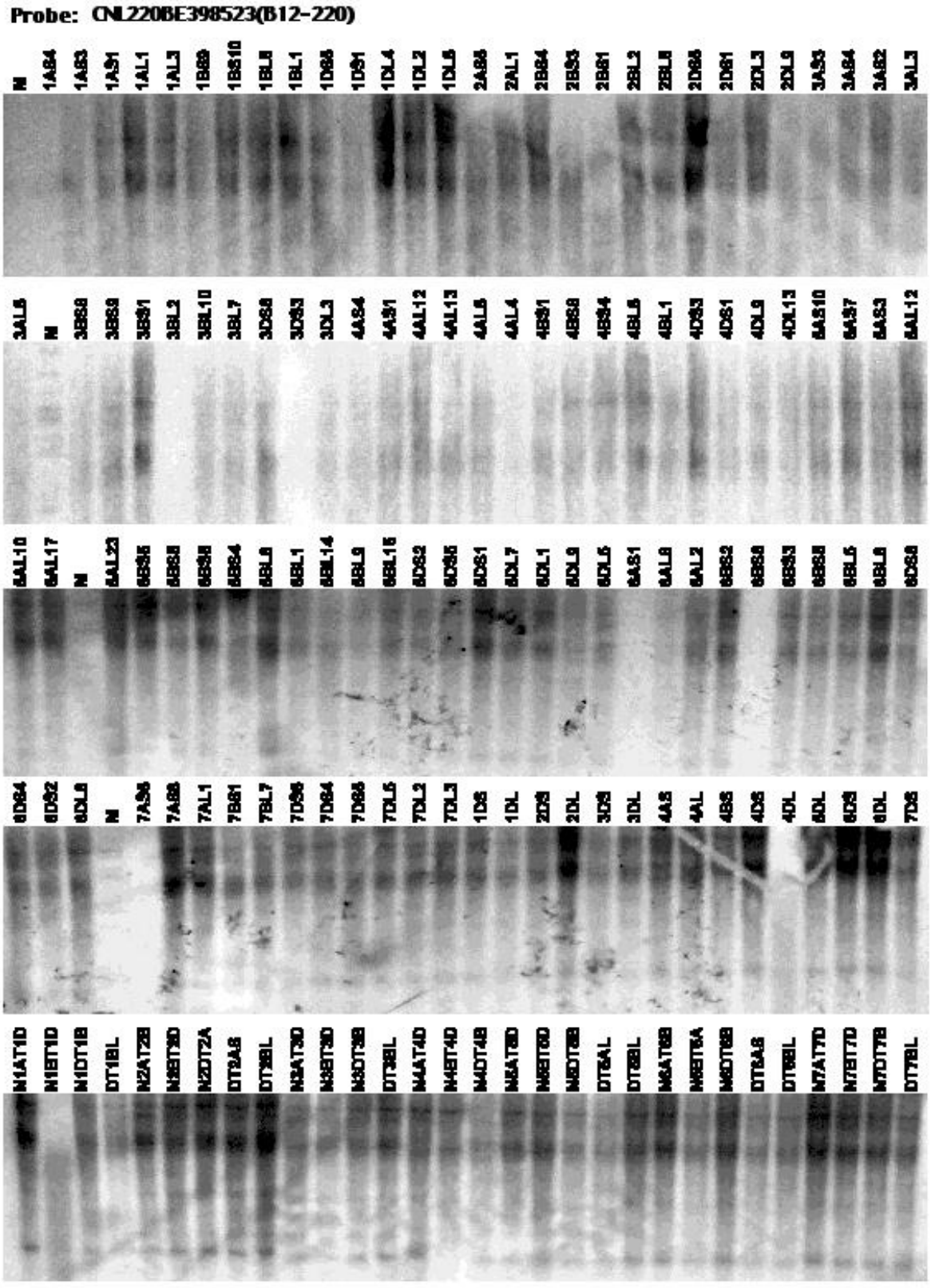


Figure V-2 The mapping data generated with probe BE398523 (Probe 2).

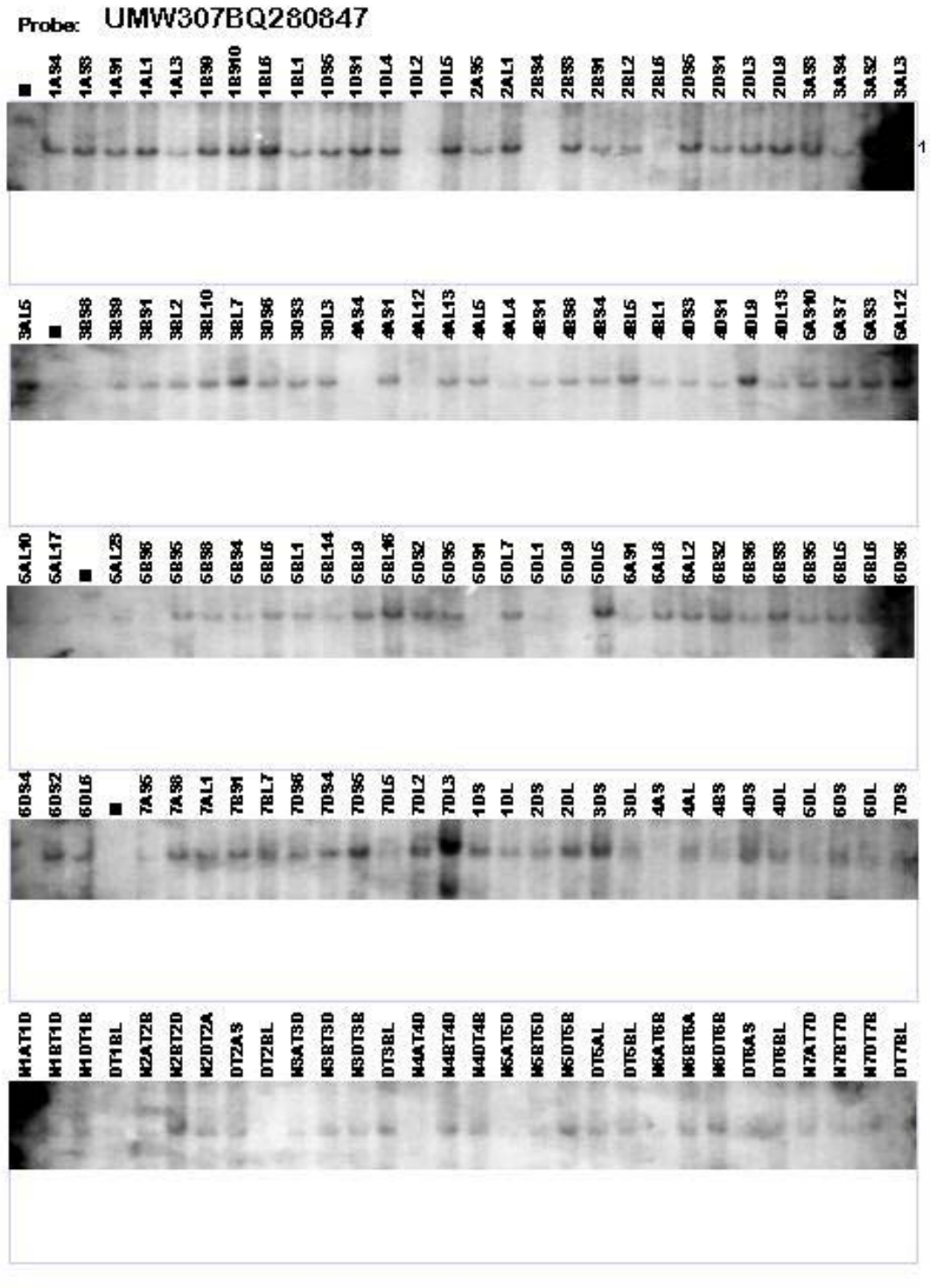


Figure V-3 The mapping data generated with probe BQ280847 (Probe 3).

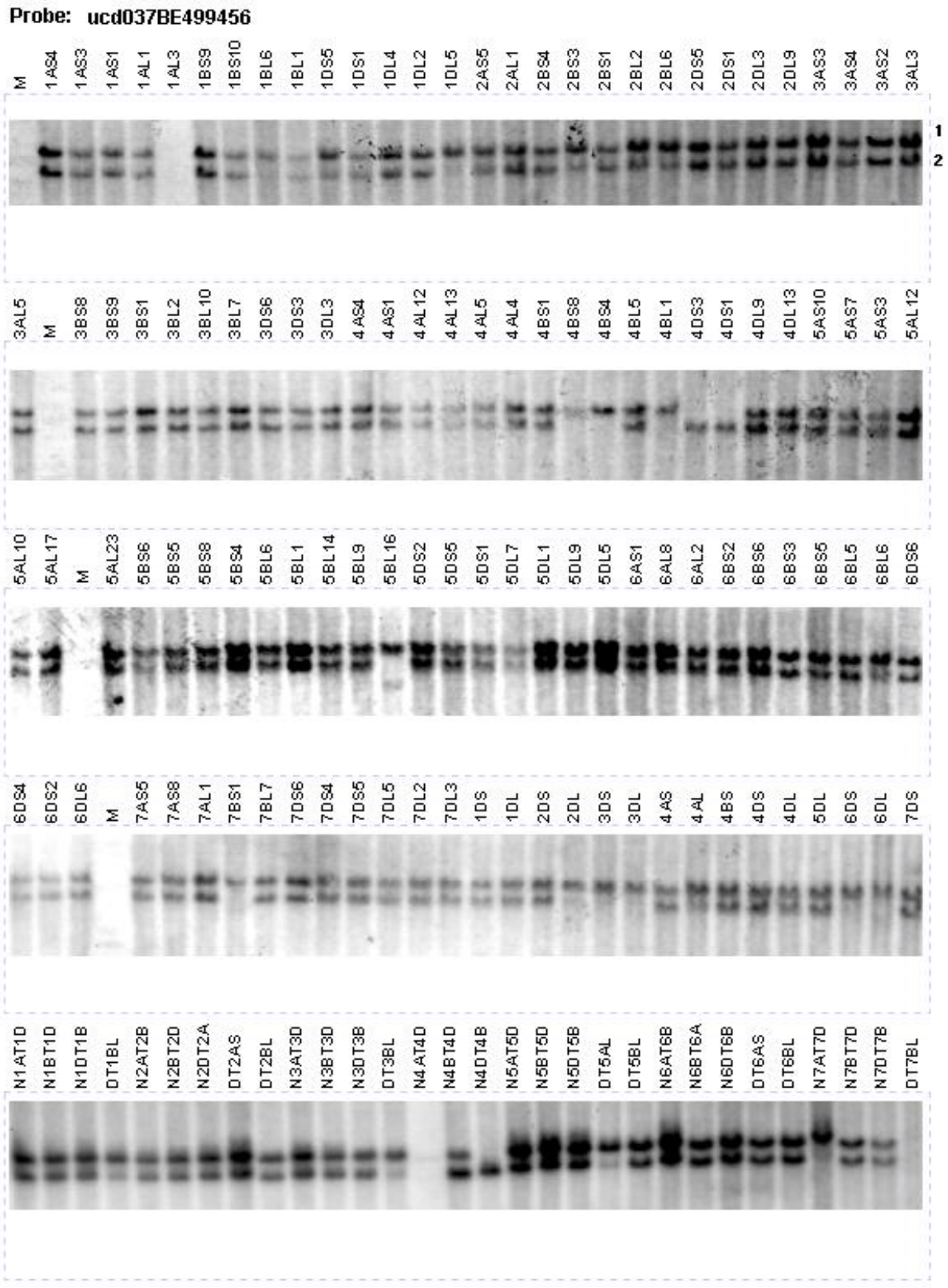


Figure V-4 The mapping data generated with probe BE499456 (Probe 4).

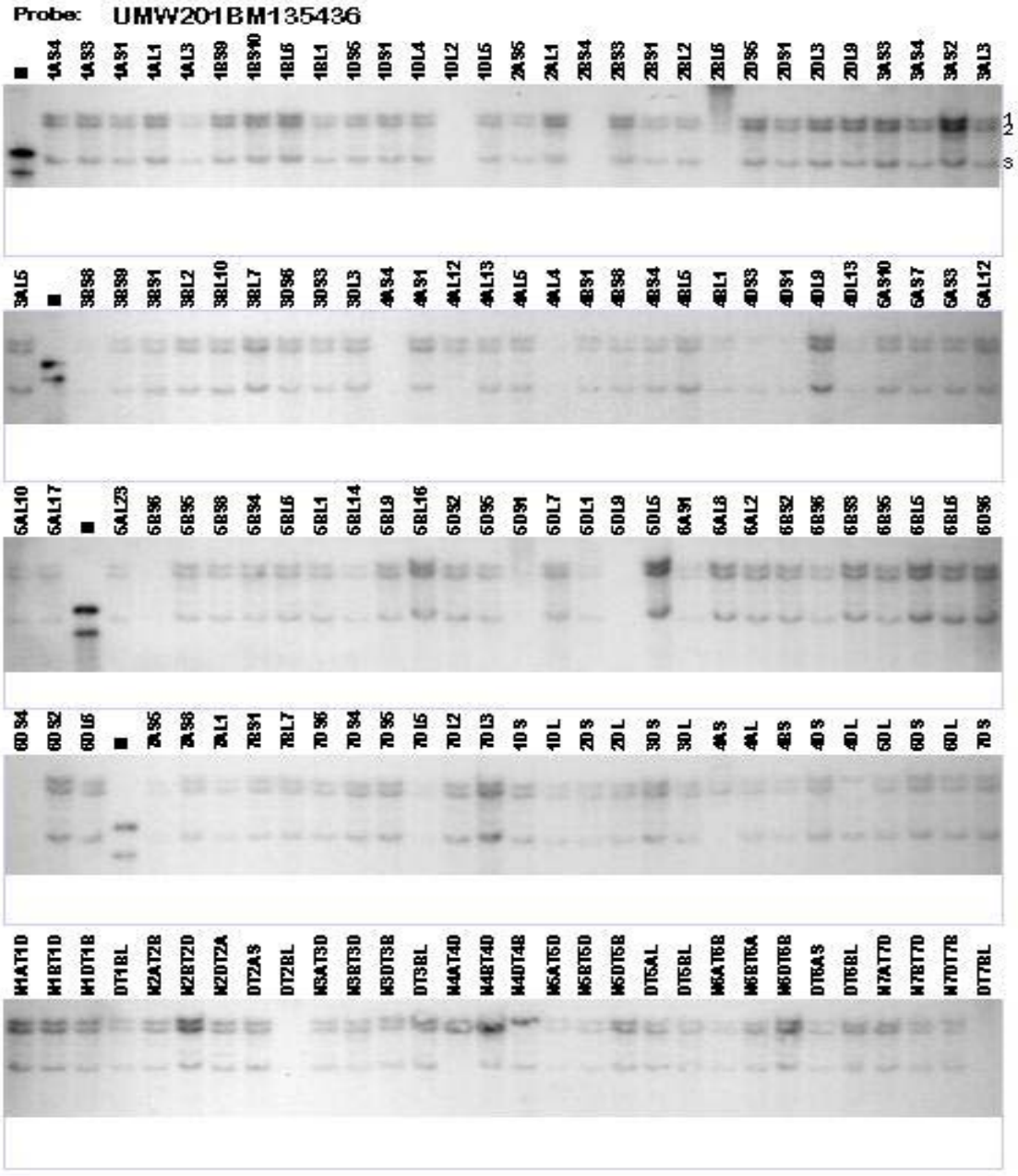


Figure V-5 The mapping data generated with probe BM135436 (Probe 5).

Legend for Figures V1-V5

This data was generated as part of the US Wheat EST project and was accessed from the GrainGenes-SQL database (http://wheat.pw.usda.gov/cgi-bin/westsql/map_locus.cgi). Details of the probes and the mapped loci are presented in Table 5.2.